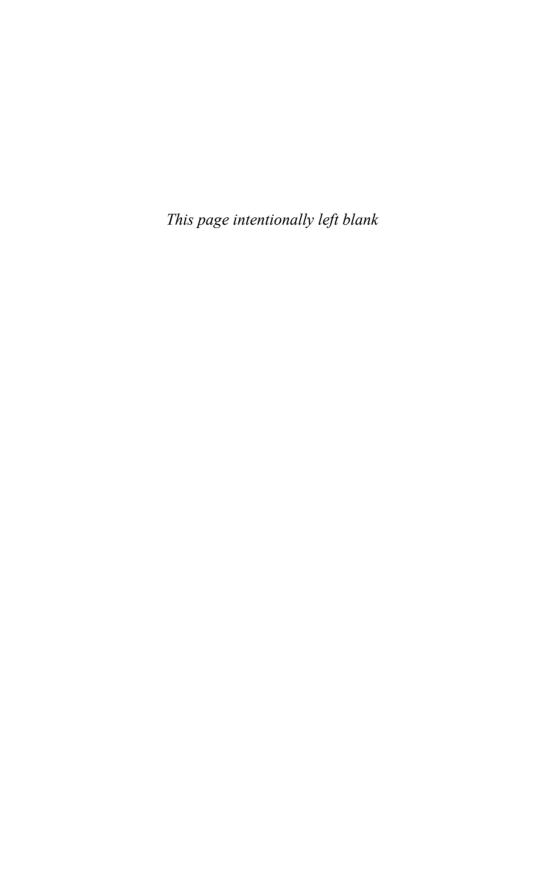# RATIONALITY AND THE GOOD

Critical Essays on the Ethics and Epistemology of Robert Audi

EDITED BY

Mark Timmons
John Greco
Alfred R. Mele

Rationality and the Good

Edited by
Mark Timmons
John Greco
Alfred R. Mele

This page intentionally left blank

Rationality and the Good

Critical Essays on the Ethics and Epistemology
of Robert Audi

Edited by
Mark Timmons
John Greco
Alfred R. Mele

OXFORD
UNIVERSITY PRESS

Preface

Very Brief Overview

For more than thirty years, Robert Audi has been one of the most creative and influential philosophical voices on a broad range of topics in the fields of ethics, epistemology, philosophy of mind and action, and philosophy of religion. This volume features thirteen chapters by renowned scholars plus new writings by Audi. Each paper presents both a position of its author and a critical treatment of related ideas of Audi's, and he responds to each of the other contributors in a way that provides a lively dialogue on the topic.

The book begins with an introduction by Audi that presents a thematic overview of his philosophy and connects his views in ethics, epistemology, and philosophy of mind and action. Each of the thirteen chapters that follow concentrates on one or another of these three main areas. The chapters are followed by Audi's replies. The exchanges between Audi and his critics in any one of the areas provides ample material for seminar discussions or researches in that field.

Ethics. Audi is the leading contemporary proponent of moral intuitionism. His 2004 book, *The Good in the Right*, defends a systematic ethical theory that provides a moderate intuitionist account of moral justification and knowledge together with a conception of morality and its pluralist structure that combines elements from the moral philosophies of Ross and Kant. Part 1 of this volume, "Problems and Prospects for Intuitionist Ethics," includes essays by Walter Sinnott-Armstrong, Roger Crisp, and Hugh J. McCann that challenge various key elements in Audi's moral intuitionism, especially its epistemology. Sinnott-Armstrong challenges Audi's distinction between "conclusions of reflection" and "conclusions of inference"—a distinction that plays an important role in Audi's defense of moral intuitionism. Crisp raises problems about the bearing of actual and hypothetical disagreement on the plausibility of Audi's intuitionism. McCann, though generally sympathetic to moral intuitionism, proposes to develop what may be described as a 'conativist' version of moral intuitionism that he presents as a corrective to the sort of 'cognitivist' view held by Audi.

The other chapters in part 1, by Bernard Gert, Thomas Hurka, and Candace Vogler, concentrate on Audi's conception of morality and his attempt to integrate Ross's moral pluralism with a Kantian unification of morality under the categorical imperative. Gert contrasts two conceptions of morality—a 'wide' conception prominent in such philosophers as Aristotle, Kant, Ross, and now Audi, and a 'narrower' conception to be found in the writings of Hobbes and Mill, and which Gert himself defends. One central element in Audi's normative moral theory is his attempt to integrate a plurality of Rossian moral principles with the Kantian categorical imperative as he interprets it. Audi calls his view *Kantian intuitionism*. In his paper, Hurka argues that this "marriage" of Kant with Ross does not yield the advantages to a Ross-style ethical pluralism that Audi claims. Finally, Vogler's paper challenges Audi's Kantian intuitionism by arguing that it fails to make proper contact with the views of either Kant or Ross.

Epistemology. Audi's epistemology is experientialist, moderately rationalist, foundationalist, realist, and aimed at being throughout consonant with a plausible philosophy of mind. Part 2, "Knowledge, Justification, and Acceptance," features essays by Laurence BonJour, Elizabeth Fricker, Timothy Williamson, and William Alston. Although BonJour shares Audi's epistemological foundationalism, he is critical of Audi's view that perceptual beliefs are among the types of foundational belief. Another important element of Audi's epistemology is his view about the social sources of justification and knowledge, particularly the epistemic status of testimony. The central epistemological question about testimony is how justification and knowledge arise from it. Fricker's paper is critical of Audi's view on this matter and defends her alternative against Audi's. Internalism and externalism are typically conceived as competing views. Audi, however, has defended an internalist view of justification and an externalist view of knowledge. Audi's blend of these views is the focus of Timothy Williamson's paper in which, among other things, he argues that this particular blend is unstable and, on some points, in error. Another element of Audi's overall epistemology is his conception of rationality in relation to religious faith. Audi has proposed a conception of such faith that expands the scope of rationality in the realm of cognitive attitudes to include what he calls 'nondoxastic faith', a fiduciary attitude that has less stringent rationality conditions than faith as usually understood. Although Alston agrees with Audi that this kind of positive attitude toward religious propositions is distinct from belief, Alston argues that Audi has not properly characterized the attitude in question.

Action, Mind, and Practical Rationality. Audi has also developed one of the most comprehensive and nuanced accounts of rational action and practical reasoning—an account that includes views on the concepts of intention and reasons for action that are crucial for ethics, particularly as they bear on matters of moral psychology. The chapters by Frederick Adams, Alfred Mele, and Raimo Tuomela in part 3 take up Audi's influential views on the topics of intention, self-deception, and reasons for action. Adams's paper is concerned with the concepts of intending and trying, arguing against Audi's view that trying is not entailed by intending. Relying partly on psychiatric studies, Mele is critical of Audi's views on self-deception and delusion. Finally, Tuomela defends a view of motivating reasons for both individuals and groups, contrasting his view with some aspects of Audi's theory of practical reasons.

In part 4, "Reason and Intuition in Thought and Action," Audi engages his critics by responding to their objections and, in many cases, refining and extending his own philosophical views. The responses are written to be read either straight through or in sections or subsections along with a single paper or section in the body of the book. Taken in the context of the many critical points Audi addresses, they constitute a rich source for continuing debate.

This page intentionally left blank

Acknowledgments

readers as a model of high-level intellectual exchange and will generate further contributions, by Audi and many others, to the central philosophical topics it addresses.

<div align="right">

M. T.
J. G.
A. M.

</div>

Contents

PART II: Knowledge, Justification, and Acceptance

PART III: Intention, Self-Deception, and Reasons for Action

PART IV: Reason and Intuition in Thought and Action

Contributors

Frederick Adams is professor of cognitive science and philosophy at the University of Delaware. He is chair of the department of linguistics and cognitive science and is director of the cognitive science program. He publishes in epistemology, philosophy of language, philosophy of mind, and theory of action.

William P. Alston has served on the philosophy faculties of the University of Michigan (1949–71), Rutgers University (1971–76), University of Illinois at Urbana-Champaign (1976–80), and Syracuse University (1980–92). From 1992 he has been professor emeritus at Syracuse University where he continued to teach until 2000. Among his books are *Divine Nature and Human Language*, *Epistemic Justification*, *Perceiving God*, *A Realist Conception of Truth*, *Illocutionary Acts and Sentence Meaning*, and *Beyond "Justification": Dimensions of Epistemic Evaluation*, all from Cornell University Press. In addition, he has published many articles in journals and books. He is a Fellow of the American Academy of Arts and Sciences.

Robert Audi is both the principal subject of this volume and author of chapters 1 and 15 through 17. Representative works of his are listed in the references following chapter 17. He is presently professor of philosophy and David E. Gallo Chair in Ethics at the University of Notre Dame.

Laurence BonJour is professor of philosophy at the University of Washington. He is the author of *The Structure of Empirical Knowledge* (1985), *In Defense of Pure Reason* (1997), and (with Ernest Sosa) *Epistemic Justification* (2003), along with many papers in epistemology and related areas.

Roger Crisp is Uehiro Fellow and Tutor in Philosophy at St. Anne's College, Oxford. He is the author of *Mill on Utilitarianism* (1997) and *Reasons and the Good* (2006).

Elizabeth Fricker is Fellow and Tutor in Philosophy, Magdalen College, Oxford. She has published extensively on testimony, and also in the philosophy of mind.

Bernard Gert is Stone Professor of Intellectual and Moral Philosophy, Dartmouth College, and adjunct professor of psychiatry, Dartmouth Medical School. He received an NEH–NSF Sustained Development Award (1980–84), Fulbright awards (Israel,

1985–86; Argentina, Fall 1995), and a National Humanities Center Fellowship (2001–2002). He is author of *Common Morality: Deciding What to Do* (2004), paperback edition (2007); and *Morality: Its Nature and Justification*, revised edition (2005). He is first author of *Morality and the New Genetics: A Guide for Students and Health Care Providers* (1996), and *Bioethics: A Systematic Approach* (2006); and is editor of *Man and Citizen*, Thomas Hobbes's *De Cive* and *De Homine* (1972, 1991).

Thomas Hurka is Chancellor Henry N. R. Jackman Distinguished Professor of Philosophical Studies at the University of Toronto, having taught previously at the University of Calgary. He is author of *Perfectionism* (OUP, 1993), *Principles: Short Essays on Ethics* (OUP, 1993), and *Virtue, Vice, and Value* (OUP, 2001), as well as numerous articles in normative ethical theory.

Hugh J. McCann is professor of philosophy at Texas A&M University. He is the author of *The Works of Agency* (1998) and of numerous papers in action theory, the philosophy of religion, and related topics in metaphysics, value theory, and practical reasoning.

Alfred R. Mele is the William H. and Lucyle T. Werkmeister Professor of Philosophy at Florida State University. He is the author of *Irrationality* (1987), *Springs of Action* (1992), *Autonomous Agents* (1995), *Motivation and Agency* (2003), *Free Will and Luck* (2006), all published by OUP, and *Self-Deception Unmasked* (2001). He also is the editor of *The Philosophy of Action* (1997), coeditor (with John Heil) of *Mental Causation* (1993), and coeditor (with Piers Rawling) of *The Oxford Handbook of Rationality* (2004) all published by OUP.

Walter Sinnott-Armstrong is professor of philosophy and Hardy Professor of Legal Studies at Dartmouth College. He recently published a monograph *Moral Skepticisms* (OUP, 2006) and is completing a collection of three volumes of original essays on *Moral Psychology* (2007).

Raimo Tuomela is professor of philosophy at the department of social and moral philosophy, University of Helsinki, Finland. His main field of research is philosophy of social action. He is a recipient of several grants and awards, including the von Humboldt Foundation Research Award, and is a member of the editorial board of several journals and book series. His recent books include *The Importance of Us: A Philosophical Study of Basic Social Notions* (1995), *Cooperation: A Philosophical Study* (2000), and *The Philosophy of Social Practices: A Collective Acceptance View* (2002).

Candace Vogler is an associate professor of philosophy at the University of Chicago. She works in ethics, action theory, social and political philosophy, feminism, sexuality and gender studies, and philosophy and literature. She has special interests in Marx, Aquinas, Rousseau, Elizabeth Anscombe, and the fates of foundationalist accounts in ethics. Her recent publications include *Reasonably Vicious* (2002) and *John Stuart Mill's Deliberative Landscape* (2001).

Timothy Williamson is Wykeham Chair of Logic at Oxford. He is the author of *Identity and Discrimination* (1990), *Vagueness* (1994), *Knowledge and Its Limits* (OUP, 2000), and *Vagueness* [with D. Graff] (2002), and of numerous articles in the *Journal of Philosophy, Philosophical Review, Mind, Analysis, Journal of Symbolic Logic, Journal of Philosophical Logic, Notre Dame, Journal of Formal Logic, Studia Logica*, and other journals and collections. In 1997 he was elected Fellow of the British Academy and of the Royal Society of Edinburgh.

Rationality and the Good

This page intentionally left blank

Rationality and the Good

An Overview

ROBERT AUDI

A child is born. It is immediately immersed in sensation and greeted with discomfort. It is relieved by fondling and feeding. It experiences touch and taste, hearing and sight, scent and movement. The earliest sensations are doubtless blurry. But soon, discrimination begins. This is a differential responsiveness to experiences. The child reaches out for milk, smiles at Mama, cries from loud noises.

It is not clear when the first beliefs are formed. Their formation is facilitated by discrimination, but belief-formation is not entailed by discrimination. Belief requires understanding. Whereof one cannot understand, thereof one cannot believe. Understanding, in turn, requires concepts. Concepts arise in interlocking formations. Beliefs do not arise in isolation either, one doxastic atom at a time. Like concepts, they are formed, and work, in families.

We are blessed not only with a receptiveness to learning from experience but also with a capacity to learn from what we already know. One route to learning is generalization: if a small yelping dog jumps at a child, the child expects much the same of a big one. Another, overlapping, route to learning is inference. Inference, too, appears early in life. We infer certain consequences of some of the things we believe. A child told that the family cannot have animals needs no logical prowess to infer that the puppy offered by neighbors will not be accepted. We also make inferences to the best explanation. The same child may infer from canine squeals at the door that the puppy is outside.

The picture so far drawn is intellectual. But just as belief and knowledge develop spontaneously from the impact of the world upon a child's experience, conduct evolves spontaneously as the child acts upon the world. Here the rewards of success and the punishments of error are great teachers. Action enriches the content of the intellect. We learn much by doing. Action also evokes desire and aversion, and so shapes the will.

I. Belief

I have been using metaphor. It can encapsulate theories, aid memory, and stimulate imagination. Permit me to use it more. My first metaphor is architectural. Buildings have both structural and material elements; they come in many kinds; they are strong or weak; they can be changed for better or worse; they are beautiful or ugly. They also have foundations and superstructure. A body of beliefs has such elements too. Foundational beliefs are grounded in experience or reason—or, we may say, just in experience if we take reason, in its grounding role, to work through intellectual experience.[1] Good grounds are solid; but not all grounds are bedrock, and even bedrock can be altered.

   The superstructure of a building is sustained by pillars. In cognition—say, in our belief systems—this sustaining role is often played by inference. But it is also played by a process of inferential belief formation that is more automatic than what we usually call *inference*.[2] Beliefs may produce others without our focusing on premises or drawing conclusions. The plurality of superstructures is indefinitely rich. There is no limit to what we can build, especially from good foundations. This limitlessness applies to both breadth and height. We all have foundational beliefs of ample scope and potential to empower us to make numerous inferences. From a single set of premises, we may go in many directions and as far as we like.

   If indefinite cognitive extension is possible, so is unending cognitive revision. As the pressure of wind can make us reduce the height of our construction, the force of criticism—or the sheer erosion of confidence as we reconsider—can make us reject what once seemed clearly true. Moreover, foundations can be rebuilt from superstructure as well as shifted from the fulcrum of their fellows. Deduction of untoward consequences from foundational elements is a common route to rebuilding them. The same holds for inductively inferred conclusions that oppose what we believe on the basis of experience or, perhaps, on testimonial authority. That authority, as Thomas Reid so clearly saw, is a social basis of knowledge.[3]

   The architectural metaphor should make clear something still not widely realized. The stereotype of epistemological foundationalism that has fueled so many postmodernist enterprises is groundless. Foundationalism as I am sketching it concerns the structure, not the content, of a body of knowledge or justified beliefs. It does not imply that knowledge or justified beliefs must have any particular type of content; nor that foundational beliefs are indefeasibly justified; nor that only deductive inference can carry justification from foundations to superstructure. And it provides a role for coherence to play in the rationality of our beliefs.[4]

II. Desire

I have described experience as engendering beliefs—though not every experience must do so.[5] It is not just perceptual experiences that do it; "internal experiences,"

such as imaginings, can also do it. This applies to desire too. The sight of a fruit bowl can evoke an appetite; imagining a sip of fine wine can arouse desire.

Consider a second, arboreal metaphor, parallel to the first. A tree is grounded in soil, the main source of its nutrients. Its roots anchor it; they are its foundations. Its trunk and branches serve as pillars in the superstructure. Its foliage shows its scope and character. If it is well grounded—being in good soil, nourished by its roots, and carrying nutrients along the normal pathways—it flourishes. Is *our* flourishing (as a certain Humean instrumentalism says) simply the fulfillment of our basic desires? Or might there be good and bad nutrients that yield desire and good and bad ways to transmit their influence to the superstructure? Surely the latter view is more plausible. But what of desire? Isn't getting what we want—satisfying our basic desires—constitutive of a good life? Not necessarily. Desire is both fallible and manipulable.

With the arboreal metaphor in mind, let us go back to nature. From babyhood onward, pleasure and pain are among the elemental nutrients of desire. They stimulate conative growth toward the pleasurable and away from the painful. This is not to endorse hedonism. But perhaps if we were not built so as to enjoy some things and be pained by others, we would not learn to want anything. Still, genetic primacy is one thing, motivational hegemony quite another. It may be that we would not learn to value nonhedonic goods if we were not first motivated by hedonic ones; but our early years under the tutelage of pleasure and pain need not prevent our developing autonomous desires. Loving our parents may begin with their relieving our pains and giving us pleasures, but it does not end there. And if love has roots in our own pleasure and desire satisfaction, its growth requires learning to care about the well-being of others. Recall some of what Paul says in 1 Corinthians 13:

> Love is patient and kind; love is not jealous or boastful; it is not arrogant or rude. Love does not insist on its own way; it is not irritable or resentful; it does not rejoice at wrong, but rejoices in the right. Love bears all things, believes all things, hopes all things, endures all things. (vv. 4–7)

There are many kinds of trees and many shades of foliage. The foliage of a single tree changes in color and is seasonally replaced. There is continuity as well as plurality. There are also structural differences between trees. Brittle trees are hurt by the stresses of wind and the weight of snow. A tree that can bend need not break. (The implicit lessons, especially for parents and teachers, are numerous.)

There are cognitive analogues of all these points. The dogmatic, for instance, tend to have stiff trunks; the fallibilistic can be resilient. The unpredictable, many-faceted growth of trees is a metaphor for the development of desire as well as of belief. Much as, in almost any realm we are exposed to, theoretical inference generates new beliefs, practical inference generates new desires. The most pervasive kind of practical inference is instrumental.[6] Wanting a good meal can lead, by way of the instrumental belief that cooking lobsters would provide it, to the conclusion that this is the thing to do and, through that, to wanting to cook lobsters. Desires also arise without reasoning. Tasting lobster for the first time may evoke a desire for more.

III. Action

The agent so far portrayed has belief and desire and the makings of intellect and will. Experience is what I have above all emphasized. Nothing experienced, nothing discriminated. Nothing discriminated, nothing believed. No pleasure or pain, or other positive or negative experiences, nothing wanted. Nothing wanted, nothing done.

But, however many beliefs and desires we have, neither belief nor desire entails intention. Desire is a pressure toward intention; but we resist some pressures and eliminate others. Why is intention so important? As Kant saw, good will—the volitional heart of good people—is a construct from intention.[7] In our character, intention is fundamental. Both points are central for ethics. But they are significant in part because of another point.

We come now to my third metaphor: the itinerary. Intention is essential for putting a destination on our itinerary. By sheer good fortune, we may land in wonderful places; but good lives require itineraries. We can revise them often, but we need a sense of where we are going. If we simply wait for life to take us where it will, we are impoverished. We may exercise freedom, but not autonomy; we may have pleasant surprises, but not the enduring satisfactions of earned achievements.

A good itinerary requires a good map, but even the best of maps does not tell us what path to take. A good itinerary rewards us by leading us to worthwhile destinations. Our belief system is our map of the world (though it is far more). Our desires are our inclinations toward destinations. Desires that prevail in our intentions put destinations on our itinerary. If we are rational, knowledge or at least justified beliefs underlie our map of the world, and we have worthwhile destinations on our itinerary.[8] We cannot be rational without minimal rationality in both the theoretical and practical domains.

In both domains, experience has high normative authority. Sensory experience provides basic grounds for beliefs about the world; intuitive and ratiocinative experiences yield grounds for beliefs with logical and other a priori content; and logic constrains what beliefs we may hold on the *basis* of other beliefs. Rewarding experiences—most clearly (though not exclusively) those that are pleasurable or are marked by relief of pain—provide basic grounds for rational desire.

One might think, as Humean instrumentalists do, that desires do not admit of rationality, though they can be *irrational*, as where we can easily see that their objects are impossible.[9] On this view, practical rationality consists simply in maximizing the satisfaction of basic (noninstrumental) desires.[10] But instrumentalism misses a profound parallel: just as, in virtue of a clear and steadfast visual impression of faces before me, it is (prima facie) rational for me to believe there are faces, so, in virtue of being pained by the touch of a hot kettle, or of enjoying conversation with a friend, it is rational for me to want to avoid the former and to have the latter.[11] We do not flourish in just any soil; and some destinations should never appear on our itineraries.

IV. Value

The parallels I have suggested between theoretical and practical reason leave room for important differences. The broadest is perhaps this: belief is, in a certain objective

way, successful when its object is *true*; desire—in the basic cases—is, in a similar objective way, successful when its object is *good*.[12] What is objective may, of course, be internal: there are truths about matters internal to the mind, and there are objective goods internal to experience, such as the enjoyment of a silent recitation of a beautiful sonnet. The same holds for the pain of vividly recalling an injustice.

How should we think of value, the realm of the good and the bad? All my metaphors apply. Take goodness. I see it as grounded in qualities of experience. The experiential soil can be good or bad, nourishing or desiccating. There are fertile fields of grain and barren sandy deserts. And much as there are foundational beliefs and others based on them, there are basic goods and instrumental goods that lead to them.

Here I introduce a fourth metaphor: the aesthetic. Consider paintings and poems. Both can have value "in themselves." But the phrase 'in itself' is too coarse to stand alone in clarifying basic value—*intrinsic* value. The phrase does encompass the goodness of intrinsically good experiences: these are *noninstrumentally* good; they need have no relational kind of goodness. But 'good in itself' also applies to things whose goodness is not experiential. The goodness of a beautiful painting resides in its beauty, and that, in turn, is consequential upon its intrinsic properties (its nonrelational ones). This makes it natural to think of the painting as good *in itself*.

Aristotle implicitly spoke to this question. For him, one good is more "final" than another if we seek the latter for the sake of the former, and *the* good—that which makes life "choiceworthy"—is not sought for the sake of anything else.[13] We value beautiful paintings in order to *view* them with a certain kind of reward. Viewing them in that way yields an aesthetically valuable *experience* (i.e., one that is good from the aesthetic point of view, not one that is a good *object* of aesthetic appreciation, though that status is not ruled out for special cases). Experiences that have such value are intrinsically good. They are also "more final" in Aristotle's sense than their objects. Do beautiful paintings contribute to the ultimate choiceworthiness of life simply by their physical existence around us, or through our *viewing* them—hence visually experiencing them—in a way that is aesthetically good? Plainly they would not so contribute if we never viewed them or, upon viewing them, we never had a good experience. Good things are good in virtue of the experiential qualities that enable them to contribute to good lives.

Call artworks and other things that are good in themselves but not intrinsically good, *inherently good*. They can be *constituents* in, and not merely *means* to, experiences that are intrinsically good. They are thus not merely instrumental goods and are sources of noninstrumental reasons for action, for instance for viewing paintings "for their own sake."[14]

The inherent good shares another property with the intrinsically good: it is organic. The value of an organic whole need not be the sum of the values of its parts or aspects. It can have parts and aspects that have no inherent or intrinsic value, such as a blank space in a painting, a harsh dissonance in a symphony, and ellipsis marks in a poem. But the overall inherent value of these artworks may be positively affected by such elements, so that the value of the whole is greater than the sum of the values of the parts or aspects. This can hold even if all of the parts and aspects are inherently good. Similar points apply to intrinsically valuable experiences (and even to disvaluable ones).

Consider a pause in musical work: experienced in itself, it may be aesthetically empty but, as part of the overall musical experience, valuationally important. Or take Shelley's wonderful lines about Ozymandias, a king who vaingloriously sought immortality in a statue that is now decayed by the ravages of nature:

> Nothing beside remains. Round the decay
> Of that colossal wreck, boundless and bare,
> The lone and level sands stretch far away.[15]

The empty silence of its surroundings is far more memorable than the statue; and the contrast between the two heightens our sense of each. Listen to Shakespeare's sonorous invocation, when Prospero summons his supernatural minions to do his final bidding. It depicts landscape and seascape, the swift and delicate, the dance and the chase, the playful and the powerful. He says:

> Ye elves of hills, brooks, standing lakes and groves,
> And ye that on the sands with printless foot
> Do chase the ebbing Neptune and do fly him
> When he comes back; you demipuppets that
> By moonshine do the green sour ringlets make,
> Whereof the ewe not bites; and you whose pastime
> Is to make midnight mushrooms, that rejoice
> To hear the solemn curfew, by whose aid—
> Weak masters though ye be, I have bedimmed
> The noontide sun, called forth the mutinous winds,
> And 'twixt the green sea and the azured vault
> Set roaring war.

<div align="center">(The Tempest, 5.1.34–44)</div>

Some of his words are individually evocative, but others are workaday tools we use with no sense of aesthetic value. Yet, joined together in these incomparable ways, they are uniquely rewarding. In a single sentence, we have depiction, narrative, dramatic movement, and powerful resolution.

If intrinsic value is organic, and if the intrinsically good has the normative authority that goes with its grounding of reasons for action, we should find that *overall* reason for action may also be organic. Recall the role of pleasure and pain as grounds of rational desires and hence prima facie reasons for action. *In* the fact that criticizing a discouraged student's paper I would cause pain, which is intrinsically bad, I have prima facie reason *not* to do this. But, in general, in the fact that I would enjoy doing something, I have prima facie reason to *do* it. Now imagine that my mood is sadistic and I would greatly enjoy causing pain. Suppose that intrinsic value, and the practical reasons it grounds, were additive. Then, if the student's pain would be minor and my pleasure great, there might be overall reason for me to do the deed. But this pleasure ill befits its object, so much so that the *overall* value of the sadistic pleasure is negative—and *less* than that of an equally intense, equally lasting pleasure in something valuationally neutral.[16] Here, then, there would be better reason for me to avoid the sadistic deed than to perform it.

V. Obligation

If the good and the bad are sources of reasons for action, and if the pursuit of the former and the avoidance of the latter are as important in human life as they seem, might the deontic realm be subordinate to the axiological? Are the right and the wrong ultimately derivative from the good and the bad? I include the obligatory—in the widest sense of what we ought to do—in the realm of the right and the wrong; I conceive what is obligatory in this widest sense as what it would be wrong not to do.[17]

There is certainly a sense in which it is bad to do what we ought not to do and good to do what we ought to do. And isn't it true that the more good we bring about, the better? These points make it natural to believe—as consequentialists in ethics do—that the right is subordinate to the good. If, as I think, the good is organic, this projected subordination of the deontic to the axiological can be developed in a way that takes account of that point. *Maximize the good* would be our categorical imperative.

If you are imagining Kant turning over in his grave at this thought, remember that some goods are realized only in action. Excellence is achieved *in action*, for instance in intellectual and aesthetic activities. Moreover, some goods, such as a just distribution, are *moral*.[18] These points suggest a better formulation connecting the deontic and the axiological; instead of requiring maximization of the good, our imperative might be more Aristotelian: *Realize* the good. This imperative allows that the good we bring about may be not an external consequence of the action by which we produce it, but intrinsic to it.

If we can frame an adequate realization theory of the basis of obligation, we may be able to derive sound standards of moral obligation from a conception of what is needed to realize the good. I leave open, then, that from a certain kind of organic theory of the good, we might derive sound standards of the right. But since *moral* goodness would be included at the base, this would not be a consequentialist project like Mill's: a derivation of the right from the *nonmorally* good. Hence, even if the standards of right and wrong are in some way derivable from those of the good and the bad, the corresponding deontic *concepts* need not be reducible to axiological concepts.

The architectural metaphor is clarifying here. Two independent foundational girders can jointly support the same superstructure. Similarly, certain moral standards can be supported both by meeting deontic demands that *right* action must satisfy *and* by meeting axiological requirements that *good* action must fulfill. I mean standards of prima facie obligation, such as Ross's in *The Right and the Good* (and those I develop and defend in *The Good in the Right*). There are obligations of justice and noninjury, of fidelity and veracity, of beneficence and self-improvement, of reparation and gratitude, and of liberty and respectfulness.[19]

The principles expressing these obligations seem to have a kind of epistemic independence relative to axiological principles. Indeed, I consider these principles of obligation self-evident, in this sense: first, an adequate understanding of them suffices for being *justified* in believing them; second, if we believe them on the *basis* of such an understanding, we know them.[20] The self-evident need not be obvious or even uncontroversial. It is simply accessible to reason in a certain way. Many self-evident truths are never in fact accessed; at most a few are on everyone's cognitive map. But

it is a blessing of our nature that so many can be readily seen and, if sometimes with difficulty, internalized.[21]

If this moral epistemology is sound, and if moral principles provide (prima facie) reasons for action, then there are moral reasons for action, say, to avoid injustice, to keep promises, and to relieve suffering.[22] Moreover, through our understanding of the principles expressing the obligations that correspond to these reasons, the intellect can ascertain moral reasons for action. This entitles it to a practical function in guiding action.

Does the intellect, then, have executive power? Perhaps it sometimes does, but I doubt that it must.[23] Its motivational power seems to depend on the cooperation of the will or at least on supporting desires. But the will is no mere handmaiden of desire. It tends to respond to natural desires that are rational. Compassionate desire, for instance, can mitigate anger and lead to irenic intentions where enmity would have ruled. But the will is also guided by practical judgment, and some of its deliverances can guide desire and even the intellect. The will may be wayward and misguided, but in a rational person it tends to support moral judgments: not only the self-addressed ones that express a sense of obligation but also practical judgments about what is right, needful, wise, or otherwise called for.

VI. Rationality

We now have a sketch of a conception of rationality in three interconnected domains: the theoretical, the practical, and the moral. The moral, being both a realm of knowledge and a source of standards essentially concerned with guiding action, is at once theoretical and practical. Theoretical rationality is a central concern of epistemology, practical rationality of the philosophy of action. Rationality in the moral domain cuts across these realms. What grounds rational moral judgment? And what are the rationality conditions of moral action? Moral judgment is practical in content; moral action is practical in nature.

Rationality is also a property of persons themselves—this is *global rationality*. It requires both theoretical and practical rationality. Recall the vital itinerary. Without a rational cognitive map, we would find worthwhile destinations only by good fortune; without desires and intentions to go to worthwhile destinations, a good map would not help us to live a rewarding life.

In a globally rational person, the practical and the theoretical are integrated. Emotions bear on this integration. To have no feelings about what one judges to be good or condemns as wrong is to be in a certain way impoverished. To have feelings like those of anger where one sees nothing wrong is to be in a dissonant and disorienting condition. These feelings, like anxiety at the prospect of flying despite a judgment that it is safe, ill befit one's cognition. Feeling and emotion can be fitting or ill befitting to cognition, and, in a rational person, fittingness in this realm predominates. We should be pleased by our friends' successes and distressed at their misfortunes. Indignation befits the sight of a confidence trickster cheating an old man; it ill befits the experience of being asked to wait in line to be served.

Intellectualist associations commonly surround the notion of rationality. But emotions can be rational as well as irrational. They may be called for by what a

person perceives, as where the roar of an apparent landslide makes fear rational. But fear may also be inappropriate to experience, as where one should realize that the roar comes from an airplane. Emotions may be integrated or discordant with other elements in us, such as beliefs and desires. Hatred of others may be unreasonable given our theology or our factual beliefs about others, or both.

On the integration of emotion with intellect I find a poem of Emily Dickinson's the *mot juste*:

> The Heart is the Capital of the Mind—
> The Mind is a single State—
> The Heart and the Mind together make
> A single Continent.
>
> One—is the Population—
> Numerous enough—
> This ecstatic Nation
> Seek—it is yourself.[24]

The importance of emotions helps us to see why (as I have said) the will is not the handmaiden of desire; it is not even the handmaiden of the intellect. Consider weakness of will, in a paradigmatic form in which we act against our better judgment. A father might judge that he must punish his teenage daughter for staying out late and frightening her parents. But, when it comes to issuing the punishment, he looks into her fearful, distressed face and feels compassion. He turns away, feeling foolish but relieved. Might his emotions and sensitivity to the child's vulnerability run ahead of his punitive judgment—and deeper? Perhaps his judgment, though rational given his generally reasonable standards, ill befits his deepest values. His action responds, partly through the influence of his emotions, to these values; and it can be rational despite opposing his judgment. Its global integration with what is deepest in him and—we may suppose—perfectly rational outweighs its local disparity with his judgment at the time.[25]

We are naturally endowed with at least two interacting ways to determine what destinations are worth visiting. If we are fortunate, if we are loved, cared for, and educated, worthwhile experiences and activities will be prominent in our lives. These experiences produce desires and intentions. If we are unfortunate early in life, our actual path through these years will be clouded. But our itinerary can still be informed by suffering: pain is a powerful teacher, and aversion is more easily educated than desire. Cognition takes stock of what we like and dislike, and influences both desire and aversion. Emotion is responsive to both cognition and desire, but it may also beneficially influence them.

Must morality play a part in the constitution of a rational person? Must others be important to us as more than means to our own ends? This much seems very likely true. If there are (as I think) objectively ascertainable—indeed, a priori ascertainable—reasons for action, and if these have high normative authority, then moral conduct well grounded in them is never *irrational*. Still, this point does not imply that every rational person must actually believe moral principles and thereby have moral reasons for action.[26] But suppose that is so. There remains a question whether moral action is always rationally required.

People differ in their experiences and hence in their grounds for rational belief and rational desire. There is, then, *one* kind of relativity that even an objectivist account of rationality must countenance: relativity to grounds. This point leaves room to argue that for *certain* people with a certain range of experiences and beliefs, some moral conduct is rationally required. My view, however, is more measured: for people who (like most of us) live in friendly relations with others and believe that we are all alike in basic rationality, motivation, and sentience, morality is a *demand* of reason. By and large, even where it would not be irrational for such people to fail to do what morality requires, it would be *unreasonable*.[27]

Much should be said about the reasonable. Like rationality, reasonableness is a kind of responsiveness to reasons and, ultimately, to experience. But reasonableness is a stronger notion and implies a greater responsiveness. A rational person, such as a shrewdly selfish one, can be quite unreasonable. One might think that to be reasonable is above all to do good reasoning, as opposed to merely being rational. But my conception of both rationality and reasonableness is less intellectualist. Reasonable belief need not be *reasoned*; it need not be grounded in any inferential process, even when based on other beliefs.[28] And reasonable action need not be reasoned action. It need not be based on practical reasoning. The rationality and reasonableness of beliefs, desires, or actions depend on their being well grounded; their well-groundedness is a matter of the kind of basis they have, not of the process by which they are arrived at. God never *has* to reason.

<p style="text-align:center">* * *</p>

Experience is the raw material of human life.[29] But, unlike any ordinary fabric, it is neither decorative nor instrumental. It is constitutive. In its absence there would be not the statuesque forms that have inspired artists and sculptors, but a naked substratum.

Fabric can also serve as a metaphor for the organicity of goodness. Fabrics may have indefinitely many patterns or none at all. Some patterns are coherent, some not. But coherence alone is not enough for goodness. Good design is more than coherence, and even good design is enhanced by good content: by the colors and shapes whose presence and relationships yield beauty.

Even the best fabric can be indelibly stained or permanently discolored. It may be sullied by dirty hands, as well as smoothed by caring ones. It may be loosely woven and easily frayed, or tightly sewn and impossible to rip. It may befit its wearer and delight its viewer. If it fits badly, we can alter it; if it is torn, we can repair it. We can hold it up to the mirror of self-scrutiny, expose it to the appraisal of our peers, and test it in heavy wind and deep water.

Philosophy is central for achieving the widest and deepest self-scrutiny. It articulates and refines our methods of analysis. It discerns the structure of our worldviews—their foundations and their modes of construction from their base—and it probes the solidity of its own grounding and the strength of its support of its own superstructure. Philosophy draws on experience, but need not accept its content at face value: neither the deliverances of the senses nor even the intuitions of reason. There are rigorous rational criteria that guide philosophical inquiry: above all, the constitutive intellectual standards in the realms of perception, reason, introspective thought, and memory, interacting with the socially constituted standards of testimony, dialectic, and experimental inquiry.

Theoretical rationality is a kind of responsiveness to experience, a responsiveness governed, however unself-consciously, by these interconnected standards. But experience provides raw material for the will as well as for the intellect. There are rewarding experiences—especially the multifarious pleasures of human relationships, of the intellect, of the aesthetic sensibilities, of reaching spiritual heights; and there are aversive experiences—particularly those of suffering: the agony of injury, the grief at love's loss, the anxieties and fears brought by a threatening world. Desire should be directed toward rewarding experiences and away from aversive ones. Practical rationality is a kind of responsiveness to these grounds of the good and the bad. It is a responsiveness in which, often enough, rational desires rise to intentions and thereby inform the will. When theoretical and practical rationality are well integrated in us, we are globally rational. And when our rationality is informed by a deep enough knowledge of others, inspired by curiosity, and blessed by imagination, we are capable of successfully pursuing both truth and goodness.[30]

Notes

1. Intellectual experience must be understood broadly, to include any kind of abstract thinking but also our considering propositions and our making inferences. Memory is an indirect source of foundational beliefs whenever one retains a belief having lost from memory propositions ("premises") on which it was originally inferentially based. Note too that not all foundations—cognitive or physical—need be built upon.

2. The crucial distinction here is between structurally and episodically inferential beliefs. This distinction is made and elaborated in "Structural Justification," *Journal of Philosophical Research* 24 (1991): 473–92, reprinted in *The Structure of Justification* (Cambridge: Cambridge University Press, 1993). It is developed further in *Epistemology*, 2nd ed. (London: Routledge, 2003), chap. 6.

3. For a discussion of Reid's view of testimony and an indication of why testimony-based beliefs are plausibly considered noninferential, see my "The Epistemology of Testimony and the Ethics of Belief," in *God and the Ethics of Belief*, ed. Andrew Chignell and Andrew Dole (Cambridge: Cambridge University Press, 2005).

4. In *The Architecture of Reason* (Oxford: Oxford University Press, 2001), I have criticized this stereotype of foundationalism in detail and indicated why it is not even adequate to plausible versions of the foundationalism of Aristotle and Descartes, though more nearly to theirs than to mine. See esp. chaps. 1–2.

5. This distinction is developed and defended in my "Dispositional Beliefs and Dispositions to Believe," *Nous* 29 (1994): 419–34.

6. In a wide, structural sense of the term, all practical inferences are instrumental: even reasoning from a desire for a pleasant respite playing the piano and a belief that playing it now would be pleasant, to a conclusion favoring that action is instrumental if we consider the belief to express a constitutive means (as opposed to an ordinary instrumental means) to the desired pleasant break. Playing it now partly *constitutes* the desired end, hence it is not an ordinary means to it.

7. See Kant's *Groundwork* for an identification of "its willing" as intrinsic to the will.

8. Only normally because rationality is a capacity concept and I do not assume that a person must have experiences yielding the "cartographic" justification or knowledge in question, nor am I ruling out the possibility of a certain kind of Cartesian demon massively influencing its victims.

9. If, as I assume, what we believe to be impossible (logically or at least nomically) we cannot want, but only wish could be, then the kind of desire in question would not occur, given a belief to this effect. Granted, the counterpart wish could have much in common with the desire. Just how it would differ is a matter for reflection that may not yet have been done.

10. Extensive critical appraisal of instrumentalism is provided in my "The Naturalization of Practical Reason: Humean Instrumentalism and the Normative Authority of Desire," *International Journal of Philosophical Studies* 10, no. 3 (2002): 235–63.

11. The rationality in question is prima facie; I am speaking of defeasible rationality and will also presuppose defeasibility in speaking of justification. For convenience I will at times drop 'prima facie' hereinafter.

12. Two comments are needed here. First, the object may be only instrumentally good, but this is still a kind of goodness implying that there *is* something intrinsically good (on the assumption that instrumental goodness is not merely effectiveness as a means—something the expression "good as a means" does not rule out). Second, there *is* an objective sense in which a desire is successful when its object is merely realized; but this is not the kind of success analogous to truth as believed in the way I have in mind: being known or believed with a certain kind of justification. (The analogy is developed in some detail in *Architecture*.)

13. Aristotle says, "We call that which is pursued as an end in itself more final than an end which is pursued for the sake of something else." See *Nicomachean Ethics*, trans. Martin Ostwald (Indianapolis: Hackett, 1962). I take Aristotle to be referring not to mere pursuit but to a kind that is "proper." The idea of the self-sufficiency of happiness (the good), in virtue of which it makes life choiceworthy, is explicit in 1097b. Note that we can accept the idea that a life is choiceworthy in virtue of realizing intrinsic goods (and, overall, *the* good) and still question the idea—which I find plausible so far as it is clear—that the good cannot be sought for the sake of anything else.

14. Two points will add clarity here. First, we *can* view a painting instrumentally, say visually study it as a means to some further end, such as learning about the tastes of a friend; but this is not the way it is normally designed (or "supposed") to be viewed. Second, the idea of the inherently good's capacity for *contribution* to good lives should be taken broadly if we are to do full justice to the value of persons (a point not brought out in my earlier work on value). Persons (or, on a Cartesian view, minds) conceived as essential "substratal" subjects of experience, are a kind of *constitutive* contributor to good lives. They themselves may of course *also* be objects of intrinsically good experiences, including experiences of oneself.

15. Percy Bysshe Shelley, "Ozymandias," in *Seven Centuries of Verse*, ed. A. J. M. Smith (New York: Charles Scribner's Sons, 1957), 362.

16. *Schadenfreude* is treated in my "Intrinsic Value and Reasons for Action," *Southern Journal of Philosophy* 41, no. 3 (2003): 323–27, reprinted in *Metaethics after Moore*, ed. T. Horgan and M. Timmons (Oxford: Oxford University Press, 2006). This paper also indicates how the
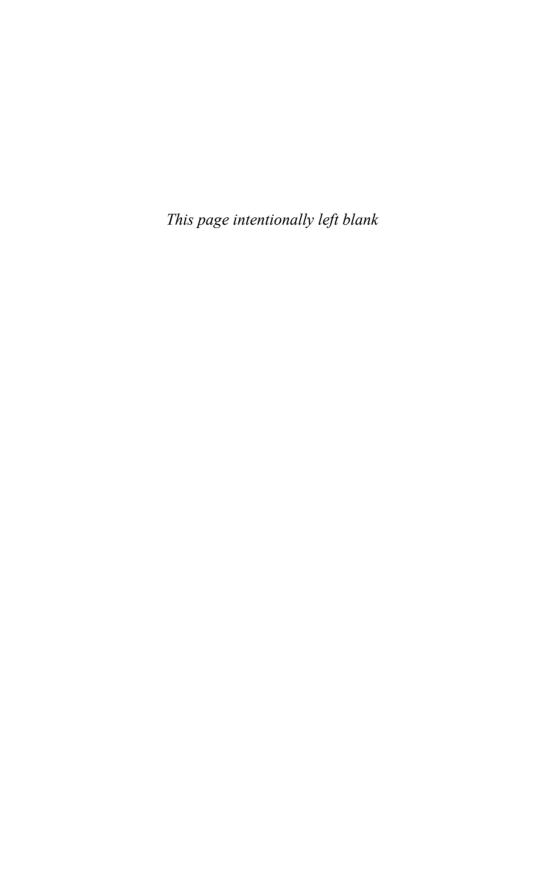
inherent value of an organic whole may be less than that of the sum of the inherent values of its parts or aspects.

17. Not all obligations, then, represent perfect duties; there are indeed things we ought to do, and perhaps in the widest sense have an obligation to do, that are not strictly duties at all, but represent the demands of what I call involuntary ideals. These points are discussed in my "Wrongs Within Rights," *Philosophical Issues* 15 (2005): 121–39.

18. That there are moral experiences and that some of these are intrinsically good is argued in my "The Axiology of Moral Experience," *Journal of Ethics* 2 (1998): 355–75.

19. I offer an explication of these prima facie obligations in *The Good in the Right* (Princeton, N.J.: Princeton University Press, 2004), esp. chap. 5.

20. I defend these points about self-evidence in my paper of that title in *Philosophical Perspectives* 13 (1999): 205–28.

21. Some moral truths and certain epistemic and logical principles are included here. I suggest that in both categories there are some easily seen but not easy to internalize, some easy on both counts (such as certain very elementary logical truths), and others difficult on both. What constitutes internalization deserves analysis; here I will only say that internalization admits of degree and it affects the guidance of thought and action.

22. For an ethical theory on which the "positive duties," such as that of beneficence, are represented as ideals rather than obligations, see Bernard Gert, *Common Morality* (Oxford: Oxford University Press, 2004).

23. I have provided an account of the motivational power of moral judgment in "Moral Judgment and Reasons for Action," in *Moral Knowledge and Ethical Character* (Oxford: Oxford University Press, 1997). This paper also addresses the conception of the practical power of intellect and supports the distinction suggested in the text between executive power—which I take to entail the power to produce intentions—with the weaker notion of motivational power, which is a matter of producing desires. I am also implicitly distinguishing in the text between an independent power and one that a "faculty" has in cooperation with another.

24. From *Final Harvest: Emily Dickinson's Poems*, ed. Thomas H. Johnson (Boston and Toronto: Little Brown and Co., 1961), 278.

25. For an earlier version of this example and a theoretical account of the possibility of rational action against one's better judgment, see my "Weakness of Will and Rational Action," *Australasian Journal of Philosophy* 68, no. 3 (1990): 270–81.

26. Chapters 6 and 7 of *Architecture* explain how a life may be such that certain a priori principles do not become part of a person's belief system and how a life might be narrow enough to make it possible not to have moral reasons (this would not entail, of course, lacking reasons for "moral actions," in the weak sense of 'actions required or permitted by sound moral standards').

27. Relativity of various kinds is treated in detail in chap. 7 of *Architecture*. Reasonableness and, in relation to it, the status of moral reasons relative to other kinds, is treated in chap. 6.

28. I have argued in many places for the possibility that a belief or action *for a reason* need not be reasoned. See, e.g., *Practical Reasoning and Ethical Decision* (London: Routledge, 2006).

29. In one way this metaphor is misleading: one can't "take off," one's experience, except by sleep or by death, and in those cases one cannot observe the result.

PROBLEMS AND PROSPECTS
FOR INTUITIONIST ETHICS

This page intentionally left blank

Reflections on Reflection in Robert Audi's Moral Intuitionism

WALTER SINNOTT-ARMSTRONG

Everyone who relies on moral intuitions—which is everyone—should welcome Robert Audi's masterpiece, *The Good in the Right* (2004).[1] This tour de force develops the most sophisticated and systematic version of moral intuitionism ever.

Each part of Audi's multifaceted gem is valuable. It is impressive how tightly they fit together and how much light they cast on nearby topics. Nonetheless, I will focus on just one small feature of Audi's version of moral intuitionism as a view in moral epistemology.

Intuitions versus Intuitionism

Audi defines a *moral intuition* as a noninferential, firm, comprehended, and pretheoretical moral belief (33–36). To call a moral belief noninferential is to make a descriptive psychological claim about its cause, namely, that it "is not—at the time it is intuitively held—believed on the basis of a premise" (33).

Moral intuit*ism*, in contrast, is a normative view about the epistemic status of some moral intuitions. The central claim of moral intuitionism is that some moral believers are justified or know in a special way: noninferentially (2, 21–22, 44, etc.). A believer is justified *noninferentially* in holding a belief when and only when the believer does not need to base the belief on any actual inference in order for that believer to be justified in holding that belief. Similarly for knowledge. When a believer is justified in holding a belief, that belief state is also said to be justified. The central claim of moral intuitionism is then that some moral intuitions do not need any actual inferential basis in order to be justified or known (25, 41).

Moral intuitionists usually also claim that certain propositions are self-evident. A proposition is self-evident if anybody who adequately understands it is justified in believing it and knows it if the belief is based on that understanding (48–49). Self-evidence is an impersonal property of propositions in the abstract, so a proposition can be self-evident even if no believer is justified in believing it (because no believer

adequately understands it). Since I am more interested in the personal question of when believers are justified in holding their moral beliefs, I will focus on the claim that some moral believers and beliefs are justified noninferentially.

This central claim is shared by all moral intuitionists, but they differ in their answers to the question of what makes such moral intuitions justified. *Reliabilists* claim that some moral beliefs are justified merely because they result from a reliable process (57–59; cf. Shafer-Landau 2003). *Experientialists* claim that some moral beliefs are justified because of experiences, appearances, or seemings that are not beliefs and, hence, cannot serve as premises in inferences (55–57; cf. Tolhurst 1990, 1998). *Contextualists* claim that the social contexts of believers make some beliefs justified independent of any inferential support (59; cf. Timmons 1999).

Reflection

In place of reliability, seeming, and context, Audi claims that reflection is what makes moral intuitions justified. Audi's basic idea is that some moral believers are justified because their moral beliefs are based on reflection of a certain kind. This reflection involves beliefs, but the resulting belief is not based on an inference from those other beliefs, so the resulting belief is still justified noninferentially.

Audi's approach depends crucially on his distinction between conclusions of reflection and conclusions of inference. Here is his main example:

> Consider reading a poem with a view to deciding whether its language is artificial. After two readings, one silent and one aloud, we might judge that the language is indeed artificial. This judgment could be a response to evidential propositions that occur to one, say that the author has manipulated words to make the lines scan. But the judgment need not so arise. If the artificiality is subtler, there may just be a stilted quality in the poem. In this second case, one judges from a global, intuitive sense of the integration of vocabulary, movement, and content. Call the first judgment of artificiality a *conclusion of inference*: it is premised on propositions noted as evidence. Call the second judgment a *conclusion of reflection*: it emerges from thinking about the poem as a whole, but not from one or more evidential premises. (45)

Audi's second example concerns appraising whether a letter of recommendation is strong (46; see also Audi 1996, 112). Although these examples do not involve any moral conclusion, Audi clearly believes that some moral judgments are conclusions of reflection in the same way as the nonmoral conclusions in these examples.[2]

This form of reflection is very different from intuitive induction, which rises from concrete instances to general conclusions (20, 63, 150, etc.). In Audi's examples, the conclusion of reflection is not general but particular: *this* language is artificial, and *this* letter is not strong. Such particular reflection is more fundamental, because intuitive induction begins with a particular judgment (which might be a conclusion of reflection) before it rises to its general conclusion. Thus, if conclusions of reflection are not justified, neither are the general conclusions reached from them by means of intuitive induction.

So, how are particular conclusions of reflection justified? As Audi says, "There is a tendency to think of intuitions as *focally grounded*: as based simply on a grasp

of the proposition taken in abstraction from one's grounds for it" (46). In contrast, "the literary and testimonial examples illustrate a sense in which intuitions may be *globally grounded*: based on an understanding of the proposition seen in the context of the overall grounds for it" (46). Audi suggests that focally grounded intuitions are not conclusions of reflection, when he makes a point that is "easily overlooked if one is thinking only of focally grounded intuitions and, as is easy in that case, neglecting the rich grounds that may be provided by the kind of reflection that underlies intuitions which constitute conclusions of reflection" (48). Thus, conclusions of reflection are only one kind of moral intuition.

Audi suggests that yet another kind of moral intuition is grounded in emotion: "Emotions may reveal what is right or wrong before judgment articulates it; they may both support ethical judgment and spur moral conduct" (87; cf. 57 and McCann in this volume). A moral belief that is based on such emotional reactions alone does not seem to be a conclusion of reflection, but it does seem to count as a moral intuition, because it is not "believed on the basis of a premise" (33).

Focally and emotionally grounded intuitions are enough to show that "by no means all moral intuitions . . . are conclusions of reflection" (47). Any moral intuition that is not a conclusion of reflection will be called an *unreflective* moral intuition.

Some unreflective moral intuitions are not justified. Many people form moral intuitions quickly after no reflection at all. As Audi admits, "An intuition can be mistaken, and a mere prejudice can masquerade as an intuition" (66). Such passages suggest that Audi would not assess *all* unreflective moral intuitions as justified.

Are *any* unreflective moral intuitions justified? Audi never answers this question. He implies that some focally and emotionally grounded moral intuitions are accurate, but he never explicitly calls them justified (unless I missed something). Audi's commitment to self-evidence of some moral propositions might seem to suggest that believers can be justified merely by understanding, but it is still not clear whether reflection of some kind is necessary for adequate understanding. If so, he might hold that beliefs in self-evident propositions are not justified without reflection.

The questions, then, are these: Are any unreflective moral intuitions justified? If so, which are justified and which are not justified? If no unreflective moral intuitions are justified, then are all justified moral intuitions conclusions of reflection like those in Audi's examples? Is reflection necessary for any moral intuition to be justified?

Inference

Suppose Audi answers that all justified moral intuitions are conclusions of reflection. Then, if conclusions of reflection are based on inference, no moral intuition is justified noninferentially. Next suppose Audi answers that some moral intuitions are justified without being conclusions of reflection. Then I will want to ask what makes those other moral intuitions justified and how they are related to inference. For now, however, I will focus solely on moral conclusions of reflection. If none of them is justified noninferentially, then Audi cannot defend the central claim of

moral intuitionism without telling a different story about how other moral intuitions can be justified noninferentially.

Either way, it is crucial to ask whether conclusions of reflection are based on inference. Audi admits that inferences occur as intermediate steps in reflection (51–52), but the conclusions of those inferences are not the ultimate conclusion of the reflection. The real question, then, is whether there is any inference from the beliefs that occur during reflection to the conclusion of that reflection.

To answer this question, we need to determine which beliefs are formed during reflection. In Audi's poetic example, "one judges from a global, intuitive sense of the integration of vocabulary, movement, and content" (45). This "sense" seems to be articulate, since the reflecter is reading and reflecting on language; so the reflecter seems to have beliefs about the vocabulary, movement, content, and their integration. If the reflecter does not have any beliefs at all about the vocabulary (that is, which words are in the poem) or the content (that is, what the poem is about) or the movement (which comes out when the poem is read aloud), then the process would not deserve the name "reflection." At the very least, the reflecter needs beliefs about which words are in the poem, or else he could not know which poem he is reflecting on. Thus, whatever they are, some beliefs are formed during reflection and are needed for that reflection. Let's call all of them together "the beliefs of reflection."

At the end of reflection, the reflecter also forms another belief: that the poem's language is artificial. This is the "conclusion of reflection." It must be connected both causally and by virtue of its content to the beliefs of reflection in order to count as a conclusion of that reflection. A random thought that happens to occur during reflection, without such connections to beliefs of reflection, would not count as a conclusion of that reflection. Moreover, the reflecter sees beliefs of reflection as grounds for the conclusion of reflection. The point of reflection is to use the beliefs of reflection as a means to reach and support the conclusion of reflection. Finally, the value of reflection also depends on connections between the beliefs of reflection and the conclusion of reflection. If the conclusion of reflection were not connected causally and by its content to the beliefs of reflection, then reflection involving those beliefs could not make its conclusion justified. Thus, the reflecter in Audi's poetic example seems to form certain beliefs (of reflection) that cause and make the reflecter justified in holding another belief in a conclusion (of reflection).

Isn't that an inference? It fits Audi's own definition of inferential grounding:

> To say that a cognition, such as a judgment or a belief of a moral principle, is *inferentially grounded* in another cognition is roughly to say that the first is *held* on the basis of the second (or cannot be properly held by the person in question apart from such an inferential connection). (141)

Although qualified by "roughly," this definition seems to imply that conclusions of reflection are inferentially grounded in beliefs of reflection.[3] The reader's belief that the poem's language is artificial is held on the basis of other beliefs about the poem's words, movement, and so on, and the conclusion cannot be properly held by the reader apart from being based on those other beliefs. Thus, conclusions of reflection are inferentially grounded according to Audi's own definition.

In his main work on inference, Audi refers to "*inference*, understood generically as a process of passing from one or more premises to a conclusion" (1993, 237). This account does not help here, because reflecters pass from beliefs of reflection to the conclusion of reflection, so the question of whether reflection involves inference comes down to the question of whether beliefs of reflection are "premises." That issue cannot be settled independently of determining whether they are part of an inference.

Other definitions, such as those by Sturgeon (2002), also count Audi's conclusions of reflection as conclusions of a kind of inference. Audi replies, "His view of inference is wider than mine" (210n12). It is still not clear, however, exactly what Audi's narrow notion is or why we should prefer it to Sturgeon's wider notion of inference. We need to know precisely which feature is necessary for an inference but missing from the cognitive transition from beliefs of reflection to conclusions of reflection.

What's missing might seem to be a certain form or structure. However, inferences come in many forms. Some are enthymematic (that is, they depend on suppressed premises). Audi also recognizes "derivations" (102, 110), "proofs" (139, 219n8), and inferences that are not deductive. The conclusions Audi has in mind do not result from inductive generalization or from inductive application of a generalization, but there are many other kinds of inductive inferences. (See Fogelin and Sinnott-Armstrong 2005, chap. 9.) Legal arguments in courts or judicial opinions, for example, are often neither deductive nor generalizations, nor applications of generalizations, yet they are still inferences (cf. 219n8). After all, the lawyers or judges say, "Therefore, . . . " or some equivalent, before they reach their conclusions. Such terms mark an inference. Thus, the failure of such a cognitive transition to fit into any standard form cannot keep it from being an inference.

This failure still might seem to keep it from being a *good* inference. Michael DePaul (in conversation) suggested that the transition to a conclusion of reflection should not be seen as an inference because it would be such a bad inference. However, bad inferences are still inferences. Besides, Audi claims that the conclusion of reflection is justified by the reflection, so it is not clear that Audi would agree that this transition would be a bad inference, if it were an inference.

In any case, I think Audi would and should agree that the transition from the beliefs of reflection to the conclusion of reflection has the form or structure of an inference. He writes, "Granted, if I *articulate* my non-inferential grounds, then they will be available to me as premises" (1996, 113; cf. 1993, 238). Again, "there is an appropriate accessible path leading (perhaps by natural inferential steps) from justificatory materials accessible to us to an occurrent justification for the proposition" (50–51). Thus, Audi's distinction between conclusions of reflection and conclusions of inference can't be a matter of form.

The distinction might, instead, rest on the reflecter's consciousness or articulation of the inferential structure or its elements. Audi suggests this when he says that a structurally inferential conclusion "need not arise from, or be sustained by, any tokening, e.g., internal recitation, of that structure which deserves the name 'inferring p from r' " (1993, 238). If "internal recitation" is necessary for a conclusion to be episodically inferential, and if no such recitation of an inferential structure occurs with a conclusion of reflection, but it does occur with all conclusions of

inference, then we have found the difference. However, the distinction can't be that simple. Audi gives "internal recitation" merely as an example ("e.g."), so he does not require "internal recitation" for inference. He seems to allow other kinds of "tokening." If the inferential structure and its elements can be tokened in some unconscious or less than fully articulate way, then consciousness or articulation cannot be what divides conclusions of reflection from conclusions of inference. Moreover, there is no reason to restrict inferences to fully articulated, conscious recitations, as Audi sometimes admits (e.g., 2001, 33 and 48). Coherentists and others who deny that beliefs are noninferentially justified do not require that every supporting inference be fully articulated or that believers recite the inference step by step. And if conclusions of reflection depend on inferences that are unconscious or unrecited, then they cannot avoid the skeptical regress that motivates moral intuitionism. Thus, Audi's distinction cannot do the work asked of it, if its main point is just about whether the inferential structure is conscious or articulated.

Another possible basis for the distinction lies in whether or how the reflecter *uses* that inferential structure. Audi seems to admit that reflecters would cite that structure if they were asked the right questions: "for every case of believing *p* for reason *r*, . . . there is an argument whose premise(s) indicate the structure of the belief(s) causally grounding the belief that *p*, and is such that S is disposed in effect to appeal to the argument should he try to explain or justify his believing that *p*" (1993, 238).[4] However, as he emphasizes, "surely my having a ground that is *expressible* in a premise does not imply that I must *use* that ground *in* a premise in order to form a belief on the basis of that ground" (1996, 112). This and other passages suggest that it is something about how the beliefs of reflection are used that keeps them from being premises in inferences.

But what is the difference in use? The idea seems to be that beliefs of reflection are used *only* to help the reflecter *understand* the conclusion of reflection. If the believed proposition is self-evident in the sense that anyone who adequately understands it is justified in believing it (48), then reflection makes the believer justified simply by providing understanding.

However, if this were the intended story, then Audi's poetic and testimonial examples would be misleading. The beliefs that certain wording is artificial and that a certain recommendation is not strong are surely not self-evident. Moreover, the beliefs of reflection in those cases do not merely help the reflecter understand what the conclusion of reflection means. The reflecter understood artificiality and weakness in recommendations prior to and independently of reflection on these particular cases. Reflection on these cases might help the reflecter understand something else—*why* the wording is artificial and *why* the recommendation is weak. But that kind of understanding works only because of how the beliefs of reflection fit into an inferential structure that provides the reasons why. The inferential structure is then not a mere accompaniment but is essential to this use of the reflection. Besides, even if reflection does help us understand why the conclusion is true, the question of why it is true assumes that it is true, so it would beg the question to use reflection aimed at that why-question in order to justify belief in the truth of the conclusion of reflection. I doubt that this is what Audi meant, since it would not serve his purposes.

This all leaves me wondering why Audi denies that the reflecter infers the conclusion of reflection from the beliefs of reflection. This cognitive transition might be enthymematic, inductive, or both, but why isn't it still an inference? Of course, whether this cognitive transition is or is not an inference depends on what an inference is or on how inference is defined. So what we really need from Audi is a definition of inference that brings out the crucial difference between conclusions of reflection and conclusions of inference.

Second-Order Beliefs and Inferences

So far I have focused on first-order inferences from some beliefs about the topic of reflection to a conclusion about that same topic. Even if reflection does not involve inference at that level, it still might involve another kind of inference at a higher level: a second-order inference from beliefs about the reflection to the conclusion of reflection. The current fashion in epistemology downplays second-order beliefs, but they still might play a role in the special process of reflection.

The need for second-order beliefs arises because not just any reflection will do. A reflecter might not have enough facts, or might not reflect long enough, or might be biased or distracted. Then the reflection is inadequate to justify its conclusion, as Audi recognizes. To make a moral belief justified, reflection must have certain properties: it must be careful and long enough, it must not be distorted by ignorance or self-interest, and so on. Beliefs like these—about when reflection is adequate— are surely common. The question is whether they are necessary.

Audi implicitly suggests the need for such second-order beliefs when he describes his poetic example (although his other examples would work as well). Audi mentions that the reflecter makes "two readings, one silent and one aloud" (45). Why did the reflecter decide to read it aloud after having just read it silently? Why does he bother with a second reading? If he just happens to find himself doing this for no reason, this random process hardly warrants the name "reflection." The most likely explanation is instead that the reflecter at least implicitly believes that one reading is *not* enough. But then why does he stop with two readings? Maybe he just got tired or ran out of time, but again that seems too arbitrary for serious reflection that justifies belief. The most likely explanation is instead that the reflecter implicitly believes that two readings *are* enough. For that reason, this reflecter seems to believe that reflection is adequate on issues like this when it is based on two readings, one silent and one aloud.

Similarly, Audi says, "one judges from a global, intuitive sense of the integration of vocabulary, movement, and content" (45). Why does the reflecter attend to *these* factors? The most likely explanation is that the reflecter believes that these factors are all relevant, so reflection on only some of them would not be adequate. But then why doesn't the reflecter go on to consider *other* factors, such as the author's intentions, poetic precursors and contemporaries, possible accents in which the poem might be read, and so on. The most likely explanation is that the reflecter believes that reflection on the original factors (the integration of vocabulary, movement, and content) is enough. In short, the reflecter needs at least some implicit

beliefs about when reflection is adequate in order to know when to stop reflecting and draw a conclusion.

Of course, not every reflecter needs to formulate such beliefs fully, explicitly, or self-consciously. However, every reflecter does need to decide what to reflect on and when to stop reflecting. Such decisions and actions are intentional, so they seem to result from beliefs (as well as desires), at least when they are rational. That makes it natural to ascribe second-order beliefs to reflecters, even when they are not fully self-conscious about what they are doing.

These second-order beliefs then seem to serve as premises in second-order inferences. A second-order belief about the adequacy of the reflection causes the reflecter to stop reflecting and to draw the conclusion of reflection. And if the conclusion of reflection were not so causally connected to this second-order belief, then the reflecter would have stopped without believing that the reflection is adequate, in which case this reflection could not make the conclusion justified (as I will argue shortly). Thus, the conclusion of reflection "is *held* on the basis of" the second-order belief about adequacy and "cannot be properly held by the person in question apart from" that belief; so the cognitive transition from that second-order belief to the conclusion of reflection fits Audi's own definition of inferential grounding (141).

As with first-order inferences, Audi might respond that he uses a narrower notion of inference. But then, as before, we need to know what that notion is and why we should prefer it to broader notions that include inferences from second-order beliefs like those that operate implicitly in drawing conclusions of reflection.

A more likely response by Audi is that the reflecter need not form even an implicit belief that the reflection is adequate in order to be justified in believing the conclusion of reflection. However, there seem to be only three ways to avoid an implicit second-order belief in adequacy: either the reflecter does not consider whether his reflection is adequate or the reflecter considers it but does not form any belief about whether his reflection is adequate or he considers it and forms the belief that his reflection is inadequate. Let's consider these possibilities in reverse order.

First, if the reflecter believes that his reflection is inadequate, it is hard to see why that reflecter would be justified in believing the conclusion of that reflection. Just imagine someone who says, "I did not reflect long enough, and I reflected on the wrong kinds of examples, and I am partial and ignorant. All of these factors distort my reflection and make it unreliable. Nonetheless, my reflection still makes me believe that sodomy is immoral." This person does not seem justified.

Second, what if the reflecter considers the adequacy of his reflection, but does not arrive at any considered judgment about its adequacy? Again he does not seem justified. Imagine someone who says, "Maybe I didn't reflect long enough or on the right kinds of examples. I might be partial or misinformed. I considered these sources of distortion, but I still have no views on whether my reflection is distorted or adequate. Nonetheless, my reflection makes me believe that sodomy is immoral." This reflecter also does not seem justified in holding this belief. To trust reflection in such circumstances is like trusting a thermometer that one started to test but did not finish testing, when one has no reason to trust it. That hardly seems justified.

The third possibility is that the reflecter has not even considered whether his reflection is adequate. This reflecter is not very reflective. This process is no ideal

to pursue. When risks are high and disagreements and distortions are common, as in moral beliefs (see Sinnott-Armstrong 2005), a reflecter needs to think about whether he should trust the source of his belief. Not to do so is irresponsible. Thus, if a moral reflecter does not think at all about whether his reflection is trustworthy, then he is not justified in trusting it or its conclusions.

Overall, then, if a moral reflecter does not believe that he has reflected adequately, he is not justified in believing the conclusion of his reflection. To be justified on the basis of reflection, he must believe that he has reflected adequately. This requirement might seem to lead quickly to skepticism. However, coherentists can accept this requirement without becoming skeptics, so it does not lead directly to skepticism. Moreover, my point about moral intuitionism works even if this requirement applies only to moral reflection because of special problems for morality (such as those mentioned in the preceding paragraph), and then it does not lead to general skepticism. Finally, if the result is moral skepticism, so be it. Some moderate forms of moral skepticism are defensible (Sinnott-Armstrong 2006). It would beg the question to assume otherwise here, since part of the point of moral intuitionism is to respond to moral skepticism, not just to assume that it is false.

What does follow is that conclusions of moral reflection cannot be justified apart from inference. If the moral reflecter must believe that he has reflected adequately, then the conclusion of reflection "cannot be properly held by" the reflecter apart from that second-order belief. And if that second-order belief were unconnected to the reflection, then it could hardly make the reflecter justified in believing the conclusion of that reflection. Thus, the cognitive transition from that second-order belief to the conclusion of reflection seems to count as an inference by Audi's own definition (141). As always, Audi could invoke a different notion of inference, but then we need to see what that new notion is.

An Ability to Infer

My claims so far, especially in the preceding section, are likely to be controversial. Still, even if some moral conclusions of reflection are justified apart from any *actual* inference—either from beliefs of reflection or from a second-order belief about the adequacy of reflection—a more basic problem arises: justified conclusions of reflection still might depend on an *ability* to infer them.

Audi himself admits that reflecters are *able* to infer conclusions of reflection: "I grant that in principle, where one arrives at a conclusion of reflection, one *could* figure out why and *then* formulate, in explicit premises, one's basis for so concluding" (47; cf. 64 "in a position to"). What we could do, we are able to do; so, if reflecters *could* formulate explicit premises and *could* reach their conclusions of reflection on the basis of those premises, then they are *able* in the relevant sense to infer those conclusions. This admission is repeated in earlier work where Audi suggests that reflection involves "a ground that is *expressible* in a premise" (1996, 113). We are able to express what is express*ible*, and if reflecters are able to express their beliefs of reflection as premises and are able to base their conclusions of reflection on those premises, then reflecters are able in the relevant way to infer their conclusions. Again, Audi

admits that when we believe a conclusion of reflection, we have what he calls structural justification for that belief: "there is an appropriate accessible path leading (perhaps by natural inferential steps) from justificatory materials accessible to us to an occurrent justification for the proposition" (50–51; cf. 1993, 274–96).[5] If the materials and the inferential path are both access*ible* to reflecters, then reflecters are *able* in the relevant sense to infer their conclusions of reflection. Thus, Audi seems to grant (as he should) that some weak kind of ability to infer the conclusion is necessary every time a reflecter is justified in reaching a conclusion of reflection.

Audi thinks that this concession is minor, because "the point that a ground of intuitive judgment *can* be formulated through articulation of one's basis for judgment does not entail that the ground must do its justificatory work in an inferential way" (48; cf. 1996, 113). Granted: an ability to infer a belief need not actually be exercised or actually be used to ground that belief, even if the ability is necessary for the belief to be justified.

However, this concession is more important than Audi realizes. One of the main motivations for moral intuitionism as a view in moral epistemology is to stop the skeptical regress. Audi shares this goal (44). However, moral intuitionism cannot achieve that goal if an ability to infer a moral belief is necessary for that moral belief to be justified. The reason is that a skeptical regress arises from the need for an ability to infer.

Suppose that no believer is justified in believing any moral claim unless that believer is able to draw an inference from other beliefs to that moral belief as a conclusion. Then the ability to draw an inference is necessary for a believer to be justified. Still, the mere ability to draw an inference cannot be enough to fulfill the necessary conditions of justified belief if the believer is not justified in believing the premises of that inference. The reason is that I could be justified in believing *anything* if the ability to draw an inference from unjustified beliefs were enough for my belief to be justified. After all, any fool belief can be inferred from itself. Consequently, a mere ability to infer a belief from other beliefs is not enough to provide what is necessary for a believer to be justified in believing a conclusion unless the believer is also justified in believing the premises of that potential inference. But then the believer needs to be able to infer the premises in order to be justified in believing the original conclusion. And so on. This requirement is enough to start the very skeptical regress that moral intuitionists were trying to avoid. Thus, Audi cannot avoid the skeptical regress, so he cannot achieve his goal, if he claims only that believers can be justified independently of actual inference.

To stop the skeptical regress, Audi and other moral intuitionists need to make the stronger claim that a belief can be justified independently of any actual inference and also independently of any ability to draw any inference. This claim would be enough to stop the regress, because it denies any dependence on any inference with any new premise that needs to be justified.

However, if *reflection* is what makes a belief justified, then that belief is not justified independently of an *ability* to infer it. Audi conceded as much in the quotations five paragraphs above. That concession is unavoidable, since a reflecter who makes the cognitive transition from beliefs of reflection to a conclusion of that reflection must at least be able in a weak sense to draw an inference from those beliefs of reflection to that conclusion of reflection. Thus, even if conclusions of

reflection were justified independently of any actual inference, they still would not be justified independently of an ability to infer them. Hence, they would not provide an adequate solution to the skeptical regress problem, as Audi and other moral intuitionists want.

Questions

In sum, I want to ask Audi these questions:

(a) Are any unreflective moral intuitions justified? (If so, what makes them justified? If not, is reflection necessary for justified moral intuition?)

(b) Why doesn't the cognitive transition from the beliefs of reflection to the conclusion of reflection count as an inference? Exactly which essential feature of inference is missing?

(c) Can a conclusion of reflection be justified without even an implicit second-order belief that the reflection is reliable? If so, why is it justified? If not, why isn't the cognitive transition from such second-order beliefs to the conclusion of reflection another inference?

(d) Can a conclusion of reflection be justified without even an ability to infer it? If not, can the skeptical regress really be stopped by the weak form of moral intuitionism that claims independence of actual inference but not of ability to infer? If so, how? If not, what is left of the motivation for moral intuitionism as a position in moral epistemology?

I have no doubt that Audi can answer these questions in illuminating ways. I look forward to his answers.[6]

Notes

1. All references are to this work unless otherwise indicated.

2. Unfortunately, I could not find any place where Audi explicitly labels any moral judgment a conclusion of reflection. He does give a moral example when he discusses conclusions of reflection: "The sight of a soldier in an occupying army violently slapping an old man may serve as a paradigm by which the wrongness of injury is memorably seen" (46). However, this strikes me as intuitive induction because it rises to a general conclusion about the wrongness of injury. That passage focuses on the "contrast between abstract and concrete intuitions" (46), and Audi does not call either intuition in this example a conclusion of reflection.

3. It is not clear how to read the parenthetical disjunction, but Audi's definition at least implies that a cognition is inferentially grounded when *both* disjuncts are true. To be safe and fair, whenever I claim that a cognitive transition is an inference by Audi's definition, I will argue that both conditions are met.

4. Although it is not completely clear, I assume that reflection is one instance of the kind of "believing for a reason" that Audi is discussing in this quotation.

5. The parentheses make this quotation ambiguous, but I read Audi as saying that the inferential path is always available, even though its steps are only "perhaps" natural.

References

Audi, Robert. 1993. *The Structure of Justification*. Cambridge: Cambridge University Press.
——. 1996. Intuitionism, Pluralism, and the Foundations of Ethics. In *Moral Knowledge?*, ed. W. Sinnott-Armstrong and M. Timmons. New York: Oxford University Press.
——. 2001. *The Architecture of Reason*. New York: Oxford University Press.
——. 2004. *The Good in the Right: A Theory of Intuition and Intrinsic Value*. Princeton, N.J.: Princeton University Press.
Fogelin, Robert, and Walter Sinnott-Armstrong. 2005. *Understanding Arguments*, 7th ed. Belmont, Calif.: Thomson/Wadsworth.
Shafer-Landau, Russ. 2003. *Moral Realism: A Defense*. New York: Oxford University Press.
Sinnott-Armstrong, Walter. 2005. Moral Intuitionism Meets Empirical Psychology. In *Metaethics after Moore*, ed. M. Timmons and T. Horgan. New York: Oxford University Press.
——. 2006. *Moral Skepticisms*. New York: Oxford University Press.
Sturgeon, Nicholas. 2002. Ethical Intuitionism and Ethical Naturalism. In *Ethical Intuitionism: Re-evaluations*, ed. P. Stratton-Lake. Oxford: Clarendon.
Timmons, Mark. 1999. *Morality without Foundations*. Oxford: Oxford University Press.
Tolhurst, William. 1990. On the Epistemic Value of Moral Experience. *Southern Journal of Philosophy*. Suppl. no. 29: 67–87.
——. 1998. Seemings. *American Philosophical Quarterly* 35: 293–302.

Intuitionism and Disagreement

ROGER CRISP

Robert Audi's magisterial *The Good in the Right* offers the most comprehensive and developed account of rational ethical intuitionism to date.[1] Audi himself is attracted to a Rossian version of the position in which Ross's substantive principles are integrated within and systematized by a more general Kantian theory (chaps. 2–3). He finds three elements in Ross's view "appropriate to the metaethics of any full-blooded version of intuitionism: the claim of irreducible plurality for basic moral principles; the association of each principle with a different kind of duty; and the thesis that each principle is in some sense intuitively knowable by those who appropriately understand it" (40).

For the sake of clarity, I shall concentrate on the third element only, restricting the notion of intuitionism to epistemological views according to which normative principles—that is, principles stating reasons for actions—are self-evident. My main claim is that, although this epistemology is indeed plausible, the implications of continuing and deep-seated normative disagreement are such that the prospects for supporting any specific first-order view—such as Ross's—are considerably bleaker than Audi takes them to be.

By an intuition, Audi means something like a belief—indeed perhaps a kind of belief. He identifies four salient characteristics of intuitions (32–36). First, the *noninferentiality requirement*: an intuition is noninferential, in the sense that it is not based on inference from any premise or set of premises. Second, the *firmness requirement*: intuitions must be accompanied by a reasonably strong sense that the proposition in question holds. Third, the *comprehension requirement*: intuitions must be held on the basis of an adequate understanding of their content, and in some cases this may demand a certain degree of reflection. Finally, the *pretheoreticality requirement*: intuitions, as intuitions, are neither dependent on evidence provided by theories nor themselves held as theoretical hypotheses.[2]

Certain moral principles, Audi believes, are self-evident. They may be comprehensible by a rational individual who does not in fact accept them; but they can be accepted noninferentially, on the basis of an appropriate understanding of them.

And a rational individual whose acceptance of them in this way meets both the firmness and the comprehension requirements can be said to know them, and to be justified in accepting them (43–44).

This justification, however, is not indefeasible (45). That is to say, if I intuitively accept some moral principle, and my intuition meets all of Audi's requirements, I am justified in accepting this principle, and if it is true then I know it. But it may not be true. So such principles are not entirely certain, and I cannot be entirely certain about them. And now the question arises what might be—to quote Henry Sidgwick—the "conditions, the complete fulfilment of which would establish a significant proposition, apparently self-evident, in the highest degree of certainty attainable."[3]

Sidgwick himself suggests four conditions.[4] The first two might be understood to be included within Audi's comprehension requirement: the terms of the proposition must be clear and precise, and acceptance of the proposition must be grounded in careful reflection. The third is uncontroversial: no individual is entitled to claim genuine certainty or self-evidence for any two intuitions that are seen by that individual as inconsistent with each other. This condition would of course be accepted by Audi, and might again be covered by an appropriate extension of the comprehension requirement from the content of propositions to the logic that governs the assertion of them. Someone who avows two propositions she herself believes to be inconsistent, even if she manifests apparent understanding of the content of each of the two propositions in question, misconceives the logic of assertion to the point that one must question her overall capacity for understanding the import of what she is saying.

Sidgwick's fourth condition—which one might call the *consensus condition*—concerns interpersonal consistency, and is worth quoting:

> Since it is implied in the very notion of Truth that it is essentially the same for all minds, the denial by another of a proposition that I have affirmed has a tendency to impair my confidence in its validity. . . . And it will easily be seen that the absence of . . . disagreement must remain an indispensable negative condition of the certainty of our beliefs. For if I find any of my judgments . . . in direct conflict with a judgment of some other mind, there must be error somewhere: and if I have no more reason to suspect error in the other mind than in my own, reflective comparison between the two judgments necessarily reduces me temporarily to a state of neutrality. And though the total result in my mind is not exactly suspense of judgment, but an alternation and conflict between positive affirmation by one act of thought and the neutrality that is the result of another, it is obviously something very different from scientific certitude.[5]

Something like the consensus condition might also be included within the comprehension requirement, embedded in the account of reflection (see 35). In discussing the pretheoreticality requirement, Audi suggests that intuitionism need not require that biases be entirely avoided: "It is enough if . . . they are always correctable by further reflection. Such reflection may include comparison with the intuitions of *others*, just as in scientific enquiry one might compare one's observations with those of co-investigators" (37).

Proper reflection, then, might consist partly in one's assessing one's own apparently self-evident ethical beliefs against those of others: "Given how intuitions are understood—as deriving from the exercise of reason and as having evidential

weight—conscientious intuitionists will try to factor into their moral thinking, especially on controversial issues, the apparent intuitions of *others*" (47; see also 66–67).

But what exactly is the consensus condition, and what are its implications? Consider the situation of an ethical thinker, R, who has a normative intuition that *p*, which meets all of Audi's requirements and the first three of Sidgwick's conditions. She accepts *p* not, or not only, on the basis of any inference; she has a strong conviction that *p*; she fully understands *p*; she does not accept *p* on the basis of any theory, nor is she postulating it as part of any theory; the terms of *p* are clear and precise; she holds *p* on the basis of serious, unbiased, and mature reflection on *p* and its implications; and *p* is consistent with all her other beliefs. But now R comes across another thinker, S, who asserts *not-p*. Immediately, according to the consensus condition, this should impair her confidence in *p*. But she may regain her confidence in proportion to the degree to which S fails to meet certain of the various conditions and requirements just mentioned. If S is tentative, demonstrates a lack of understanding of *not-p*, or has not adequately reflected upon the proposition, R's confidence in *p* may begin to return. But unless S's epistemic state is significantly faulty in one or other of these ways, R may well feel that her trust in *p* has been dented to the point that she could not claim to hold it "in the highest degree of certainty attainable."

Now imagine that it becomes clear to R that she and S meet the various requirements and conditions roughly to the same degree, so far as that is ascertainable.[6] At this point, according to Sidgwick, her confidence in *p* must disappear entirely and she must be neutral between *p* and *not-p*. Sidgwick says that "this is not exactly suspense of judgment," since R may well continue to affirm *p* "by one act of thought" while remaining neutral in another. But this sounds like a description of a kind of epistemological schizophrenia in which the contradictory affirmations are somehow kept apart. In fact, I suggest, what should emerge, on a proper understanding of the consensus condition, is indeed suspension of judgment, alongside an affirmation of a mere appearance. That is to say, if we ask R about her judgments regarding *p* and *not-p*, she will, if she understands the consensus condition, be neutral between them, but admit that it still *seems* to her that *p* is true.[7] If two people who believe themselves to be in roughly the same epistemic circumstances are confronted by what one of them takes to be a redwing, the other a song thrush, they should suspend judgment on which of the two the bird is. But the bird's appearance to each as one or the other may not change.

On this account, whether R is justified in her basic normative beliefs depends on the views of others, and how well grounded those views turn out to be. Here someone hostile even to this degree of contextualism in justification might object that epistemic entitlement cannot be so contingent. In fact, R does not need to find out whether others also accept *p*. Rather, she has merely to ask herself whether it is possible that another person, in as epistemically good a state as she is, might reasonably hold *not-p*. If it is possible, then she should suspend judgment on *p*.

This objection rests on a dim view of human epistemic capacities. A philosophical intuitionist is likely to put some faith in the notion that our intuitive faculty is to some extent truth-tracking. This claim might be supported by reference either to consensus on certain nonethical truths, such as those of basic arithmetic or logic, or to consensus on apparent normative truths, such as the following:[8]

> *Pain*: Non-deserved suffering of any agent A that would be caused in or by some
> action of A counts (though not always decisively), for A, against the performance of
> that action by A.

It is plausible to think that the reason that nearly all human beings accept *Pain*
or something like it is that *Pain* is true and that our capacity for detecting such
truths explains our convergence upon it.[9] So contingent consensus on some nor-
mative principle provides, in the appropriate circumstances, some justification for
certainty about such principles, although this is not, of course, to claim indefeasi-
bility for them (see 31–32 and *passim*). Likewise, contingent dissensus among epis-
temically similarly situated thinkers not only precludes certainty but also requires
suspension of judgment. It is true, however, that numbers count.[10] To return to our
example, if R finds that S alone accepts *not-p*, and that T, U, V, and others agree
with her about *p*, then her judgment need no longer be suspended, and once the
numbers on her side are large enough, *p* may plausibly be said to be established "in
[almost] the highest degree of certainty attainable."

The problem for Audi, and for any proponent of an intuitionist epistemology in
ethics who wishes to advocate some first-order principles, is that the degree of con-
sensus in philosophical ethics is very small, and Sidgwick's fourth condition soon
enters to require suspension of judgment. One might expect near consensus on
something like the following:

> *Pain 2*: Non-deserved suffering of any sentient being that would be caused in or by
> some action of any agent A counts (though not always decisively), for A, against the
> performance of that action by A.

Pain 2 will be disputed by philosophical egoists. But in fact *Pain 2* anyway does
not capture an ethical principle with the sort of generality and explanatory force
that we seek in ethics.[11] What we shall ask anyone who advocates *Pain 2* is what prin-
ciple lies behind it, and there will then be radical disagreement between Kantians,
who might put weight on notions such as dignity and respect, utilitarians, who
emphasize the rationality of maximizing well-being, Rossian pluralists,[12] virtue eth-
icists, feminist ethicists, social contract theorists, and so on.[13] Numbers are again rel-
evant, especially when making certain comparisons between positions. A utilitarian
might be justified in thinking that it is more likely—though perhaps only slightly
more likely—that utilitarianism is true than that a somewhat esoteric trust-based
ethics is true. But when it comes to whether utilitarianism or, say, a Rossian view is
correct, suspension of judgment will be called for.

It might be said that usually, when confronted by someone with a different eth-
ical viewpoint from ourselves, we are inclined to think that he or she has failed to
see something that we have appreciated. But here we should remember that he or
she will think the same about us, and the question is whether either of us is justi-
fied in thinking that the other is in a worse epistemic position. I suggest that such
justification in ethics is really quite rare, and that to this extent philosophical ethics
is characterized by an unjustified dogmatism. Nor should it be thought that the
consensus condition does not apply to dogmatists, who hold without justification
that they are in a better epistemic position than others. Sidgwick may state it in the

form of autobiographical description: "if I have no more reason to suspect error in the other mind than in my own, reflective comparison between the two judgments necessarily reduces me temporarily to a state of neutrality." But he clearly means his claim to imply that someone who is not so reduced is in error.

A further question worth considering is whether the consensus condition applies only within an intuitionistic epistemology. In fact, the condition anyway does not apply within all such epistemologies. Consider an intuitionist who is a reliabilist. It may be that because you are in touch with the moral truth you have knowledge of your moral principles and I do not, though neither of us has any better evidence that the other is in a worse epistemic position. And it does seem that there will be other epistemologies in which the consensus condition would be out of place, such as a version of relativism about knowledge in which two people whose beliefs are contradictory can both be said to possess knowledge. Or we might imagine a coherentist who puts all the weight on intrapersonal coherence of beliefs, so that consensus becomes irrelevant. Nevertheless, it does seem that the consensus condition, partly because it is so commonsensical, is likely to apply within many nonintuitionist epistemologies. That is not, however, to play down its importance for intuitionists. If anything, it highlights that importance.

Audi himself (60–63) does discuss the problem of ethical disagreement in the context of what he calls the *dissensus objection*: If the basic principles of ethics are self-evident, why is there so much disagreement about them? First, Audi claims that, since ethical intuitionists need only 'soft' self-evidence, we should not expect a high degree of consensus even after discussion. Soft self-evidence is that which lacks the properties of hard: strong axiomaticity, immediacy, indefeasibility, and compellingness (53). Ross's theory, for example, is quite hard to understand, and since it postulates fairly stringent moral demands, people will anyway be inclined to resist it.

This point goes some way to explaining disagreement among the less reflective and the morally weak. But it does not touch disagreement between, say, Rossians and utilitarians, many of whom presumably have a good grasp of the various positions in play, and are both advocating potentially stringent views.

Audi's second argument, however, does focus on philosophers. He suggests that philosophers may hesitate to accept certain moral principles because they are unwilling to accept that the self-evidence or necessity of these principles is itself self-evident. That may be true. But there seems little reason to think that, if all those involved in philosophical disagreement were fully informed of Audi's claim that self-evidence need not itself be self-evident and the moral views of those who did not accept that claim were then discounted, there would not remain sufficiently widespread and deep disagreement for suspension of judgment to be required.

A third suggestion of Audi's is that, although people disagree about the truth of the Rossian principles, they might nevertheless agree about the basic 'moral force' of the considerations to which they refer:

> For instance, whether or not we accept Ross's principle concerning promising, we might, both in our reflection and in regulating our conduct, take our having promised to do something as a basic moral reason to do it—*basic* in the sense that its reason-giving force does not derive from some other reason. This is a case of agreement *in* reasons. (61)

The distinction Audi has in mind here is somewhat hard to grasp. I myself would be ready to describe someone as broadly Rossian who, both in reflection and in deliberation, accepts one's having promised to φ as a basic reason to φ. So perhaps the distinction is between someone who is narrowly Rossian, holding to the very letter of Ross's principle, and someone more broadly Rossian. But of course there are many positions on promising other than these two, and many thinkers would disagree with both forms of Rossianism.

Audi goes on to outline a threefold distinction intended to provide further elucidation, between *accepting* reasons, accepting them *as* reasons, and *conceptualizing* them as reasons. An example of the first is my doing something simply because I promised to; of the second, my accepting a student's explanation of his lateness; and of the third, my wondering whether disliking people is a reason against recommending them for jobs. Audi suggests that agreement in reasons—"operative" agreement—is found at the first two levels.

The kind of agreement in the promising case quoted above seems different from mere acceptance of promising as a reason, since it involved taking having promised to φ as a basic reason to φ. First-level agreement in reasons could be found between, say, an unreflective 'plain man', a reflective Rossian, and an act-utilitarian: each may see moral force in the fact that one has promised to φ. But of course there will be no agreement at the more general explanatory and justificatory level of ethical theory, so operative agreement here cannot help an ethical theorist meet the consensus condition. And the same point goes for the second level of operative agreement: accepting as. You and I may both accept a student's excuse as a reason for his not having arrived earlier, but we may differ at the level of theory in our account of how illness affects normative reasons.

Audi suggests that part of the appeal of intuitionism lies in the fact that there is agreement *in* reasons among thoughtful people (62). But intuitionists are primarily interested in the level of agreement *on* reasons, and if there is little agreement among thoughtful people, that might be thought to make intuitionism unattractive. My suggestion in this paper is essentially that there is enough agreement on certain reasons for intuitionism to appear quite plausible, but that there is still a huge amount of disagreement that the intuitionist must face up to.

Let me return to *R*, and stipulate that *p*, in her case, is a Rossian form of ethical pluralism.[14] We now know that, on the most plausible version of ethical intuitionism, one that properly respects Sidgwick's consensus condition, *R* is required to suspend judgment about whether the Rossian view is correct, as opposed to utilitarianism, Kantianism, and various other serious contenders in contemporary ethics. She is not required to place these views on the same level as the view that it is wrong to turn NNE after turning SSW.[15] There is enough consensus on the view that this position is mistaken for it to be rejected, even if one or two eccentric supporters remain. Nor is she required to adopt any kind of normative skepticism. There is sufficient consensus on certain normative principles—such as *Pain*—to provide a sound basis for ethical intuitionism. But two questions remain. How should *R* proceed in philosophy? And how should she live?

As regards philosophy, there are at least three strategies worth considering: resignation, impartiality, and debate. The first would involve *R*'s refraining from any

kind of philosophical debate about moral theory. This strategy, however, is likely to seem to most intuitionists excessively pessimistic about the prospects for convergence on the truth in ethics and resulting ethical progress. The case of *Pain* suggests that truth is available, and there are many standard examples of apparent ethical progress, such as the recognition of racism, sexism, and speciesism.

So, on the assumption that progress in ethics is possible, *R* might seek to engage in impartial consideration of the various current normative theories. There is more to be said for this strategy than for resignation, and it is certainly likely to be part of the best overall philosophical package. But the fact remains that *R* is still attached to her Rossian principles, and she is thus in an especially good position to spell out such a position and display its advantages to others.

So this leaves us with the final strategy, in which debate between the advocates of the different normative theories continues. But carried out between intuitionists who have suspended judgment as to the correctness or otherwise of the view that they themselves find attractive, such debate would be less adversarial and more constructive than much in philosophy at present. This would have several significant advantages. First, each participant would be more likely to see the faults in her own position and the advantages in those of others. Second, philosophers would see that there is often greater epistemic benefit in discussing issues with those of radically different views than with some clique of one's own. Third, the aim of debate would be not the victory of one's own position but convergence on some truth, which might well be a conglomeration of various elements from several existing ethical theories. It is worth noting that these virtues of open-mindedness are to be found to some extent in the work of contemporary philosophers, including Audi himself, whose own eclectic view not only systematizes Rossian principles within a Kantian framework but also makes room for traditional utilitarian elements, and notions from virtue ethics, theories of rights, and so on.

Finally, how should *R*'s suspension of judgment change the way she lives her life as a whole? The implications may be significant. She may have started with a belief in certain apparently self-evident comprehensive ethical principles, the force of which she allowed to guide her practical decisions. Once she realizes that these principles fail to meet the consensus condition, it will become clear to her that they have no claim to self-evidence, to being objects of knowledge, or to any directly justified role in her deliberation. Because she is no skeptic, however, she cannot aim for the *ataraxia* or tranquillity of the Pyrrhonist skeptic.[16] At the end of the *Methods of Ethics*, Sidgwick—who himself fails to face up to the implications of his own consensus condition—runs into the problem that he himself cannot decide between egoism and utilitarianism, and says:

> [In] the . . . cases of a recognised conflict between self-interest and duty, practical reason, being divided against itself, would cease to be a motive on either side; the conflict would have to be decided by the comparative preponderance of one or other of two groups of non-rational impulses.[17]

In many cases, there is a good deal of first-order or substantive agreement among ethical theorists. To use Audi's example, most will agree that in most circumstances one should not stab another person. But when there is a conflict, should *R* merely

consult her own tastes and preferences? If we accept that behind *Pain* lies a general principle that each person has a *pro tanto* reason to promote her own good, then the answer to this question is that she should pursue her own tastes only if they most promote her own good. And here two issues are likely to be salient. First, going against her own view may impose costs in terms of guilt, shame, regret, depression, lack of self-esteem, and so on. But, second, going with her own view as opposed to that of others may damage her well-being in other ways, especially if that view is an especially demanding one.

That ethical intuitionism should lead to a result in which each of us is left deciding which ethical theory to follow on the basis of how much each will advance our own self-interest is somewhat odd. But if normative philosophical ethics can progress to the point of widespread convergence on a comprehensive view, it would provide a practical role in our lives for ethical theory to which such theory has not yet earned any entitlement.[18]

Notes

1. R. Audi, *The Good in the Right* (Princeton, N.J.: Princeton University Press, 2004). Unattributed page references in the text are to this work.

2. If we understand the notion of inference sufficiently broadly, then the pretheoreticality requirement may be said to follow from the noninferentiality requirement.

3. H. Sidgwick, *The Methods of Ethics*, 7th ed. (London: Macmillan, 1907), 338.

4. Ibid., 338–42.

5. Ibid., 342.

6. S is, that is to say, an "epistemic peer" of R. See S. McGrath, "Moral Disagreement," *Oxford Studies in Metaethics* 3, ed. R. Shafer-Landau (Oxford: Oxford University Press, forthcoming), sect. 3. McGrath refers to work by T. Kelly and by R. Feldman, noting that Kelly attributes the term to G. Gutting.

7. See Sextus Empiricus, *Outlines of Scepticism*, trans. J. Annas and J. Barnes (Cambridge: Cambridge University Press, 1994), sects. [15], [31], [151], [196], etc.; also G. Striker, "Historical Reflections on Classical Pyrrhonism and Neo-Pyrrhonism," in *Pyrrhonian Scepticism*, ed. W. Sinnott-Armstrong (New York: Oxford University Press, 2004), 13–24, at 16. It is worth noting that since the consensus condition itself is controversial, its advocates may have to suspend judgment on it. But, as we shall see below, this does not prevent their continuing to advocate it.

8. See Audi, *Moral Knowledge and Ethical Character* (New York: Oxford University Press, 1997), 80–81.

9. See David Wiggins, "Truth, and Truth as Predicated of Moral Judgments" in *Needs, Values, Truth* (Oxford: Blackwell, 1987), 139–84, at 149–51.

10. See W. Sinnott-Armstrong, "Moral Relativity and Intuitionism," *Philosophical Issues* 12 (2002): 305–28, at 312.

11. My argument here applies also to Audi's suggestion that there is near consensus on the principle that one should not stab others (67).

12. It might be claimed that strict Rossians are merely advocating certain midlevel principles, on which there is likely to be consensus, with no overarching general theory. But of course, strict Rossians will advocate these principles as ultimate—that is, they are as general as principles in ethics get. And under that description they will be highly contested.

13. See Klemens Kappel, "Challenges to Audi's Ethical Intuitionism," *Ethical Theory and Moral Practice* 5 (2002): 391–413, at 399.

14. Ross himself certainly took his view to be self-evident. See, e.g., *The Right and the Good* (Oxford: Clarendon, 1930), 29–31.

15. Philippa Foot, "Moral beliefs," repr. in *Virtues and Vices* (Oxford: Blackwell, 1977), 110–31, at 111.

16. See Sextus, *Outlines*, [10], [12], [18], [25–31], etc.

17. *Methods of Ethics*, 508. It is worth noting that Audi himself succeeds in avoiding any head-on conflict of the kind found in Sidgwick between a purely self-regarding theory of practical reason and one that allows weight to the interests of others than the agent, through allowing reasons to be grounded in the interests of either the agent or others and to be weighable against one another in particular cases. See, e.g., *The Architecture of Reason* (New York: Oxford University Press), chap. 6.

18. I am grateful to Brad Hooker for presenting on my behalf an earlier draft of this paper to the symposium at Notre Dame on the work of Robert Audi; to Joshua Gert for his challenging response delivered at the symposium; and for helpful comments and discussion to Robert Audi, Larry BonJour, Michael DePaul, John Greco, Brad Hooker, Mark Timmons, Pete Tramel, and Nick Zangwill.

Metaethical Reflections on Robert Audi's Moral Intuitionism

HUGH J. MCCANN

Among Robert Audi's many philosophical contributions is a penetrating and subtle development of moral intuitionism. In several papers,[1] and especially in *The Good in the Right*,[2] Audi builds on a foundation laid primarily by W. David Ross, working out a clear and sensitive treatment of the intuitionist position, and defending it against numerous objections. I find much that is attractive in the resulting view—in particular its appreciation of the importance of moral experience in the development of ethical norms, and Audi's demonstration that, whatever Ross may have thought about the matter, an intuitionist stance concerning moral epistemology is in no way incompatible with, and can in fact be reinforced by, a unifying and comprehensive moral theory. But a good deal of what Audi says is to me congruent with and at times even suggestive of a somewhat different perspective that I think would be of use in further addressing challenges to intuitionism. And since I believe the perspective in question is also indispensable for the related task of making sense of practical decision-making and action,[3] its fit with Audi's views seems especially deserving of attention. Let me begin, then, by extolling some of the virtues of Audi's intuitionism, as well as looking at a potential criticism or two; I shall then try to spell out the alteration in perspective I would propose, and the ways I think it can be helpful.

I. Intuitionism and the Aims of Metaethics

David Solomon has observed that most of analytic metaethics may be seen as concerned with solving two problems.[4] The first is the problem of *justification*, of showing how it is that judgments of right and wrong are tied to factual judgments—that is, to judgments about the natural features of actions, and the contexts in which they occur. The second problem is that of *motivation*, of how a moral judgment or belief, once held, is able to motivate rational conduct. Progress in addressing one of these problems seems often to be at the cost of diminished prospects for the other. Thus, an ethical naturalist would have the easiest time with the justification problem,

since for him the valuational features of actions are simply entailed by their naturalistic descriptions. There are, however, few naturalists, and once the gap between naturalistic description and moral judgment is admitted, this problem becomes very difficult to solve. As for the motivation problem, to treat value judgments as propositional at all is to make their relation to moral motivation difficult to discern. By contrast, the motivation problem poses little difficulty for the classical noncognitivist, since for him so-called "value judgments" just *are* expressions of what amount to motivational states: conative attitudes, not cognitive judgments. But this appears to run afoul of obvious facts: that value judgments, though perhaps at times a vehicle for venting emotions, are nonetheless in themselves propositions seems as evident as anything could be, and people who disagree about them do not take themselves to be in a conflict of emotions. Worse, disaster appears in the offing on the issue of justification, since noncognitivism threatens to collapse into mere subjectivism, which would leave value judgments with no factual worth whatever.

How does intuitionism fare in addressing this implicit dilemma? Ignoring for the moment the problem of motivation, intuitionism is best understood as an effort to develop principles of normative ethics in such a way that the issue of justification is directly dealt with. In both Ross's and Audi's accounts, the theory is centered upon a set of principles of prima facie duty. These are midlevel principles, such as that we ought to keep our promises, or that others are to be treated justly. Each covers a fairly large piece of ethical terrain, and can serve as a basis for generating subprinciples (we must not swear falsely, it is wrong to cheat). But no one principle is given as fundamental to all of ethics, and no claim of exhaustiveness is made for the list as a whole. Rather, prima facie duties in different areas are on a roughly equal footing, and the list of them could in principle be extended. Most important, prima facie duties are held to be self-evident: that is, they may be known intuitively by anyone who properly understands them.

This last claim is the one we need to examine, but before doing so we should be aware of the advantages of this approach to ethics. One of them lies in the theory's focus on providing principles for moral decision-making that are neither too abstract, nor tied to any one overarching normative theory. In this it displays a strong affinity with the case study approach found in much of applied ethics—an approach, Audi observes at one point, that is at least intuitivist if not intuitionist.[5] The fact is that very little in the way of everyday moral decision-making appears to be theory based. Most people know little if any moral theory, and those who do know it need not appeal to it in their decisions. We do not inculcate ethics in the young by teaching them theory, and when we disagree with others on some issue we may not mention theoretical considerations at all. As likely as not, we will ourselves deploy the case study strategy, either formulating examples or citing actual cases, in an effort to display the types of factual considerations we deem relevant to evaluating actions of the type at issue.

What these points suggest is that both in the task of teaching morality and in that of seeking agreement upon its principles, our first reliance is upon the moral intuitions—whatever those come to—of our audience. And that is as it should be. People in general cannot be expected to have a reliable grasp of ethical theory, and children certainly cannot. That is for specialized audiences—who, by the way, may be even less likely to agree on theoretical matters than on applied ones. Moreover,

even where we are able to proceed by citing principles, the principles are not likely to be very well absorbed unless those we address have intuitions that correspond to them. I may tell a child that it is wrong to take others' things, but the lesson will be better learned if he has suffered this kind of loss himself—or, perhaps, observed the grief of a sibling who has—than if he has had no such experience. If I want to convince you that investing in tobacco companies is wrong, I may succeed much more readily by giving you a vivid account of what it is like to die from lung cancer than by bringing forth all the theoretical means at my disposal.

Moral intuitionism fits well, then, with the distinctly nontheoretical tone of much of everyday ethical life. But it hardly follows that an intuitionist ought to eschew ethical theory. Ross himself did so, considering prima facie duties to be self-evident but not provable, and denying that there was any one general characteristic that makes right acts right. But the fact, if it is one, that midlevel principles of duty may be known intuitively no more prevents us from developing ethical theories than the fact that truths about the natural world may be known by observation would rule out the development of science. Indeed, it is open to the intuitionist to hold that the relationship between experience and theory in these two realms is quite analogous: that in both, invoking very general theoretical principles enables us to unify and harmonize otherwise disparate experiential data, resolve implicit conflicts at lower levels, and disclose common elements that might otherwise go unnoticed.[6] The result should be to reinforce our confidence in principles at the mid and lower levels, to correct putative intuitions that were in fact mistaken, and to move in the direction of what Audi calls *normative completeness*: a situation in which our principles cover the terrain of ethics completely, prescribing all the duties that we have.[7] An impressive feature of Audi's intuitionism is the progress it makes in this direction, by seeking to integrate Rossian duties according to Kantian precepts, particularly the duty of treating rational nature as an end.

Perhaps the greatest potential advantage of intuitionism, however, is the progress it promises with the problem of justification. If the claim of self-evidence for prima facie duties can be made out convincingly, it may be possible to set this problem aside. Midlevel principles of ethics would then be secure. Higher level theoretical precepts, if not themselves intuitively obvious, would have the inductive support of the midlevel; and our obligations in particular circumstances could be ascertained simply by getting the facts straight, so that we would know which midlevel rules apply. It is, however, precisely on the matter of self-evidence that difficulty begins to arise for any intuitionist theory. Let us see why.

II. Objections to Intuitionism

The basic kind of self-evident proposition, according to Audi, may roughly be characterized as

> a truth such that an adequate understanding of it meets two conditions. First, in virtue of that understanding, one is justified in believing the proposition (i.e., has justification for believing it, whether one in fact believes it or not. . . . Second, if one believes the proposition on the basis of that understanding, then one knows it.[8]

A moral skeptic might respond to this characterization with the objection that if things are this simple, then knowledge of prima facie duties should be an easy matter. Yet moral principles are notoriously matters of disagreement—too much disagreement, the skeptic might argue, to allow claims of self-evidence for them. But Audi has more than one line of reply to this objection. First, note that all that is claimed to be self-evident here is the proposition at issue, the moral judgment itself, not the fact that it *is* self-evident.[9] So if the question is, say, whether it is prima facie wrong to cause others undeserved suffering, I may feel strongly disposed to say it is, yet not really be aware that the basis for my disposition is that this proposition is self-evident—if indeed it is such. Second, it is implicit in the passage cited above that although the self-evidence of the proposition might justify me in believing it, I may actually *not* believe it, even given a sufficient understanding of it. One possibility here might be that I am myself a moral skeptic, and hence unprepared to grant truth (much less self-evidence) to *any* moral claim, even that it is wrong to cause others unmerited suffering. In other cases—the duty to help the needy could be an example—I might be inhibited by prejudice, or insecurity about my own well-being. Finally, there is the possibility that I might not understand the claim sufficiently. Even Ross did not hold that prima facie duties would be self-evident to persons who are immature, or insufficiently attentive to the proposition at issue. And it can be expected that in some cases, anyway, failure to recognize a prima facie duty will be owing to a lack of maturity, training, or sensitivity on the part of the perceiver.[10]

In short, as he himself remarks, Audi's definition of self-evidence makes it possible for a claim of duty to be self-evident, yet not at all obvious.[11] Depending on an individual's circumstances, any of a number of factors might prevent his seeing the truth of such a claim. Discerning prima facie duties is not, then, all *that* easy, and we should not expect universal agreement about them. But now I think the skeptic would be well served to switch to another objection. This sense of self-evidence, he might argue, is beginning to sound a little thin. Not that there is nothing at all to be said for it. It is hard, for example, to see what sort of evidence would ever be allowed to count *against* the claim that it is wrong to cause others needless suffering.[12] The trouble is, the basis for the claim is too unclear. Perhaps it *is* a kind of self-evidence—something, in Audi's terms, about the claim itself—that justifies our believing it once it is grasped. Perhaps, on the other hand, all that is at work here is the universal human fear of suffering, which doesn't sound like cognitive self-evidence at all. In any case, the argument would run, until the basis for such claims is made clear, statements expressing Rossian duties run the risk—however convincing they may be to those making them—of being no more than arbitrary assertions of one or another moral perspective. Faced with the skeptic's doubt, the intuitionist dogmatically asserts that his claims are true, as anyone of due maturity and decent background can see, and that is that.

It is important to be clear about the nature of this objection, because Audi's version of moral intuitionism is well able to defend itself against some allegations of arbitrariness. As long as the focus of the objection is relatively narrow—targeting, perhaps, the claims of a particular person or group, or a particular duty or range of duties said to be self-evident—the intuitionist may find support for his views. If the testimony of a particular person or group is questioned, broader support may be

available. That there is disagreement over many ethical questions cannot be denied; but on many issues there is widespread agreement, and where consensus can be found it surely counts against accusations of dogmatism on the part of any particular person or party. The background of an individual can count as well: the judgments of an experienced health professional on questions of medical ethics are on the whole more to be trusted than those of a layperson. Above all, Audi's version of intuitionism can appeal to the support of moral theory. Intuitive judgments about acceptable treatment of prisoners of war, for example, are surely more reliable if they have the support of moral theory admired—as is Kant's—for the regard in which it holds the individual. I would urge, however, that replies of this kind are of very little use against the arbitrariness objection as framed above—which is aimed not at any particular range or subset of supposedly self-evident moral intuitions, but at the very idea of such a thing. The concern of the moral skeptic is a blanket one; the worry he poses is not that one or a few moral intuitions may lack appropriate support, but that the entire gamut of our cognitive moral judgments is without objective backing, that in fact there is no objective reality at all that answers to them. That we should agree on such judgments, or that they may effectively be organized in terms of some theory is of little avail against this accusation, and neither is a bare claim of self-evidence. What is needed is that the claim be given substance, that we have some account of how such a thing could be.

What, then, could be the basis for a claim that true propositions expressing prima facie duties are, in themselves, intuitively self-evident? An initially promising answer is that such principles are in fact necessary truths, in the broadly logical sense—that is, propositions whose truth can be discerned simply from a clear understanding of the concepts they employ. This was Ross's understanding of the matter, and Audi too expresses a preference for a rationalist version of intuitionism, though he does not pursue the matter very far.[13] It is possible, moreover, to make considerable progress along these lines. Consider, for example, the proposition that it is prima facie wrong to take what belongs to others. In the sense of the term at issue here, to say something "belongs" to another is to say it is the other's property—something to the possession and use of which he has a right, other things being equal. But if he has that right, then certainly it would be wrong in ordinary circumstances—that is, prima facie unjust—to deprive him of it. Or, consider a principle close to the example we have been using: that it is wrong to cause unwarranted suffering. The idea of warrant implies justification. Supposing, therefore, that a certain instance of suffering would be *un*warranted were it to occur, it must lack justification, and we would be wrong to cause it. Numerous other principles, especially principles of justice, can be given the same treatment.

Despite its initial appearance of success, however, the rationalist approach is unlikely to succeed against the skeptic. One reason is just that it doesn't always seem to work. Suppose we modify our second example to speak not of unwarranted but of unnecessary suffering. Is it wrong prima facie to cause unnecessary or avoidable suffering? Intuitionists are apt to say it is.[14] But it is hard to see how terms like "unnecessary" or "avoidable" can bear the moral weight that "unwarranted" does. We could try to insist on the point: to claim, perhaps, that in this case "unnecessary" means morally unnecessary, or unnecessary for any justified end. But this kind of move will

be convincing only to the choir. And if we drop all talk of warrant or avoidability, the situation gets worse. Consider the claim that it is wrong to cause someone pain just for fun. For the intuitionist, I think this must be self-evident; indeed, if there is a paradigm case of causing unwarranted suffering, this has to be it. Here, however, there seems no hope of finding the concept of moral wrongness implicit in the description of the act, and lacking a conceptual tie the intuitionist, and for that matter any moral realist, must find another way to make the connection.[15]

This challenge is symptomatic of a deep-running and pervasive problem—that of what J. L. Mackie calls the "queerness" of moral properties—a problem that threatens to undermine even moral principles that are fairly taken as necessary. It is possible to know that a proposition is a conceptual truth, yet be in doubt as to whether it applies to the world. Thus, we can know from concepts alone that all unicorns are one-horned, yet wonder whether there are any unicorns. And in the same way, we can be convinced that to cause unwarranted suffering would be wrong, yet wonder whether there truly could be such a thing as suffering that is objectively unwarranted—suffering that, of itself, has no moral business occurring. The same goes for the concept of property. Granted, we may say, it would be wrong to take another's property. But is there really such stuff—stuff that belongs to others just by some sort of natural right, however circuitously derived? There is, moreover, an important difference between the case of unicorns and cases of duty. With the former, we can tell whether a principle true by definition takes hold of reality simply by observation, by looking to see whether the essence *unicorn* is instantiated in the natural world. With the ethical cases the situation is different because, unlike *unicorn*, concepts like *unwarranted suffering* and *property* do not pertain straightforwardly to natural observables. They are, we might say, *deontically laden*: they have to do, not with what is or is not, but with what is permissible and impermissible, with what may or must be done. And it is in no way obvious that this is a matter for straightforward cognitive observation. This is not to say the rationalist approach to principles of prima facie duty is simply misguided. Such principles are at times conceptually true, and when they are, conceptual analysis will tell us our duty—as long as we can see that the principles have application to the world. The difficulty is to see that they have application. That is not a conceptual matter, and hence not one conceptual analysis can address. Neither, of course, does this approach help with principles such as that it is wrong to cause pain just for fun, which is not a conceptual truth, yet seems to have a pretty fair claim to being considered intuitively obvious. How should the intuitionist react to this problem? One possibility is simply to dig in one's heels, to insist that we simply do see that causing suffering for fun is wrong, that it is a matter of mere observation that such suffering is unwarranted, that the same holds for taking things that others physically possess, and similarly for the entire list of Rossian duties. To do so, however, would in effect be to give up the struggle, to lapse into the very dogmatism the skeptic is protesting. And there is another reason to avoid this type of move. It is generally allowed that the valuational properties of things supervene on their natural properties: that is, value properties are determined by natural properties, and can vary only if the natural properties vary as well. In the ethical realm, this means that the permissibility or forbiddenness of an action must arise out of the description of the action naturalistically considered—that is, out of its description shorn of all valuational content. Thus, the

forbiddenness of causing pain for fun has to arise out of the descriptive nature of that action. Now the problem we have been exploring—that of the basis for intuitionist claims of self-evidence—is closely related to a problem about supervenience: that of just *how* it is that the deontic properties of actions supervene on their natural properties. Any substantive solution to one of these questions is likely to shed light on the other. Equally, to give up on either problem is to threaten the other with obscurity. Thus, if we decline to say how the intuitionist's judgments of duty are grounded and simply insist that they are obvious, we threaten to turn moral supervenience into an occult relation, to force ourselves to say the moral properties of actions just do arise out of their natural ones, and that is that. A better solution is needed on both fronts.

III. Felt Obligation

I think a better solution can be found if we begin with the will: specifically, with what we may call *felt obligation*—that is, the sense of duty each of us is prone to feel when we consider courses of action we take to be demanded of us.[16] Motivationally speaking, felt obligation lies at the heart of moral behavior—that is, of acting from duty—and so is quite a frequent phenomenon. It is a kind of movement of will or emotion, the thrust of which is that although, subjectively, there is a choice to be made, there is a certain objective settledness about the options, a demand that one course be followed. I want to suggest that although felt obligation is at least largely a conative state, it may legitimately be viewed as a form of *objective* awareness, and hence as an appropriate ground for judgments of duty, judgments that, depending on the circumstances, may be either prima facie or final. There are good reasons for viewing felt obligation in this way, and when we do, most of the difficulties usually thought to attend moral intuitionism fade away.

A view of this kind is in fact suggested by Audi himself, in his claim that moral emotion can serve as a noninferential ground of cognitive moral judgment:

> Moral sensitivity can run ahead of judgment. We may sense a duty, say, to help someone, without any good idea of whether the duty derives from a tacit promise or from beneficence or both. We may thus *respond* to appropriate grounds before forming a *belief* that they are present or making the corresponding judgment, here the judgment that we should help. This is one kind of case in which emotion, say compassion or indignation, can be both morally evidential and morally motivating. Emotions may reveal what is right or wrong before judgment articulates it; and they may both support ethical judgment and spur moral conduct.[17]

The idea here is that the sorts of grounds that evoke moral judgment—that is, I take it, the base features on which the rightness or wrongness of an act supervenes—are able to arouse in us, perhaps in the form of some specific moral emotion, a sense of duty. This sense of duty is strongly conative in nature, and hence intrinsically motivating; but it can also serve as a legitimate basis for moral judgment—that is, for an act of belief formation that a certain course of action is right or wrong—without the aid of intervening beliefs that might serve as supporting premises. Thus, "Emotion can sometimes interpret people or situations before a search for propositional evidence

can find the crucial points, or even where it cannot."[18] This seems to me precisely correct, only understated. I want to suggest that this phenomenon is far more frequent than is implied in Audi's treatment, and has a claim to being viewed as the primary experience through which we become aware of right and wrong.

This may seem a bold and unrealistic thesis, since we are prone to think of conative states as strictly subjective in origin and import, and hence without bearing on questions about what is objectively real. There are, however, good reasons for taking the suggestion seriously. We have spoken of the unusual character of value properties—in Mackie's term, their "queerness." What this comes to, really, is simply that value properties are teleological. They have to do with legitimate objectives: in the case of the good, with what is worth having or doing; in the case of the right, with what, deontically speaking, must be sought or performed. And we have taken note of the centuries-old difficulty of how such properties might be made available for strictly cognitive recognition. It is hard to understand how our cognitive faculties can read directly in what is present in the world—in what we actually see—what it would be good to have there, or what we ought to put there. As long, however, as we are prepared to think the good and the right are genuine, objective realities, rather than illusions or mere human inventions, it is not unreasonable—in fact, it is appropriate—to think they are detected through experiences of will and emotion. Indeed, it is hard to see how we *could* become aware of objective ends except through the experience of being drawn or even shoved toward them. And that is exactly what conative awareness is about: we feel the tug of the good in desire, and in felt obligation we are impelled toward what is obligatory, and driven away from what is forbidden.[19]

It is the latter kind of experience that concerns us here.[20] The experience of felt obligation is, I think, familiar to all. Indeed, if it were not, it is hard to see how moral uses of the word *ought* could have real meaning for us. To sense obligation is to feel bound to seek some end or perform some action—not in the sense of having an overwhelming desire, but in the irreducibly deontic sense that the end or action is categorically prescribed. The experience is often expressed by the one who has it saying "I must do this," and there is indeed a kind of necessity involved—not that of causality, but the deontic necessity associated with duty, the necessity of an action that will occur in any permissible extension of the present world.[21] Despite the fact that the experience is likely to be expressed propositionally, however, it is not primarily cognitive. It is a movement of the will, an experience of being impelled to act, of the sort commonly had upon receiving a command from an acknowledged authority.[22] It is, no doubt, this aspect of felt obligation that has led philosophers in the Kantian tradition to frame the deliverances of morality as imperatives rather than declarative statements. That is very much in keeping with the position I wish to defend, except that the imperatives, though perhaps testable with the aid of pure reason, should not in my view be understood as handed down by it.[23] On the contrary: they are as available to children as to moral philosophers, because they arise not from reasoning but from conative appreciation of the acceptable ends implicit in the descriptive features of situations that call for a moral response.

Felt obligation, though conative, is to be distinguished from mere desire, even a desire to do one's duty. We think of desire as impelling us to act only when it is unusually strong. But even then, the force we associate with desire is quite different

from that associated with duty. When we speak of the force of desire we have in mind its motivational strength: that is, the degree to which our subjective preference is focused on the end in question, and our consequent tendency either to pursue the end, or to be disappointed if we forego it. The force of felt obligation is not a matter of subjective preference or impulse but of objective demand.[24] That force—we might call it the *mustness* of obligation—is present even when the motivating power of duty, the degree to which our sense of obligation dominates our deliberations and, perhaps, our proneness to act on it, is less than that associated with our desires of the moment. All the same, the objective necessitation implicit in the sense of duty is supposed to prevail in our decisions even when the motivational force of desire is stronger. The whole point of duty is to overrule desire where the two conflict. Accordingly, it would be a mistake to treat felt obligation as just one more kind of desire—as if, in order for the notion of duty to fulfill its function in deliberation, I must have a desire to do my duty that outweighs other desires. On the contrary: when, out of felt obligation, I return a borrowed book, I act out of a motive whose content is *to return the book*. This is quite different from the content of a desire to do my duty—which is simply the notion of duty itself, apprehended as attractive. It is, no doubt, possible to act out of such a desire, but the idea of so doing has a certain fatuous ring to it—nothing like the seriousness of acting from a feeling of obligation.

Since the sense of duty is primarily conative, it would also be a mistake to identify experiences of felt obligation with judgments of duty. To feel morally bound to return a borrowed book is not in itself to judge that I am obligated to do so.[25] It is, certainly, a natural next step to form such a judgment upon experiencing the feeling of obligation. And if we are prepared to grant that such experiences can be a means of discerning objective duty, the experience can justify the judgment. But the two are different. If I feel obliged to return a borrowed book I am, in that very experience, necessarily motivated to do so, for conative experiences just *are* motivational states. If I only judge that I am obliged to return it, it is highly debatable whether I must be so motivated.[26] Perhaps anyone serious enough about ethical matters to take up the question of what he ought to do in a given case is likely to be at least somewhat motivated by the answer. It must be remembered, however, that even if, as I am claiming, felt obligation is a legitimate ground for judgments of duty, our beliefs about morality can derive from many other sources as well: formal instruction, the advice of friends, theoretical reflection, and so on. In light of this, it is not at all clear that a judgment of obligation can be understood to guarantee corresponding motivation to act. It ought to be at least possible for someone to judge or believe it to be a duty to perform a certain action, yet not feel obligated to do so.

Normally, however, one does feel obligated in such situations, and if the view I am suggesting is correct, we can begin to see why. If the primary ground for judgments of obligation is our conative sense of duty, then truly to *apprehend* an action as obligatory is not to judge it to be in the category of prescribed conduct, or even to judge it to be so on the basis of its pertinent descriptive features. Rather, it is for those features to *evoke* in us a sense of duty, the conative sense that the action must be performed. Lacking that—as we might if, for example, we adopt a moral belief solely because it comes from a respected authority, instead of gaining it by understanding

the action itself—the belief will be hollow, rather like believing some vitamin is good for us when we have no idea what it does. Morality is best taught by examples because if we rely solely on received judgments of duty, our ethical beliefs are apt to lack the appropriate conative foundation, and so fail to evoke in us the kind of motivation that will guide us to act from duty when that is called for. Accordingly, moral advice is very often accompanied by reminders of the features of the action in question that justify it—the very features likely to invoke in us a sense of duty about the action.

IV. Intuitionism with a Conative Basis

The brand of intuitionism I am proposing is like Audi's, then, in being consonant with the case-oriented character of most moral instruction and day-to-day ethical discourse. It also preserves the other advantages of Audi's approach. In itself it is theory neutral, but it is perfectly agreeable to our developing moral theories, and stands to profit from the capacity of such theories to systematize, unify, and at times correct ground-level intuitions about right and wrong. It also allows for the intuitions of individuals to be checked against one another. We need not insist that any one person's sense of duty will always be well directed. Felt obligation depends on our cognitive grasp of the descriptive features of whatever action is under consideration, which need not always be accurate. In addition, one's sense of duty can become confused or biased through poor education or harmful experience, weakened by the formation of bad habits—even, in some cases, depraved through serious misconduct, devotion to evil ideologies, or the like. When such things occur, appeals to both collective experience and theoretical reflection are legitimate correctives that intuitionism can and should accept.

To be sure, this kind of correction requires that collective moral experience be trustworthy—something that may be doubted in light of the disagreement about values that exists from individual to individual and from culture to culture. The fact is, however, that in the end there just *is* no other basis for moral correction and advancement than the common heritage of human moral experience, and the theoretical reflection through which it is codified and organized. There is no other way to proceed, nor should we despair of success in the project. Perhaps there is no such thing as an entirely impartial moral observer, and certainly we cannot specify theoretically who would count as such an observer without invoking values. But we can work toward agreement at the theoretical level. Far more important, however, is the fact that theoretical specification need not be necessary: serious moral advice is much more a matter of practical wisdom than of theory, and we may find it far easier to agree that one or another individual is worth listening to on moral matters than to agree on the theoretical basis for the claim. Equally, once we agree on the descriptive facts about a given action and the circumstances in which it occurs, we may find it far easier to agree on the morality of the act than on what theory should be invoked to explain our agreement.

A conatively based moral intuitionism would not, then, invest individual experiences of felt obligation with any kind of incorrigibility. It is, however, against the accusation that intuitionism involves a blanket kind of arbitrariness that the view

I am suggesting offers real progress. We saw earlier that a strictly cognitive intuition-ism is open to the charge that it is unable to explain *how* it is that we are able to intuit directly the moral rightness or wrongness of an action simply by considering its descriptive features. Descriptive features do not entail evaluative ones, and it is hard to see how they can be directly intuited by our cognitive faculties—unless, of course, we wish to postulate a unique moral sense, which is plainly an ad hoc maneuver. The solution to this problem, on the view I am defending, lies in the fact that the moral features of actions, and the principles of duty that go with them, are not reached by direct cognitive intuition. Our cognitive recognition of them is *indirect*, grounded in the conative experience of felt obligation, in which we apprehend our duty by feeling impelled toward or away from a given course of action by our cona-tive sense that the action is commanded or forbidden. It is through this kind of expe-rience that duty is first grasped, and only in light of this experience is it judged to be right or wrong. The principles that emerge, since they are propositions, must still be treated as cognitive intuitions, but the intuition occurs only by means of conative experience, rather than through a direct appreciation of the ethical features of actions that is somehow to be found in their natural description alone. If this is cor-rect, we can readily see how a principle such as that which says it is wrong to cause pain just for fun can appear self-evident even though it is not a conceptual truth. The principle *is* self-evident, but not to our unaided intellect. It is evident first to conative moral awareness, in the abhorrence we find in this kind of act, and in light of this experience it becomes evident to the intellect. The advantage here is that we need not invoke any exotic cognitive power. Rather, sensitivity to moral truth is seen on this view as intrinsic to a form of experience with which all are familiar, and which in light of the teleological nature of moral properties is naturally taken as a response to them. A second important gain offered by the present view is the light it sheds on the issue of moral supervenience; on the way in which moral properties arise out of strictly descriptive ones. They do so in that, given their descriptive nature, certain actions are appropriately taken by us as commanded or forbidden—that is, they are appropriately experienced through our conative faculties as actions we must or must not perform. Far from being occult, the supervenience of the ethical is simply a mat-ter of the way the world presents itself to our will, which by its nature is an evalua-tive faculty. Finally, this account of moral apprehension helps us understand the close relationship between moral belief and moral motivation that has been the sub-ject of so much recent discussion. It is not that cognitive moral judgments simply *are* motives; that is just a confusion. Nor is it that such judgments of duty simply engen-der the appropriate motive in us through a kind of sheer causal hocus-pocus. It is, rather, that judgments of duty may well, in the individual in question, be grounded in such motivation to begin with, and where they are not may arouse such motiva-tion by being so presented as to call to mind the features on which the moral char-acter of the action in question supervenes.

I think, then, that the kind of position I am defending has a good deal going for it. By contrast, I see little force in the sorts of considerations that might be raised against it. One might protest that this view does nothing to dispel Mackie's complaint about the "queerness" of value properties. In my view, however, Mackie was quite right on that score: from a naturalistic perspective value properties *are* queer, in that they are

teleological. They treat the preferability of some outcomes over others, and demandedness of certain actions, as objective features of the world. And one might be disposed, as Mackie was, to reject entirely the existence of such things. One might insist that the world is a directionless place, and can present only the illusion of any one state being preferable to others. But this kind of position is not just out of keeping with the spirit of intuitionism: to take it is to give up objectivism in ethics altogether, and once that is done I see no way to provide value theory with any kind of firm foundation.

Moreover, I have no idea how it might be demonstrated that the objective features of the world are confined to those we call "naturalistic" without simply begging the question against the idea of the will as a source of value apprehension. One may, of course, have doubts on the latter score, but I would urge that such doubts are, as well as antirealistic, unrealistic. After all, like sensory perception, the experience of felt obligation is essentially passive, a psychological occurrence in which we react to real or imagined situations in which various ends might be pursued, and various actions are possible. Why assume, then, that although our perceptive faculties are able to get a grip on something real in such situations, our conative capacities grasp nothing at all? The more reasonable supposition, it seems to me, is that each faculty is attuned to the kind of reality to which it needs to be sensitive in order to assist us in understanding the world and making our way through it. In the case of the will, that assistance involves changing the world in ways that, if objectivism in ethics is correct, are suitable not just to us but also in themselves. The experience of felt obligation fits with that. To feel duty bound not to cause needless pain is not to experience myself as having a subjective behavioral impulse. It is to experience the thing itself as forbidden. Were it not so, all deontic experience would be trivialized, and we could not take ourselves seriously as moral deciders. In practical life, we are not capable of that. We have to treat our sense of duty as putting us in touch with something real. True, the reality in question does not sound "scientific." But that is as it should be, inasmuch as the reality is teleological in nature, and science does not aim at discerning such things. That, however, is not a reason for thinking the reality is not there; and as for behaving as though it were not there, we could not do that if we tried, any more than we could behave as though there were no physical world. The version of intuitionism I have defended suggests only that we theorize about ethics with the same authenticity—something very much in the spirit, I think, of Robert Audi's approach to the subject.[27]

Notes

1. Several of these are collected in Robert Audi, *Moral Knowledge and Ethical Character* (New York: Oxford University Press, 1997), parts 1 and 2.

2. Robert Audi, *The Good in the Right: A Theory of Intuition and Intrinsic Value* (Princeton, N.J.: Princeton University Press, 2004).

3. See my *The Works of Agency* (Ithaca N.Y.: Cornell University Press, 1998), chap. 11, pages 213–33.

4. David W. Solomon, "Ethical Theory," in *The Synoptic Vision: Essays on the Philosophy of Wilfrid Sellars*, ed. C. F. Delaney, M. J. Loux, G. Gutting, and W. D. Solomon (Notre Dame, Ind.: University of Notre Dame Press, 1077), 149–88, 151–52.

5. Audi, *The Good in the Right*, 162.

6. Compare ibid., 82–83.

7. Ibid., 85.

8. Ibid., 48–49.

9. Ibid., 29.

10. Ibid., 31.

11. Ibid., 49.

12. Christopher Peacocke, "Moral Rationalism," *The Journal of Philosophy* 101 (2004): 499–526, 501.

13. Audi, *The Good in the Right*, 54. Peacocke's "Moral Rationalism" also defends this view.

14. Peacocke, for example, "Moral Rationalism," 501.

15. Compare J. L. Mackie, *Ethics: Inventing Right and Wrong* (New York: Penguin, 1977), 41, from which this example is taken.

16. The importance of felt obligation was emphasized by Maurice Mandelbaum in *The Phenomenology of Moral Experience* (Baltimore: Johns Hopkins Press, 1969). More recent treatments include Terry Horgan and Mark Timmons, "Moral Phenomenology and Moral Theory," *Philosophical Issues* 15 (2005); and William Tolhurst, "Seemings," *American Philosophical Quarterly* 35 (1998): 293–302.

17. Audi, *The Good in the Right*, 87; cf. 56–57.

18. Ibid., 57.

19. Compare Mandelbaum's observation that the demandedness of an action is experienced as a force, although he gives less emphasis to the conative dimension of the experience. See *The Phenomenology of Moral Experience*, 54.

20. I have argued for the thesis that desire is a means of apprehending objective good in "The Will and the Good," in S. Nannini and C. Lumer, eds., *Deliberation, Intentionality, and Autonomy: The Action-Theoretic Basis of Practical Philosophy* (Aldershot: Ashgate Publishing, 2007).

21. For this approach to deontic semantics see Hector-Neri Castañeda, *Thinking and Doing* (Dordrecht: Reidel, 1975), chap. 8.

22. The experience of felt obligation can usefully be thought of as the experience of a kind of internalized command. An interesting aspect of this perspective is that although the force is that of a received imperative, imperatives cannot be stated in the grammatical first person. As a consequence, the experience tends to be reported in the indicative, e.g., "I must return this book to the library," a fact that reinforces the supposition—mistaken on the view I am suggesting—that the deliverances of morality are at bottom cognitive in origin.

23. Kantian tests of maxims of action in terms of one or another form of the categorical imperative can easily be extended to judgments of obligation. But of course the value of the result will depend on whether the categorical imperative itself has moral force, and how that force is gained. On the view I defend, its force must derive ultimately from our nature and circumstances as moral beings, not subjective authority.

24. Mandelbaum, *The Phenomenology of Moral Experience*, 54–55.

25. Nor is it a substitute for such a judgment. That is to say, I am not defending a version of noncognitivism, according to which what appear to be cognitive judgments are in fact effusions of emotion, disguised commands, or the like. On the contrary, conative experience *of* value is functionally quite different from cognitive judgments *about* it, and it is important to defend the integrity of both.

26. The view that judgments of duty are automatically motivating, often called "motivational internalism," has been the subject of much debate. Audi gives a penetrating evaluation of it in his "Moral Judgment and Reasons for Action," chapter 10 in *Moral Knowledge and Ethical Character*. See also Sigrún Svavarsdóttir, "Moral Cognition and Motivation," *The Philosophical Review* 108 (1999): 161–219.

27. I am grateful to Robert Audi for numerous conversations about these issues, as well as to Mark Timmons and William Tolhurst for useful comments and discussion.

Two Conceptions of Morality

BERNARD GERT

Abstract

There are two conceptions of morality, one championed by Aristotle (Ross) and Kant, the other championed by Hobbes and Mill. Audi presents a powerful version of the first tradition by using Kant as a foundation for Ross. At the end of his book, he presents ten midlevel axioms that are intended as modifications of Ross's prima facie duties. All of these self-evident middle axioms are supported by various versions of Kant's categorical imperative. I compare these ten midlevel axioms with the ten moral rules that I claim are the universally known rules of common morality. The point of this comparison is to show the great difference between the two conceptions of morality mentioned above. Not surprisingly, I strongly favor the Hobbes-Mill conception.

Introduction

It is quite surprising how similar the moral theories that Robert Audi and I put forward seem. We both start with what he calls midlevel principles or middle axioms, and what I label as moral rules and moral ideals. We both hold that the principles or rules that we start with should be accepted by every rational person without argument, but we also both hold that arguments can be put forward to support these middle axioms or rules. We also both hold that these middle axioms or rules need to be interpreted and that sometimes it is appropriate not to follow them. Indeed, looked at from a distance, or at a certain high level of abstraction, it may seem that we simply hold variations of the same theory. And in a certain sense, that is correct. But, the devil is in the details, and it is my view that Audi does not have enough of the devil in his theory. It is not that he does not deal with details; indeed, he discusses the views of Ross and Kant, as well as those of many other philosophers in great detail. What I mean by saying that he does not have enough of the devil in his

theory is that he is too respectful of the views of Ross and Kant and so accepts more of what they say than he should. It is quite likely that he believes that I have too much of the devil in my theory, that is, that I am insufficiently respectful of the views of Kant and Ross.

It is certainly true that, except for Hobbes, I think that all of the standard moral theories make fundamental mistakes. Actually, some, but not all, of what I call mistakes are not appropriately called that; rather, the philosopher in question, for example, Aristotle (Ross) or Kant, is not concerned with morality in the sense that I take to be fundamental. It is now generally acknowledged that what Aristotle discusses in the *Nichomachean Ethics* is a different subject than what we now generally call morality. It includes some of what we call moral behavior, but it also includes much more. As an Aristotelian scholar it is not surprising that Ross has a rather Aristotelian view of morality. It is not so generally recognized that what Kant's translators call "morality" is also much wider than what "morality" generally means in English-speaking countries. In German, the words for morality clearly include behavior that does not involve others even indirectly. In German, a person alone on a desert island can behave immorally, whereas in English, that is not possible, at least for those who have completed the separation of morality from religion. Of course, sex confuses matters, but neither Audi nor I think of sexual behavior as constituting a special area of morality. Like most philosophers, both of us hold that whether certain kinds of sexual behavior count as immoral depends on their violation of some nonsexual principle or rule. Because Audi develops his view by combining Ross and Kant, it is not surprising that he accepts this wider concept of morality rather than the concept of morality that I take as fundamental and that now seems dominant among those in English-speaking countries who view morality as distinct from religion.

At the end of his book, *The Good in the Right*, Audi lists ten middle axioms, which have a close relationship to what Ross calls prima facie duties, but Audi does not use the word *duty*. Partly because I list ten moral rules, I thought that it might be useful to compare my rules with his middle axioms. Both of us agree that the rules or middle axioms need interpretation and that the obligations imposed by them are prima facie. Indeed, some of Audi's middle axioms are identical to my rules; however, some of his middle axioms are very different from my rules. Comparing his middle axioms to my rules seems to me to be a good way to understand that we have very different concepts of morality. I do not think that this is a trivial matter. It is important for me to be clear and explicit from the very beginning. I am not criticizing Audi for misinterpreting Kant or Ross, or for distorting their views in order to bring them into harmony. Rather, for the purposes of this paper, I assume that Audi is correct in using Kant's categorical imperative to support Ross's prima facie duties. I am criticizing the view of morality that is shared by Audi, Kant, and Ross.

This view of morality is a much wider one than the one that I regard as appropriate. Thus my disagreements with Audi will generally not concern those middle-level axioms that parallel my moral rules. As I use the phrase "moral rules" they refer to those rules that it would be immoral to violate without an adequate justification. These rules—for example, do not kill, do not deceive, and do not cheat—can be obeyed impartially with regard to all moral agents, and unjustified violations of them

make one liable to punishment. I distinguish between moral rules and what I call moral ideals, which encourage people to prevent others from suffering harm or evil, for example, help the needy and prevent pain. These ideals cannot be followed impartially with regard to all moral agents, and failure to follow them does not make one liable to punishment, although in some situations, failure to follow them may subject one to moral criticism. These ideals differ from all other ideals in that following them sometimes provides an adequate justification for violating a moral rule, even without the consent of the person toward whom the rule is being violated. Other ideals, such as social ideals, for example, showing gratitude, are often not distinguished from moral ideals, even prompting criticism when not followed in some situations. However, unlike moral ideals, following a social ideal never justifies violating a moral rule toward someone without his consent.[1]

I will make some criticisms of Audi's middle axioms that parallel what I call moral ideals rather than moral rules. I will also criticize those middle axioms that are equivalent to what I call social ideals, but my major criticisms will concern those middle axioms that do not involve other people at all. Given that most people think that failing to do what morality requires deserves censure and perhaps even punishment, it is not a trivial matter to include more in the province of morality than belongs there.

I have almost no problems with Audi's first middle axiom, and I agree with him that it deserves pride of place.

1. PROHIBITION OF INJURY AND HARM
WE SHOULD NOT INJURE OR HARM PEOPLE.

This middle axiom contains at least the first three of my rules: "Do not kill," "Do not cause pain," and "Do not disable." It even seems to include "Do not deprive of freedom," for Audi explicitly says, "Deprivations of freedom . . . also deserve the name [harm]" (188). However, in note 21 to this page where he contrasts his view with mine, he says, "I am taking deprivation of ability to be a harm, and the other two [deprivation of freedom and pleasure], to be in general injustices" (237). If he is talking about intuitive rules or middle axioms, I do not see why he wants to distinguish these last two rules from the first three, for every rational person would support all five of these rules for the same reason, namely, that they do not want to suffer the harm involved, whether it be death, pain, disability, or loss of freedom or pleasure.

Audi and I agree that my first five rules, or his first middle axiom, are self-evident. However, Audi, as an intuitionist, regards coming to accept the middle axioms on the basis of a proper understanding or grasping of their content as adequate for being justified in accepting them, and I do not. I think that in order to be justified in accepting these rules, it is necessary to show that all appropriately described rational persons would put forward the rules as public rules governing the behavior of all rational persons. Audi seems to use Kant not to justify the middle axioms but to provide additional coherence and force for these middle axioms. He seems to hold that adopting the categorical imperative, in one or more of its formulations, aids in interpreting the middle axioms as well as providing additional support for them.

I hold that every rational person regards these rules as moral rules because he wants these rules obeyed to protect him and his friends, and realizes that if these

rules are to be accepted by all as moral rules, that is as rules that are publicly taught, all must put forward these rules as protecting all moral agents. Audi seems to accept the standard philosophical view that morality is primarily a guide to conduct that people adopt for themselves, whereas I hold that morality is primarily a public guide to conduct that people put forward for others to follow, and only sometimes for themselves. As a rational person I want everyone to follow the moral rules with regard to my friends and me, and I know that everyone else wants me to follow these rules with regard to their friends and them. So we all claim, at least publicly, that everyone should follow the rules with regard to everyone, that is, should regard them as moral rules. Given human nature, hypocrisy is not merely something that is closely related to morality, it is almost conceptually related. On the standard account of morality, which Audi seems to accept, hypocrisy is a puzzle.

It is interesting that Audi formulates this middle axiom so that it only prohibits harming people, whereas I note that there is disagreement about whether these rules should be interpreted as applying to sentient beings who are not people, that is, to nonhuman animals, or to potential people, or, as I put it, potential moral agents. In this, Audi clearly follows Aristotle (Ross) and Kant, whereas I acknowledge that some people, including some philosophers, especially consequentialists, hold that morality also prohibits causing harm to nonhuman animals and to fetuses. Thus I explain why there is an unresolvable moral disagreement about the treatment of animals and about abortion, namely, that there is an unresolvable disagreement about whether these nonhuman animals and fetuses are impartially protected, or protected at all, by the moral rules. Audi and I both hold that everyone accepts that people, by which I mean, and take him to mean, moral agents, those who are held morally responsible for their behavior, are impartially protected by morality. It may be that it is because he realizes that any enlargement of the group impartially protected, or protected at all, by morality is not self-evident, that he formulates his first middle axiom in the way he does. I do not have any disagreement if that is his motivation.

2. *VERACITY.* WE SHOULD NOT LIE.

There is an interesting disagreement between Audi's formulation of this middle axiom and my moral rule "Do not deceive." Audi explicitly argues against formulating this middle axiom as I do, because he does not view deception, per se, as even prima facie wrong. He points out, quite correctly, that deception can be unintentional, however, it seems clear that there is something wrong with unintentionally, but thoughtlessly, deceiving. I agree with Audi that lying is the clearest example of the kind of deceiving that is prima facie wrong, but I would have thought that the intent to deceive is what made lying wrong. I would regard setting up some situation involving nonverbal means in order to deceive as being as wrong as lying. Audi is correct that some lies do not succeed in deceiving, but if they were intended to deceive, they were still lies and Audi's maxim would prohibit them. If the lie was not even intended to deceive, and did not deceive, for example, those participating in a liars' club competition, then I do not see that there is anything wrong with lying. If Audi thinks that a lie must involve intending to deceive, our difference here is very small. As I interpret violating the rule prohibiting deceiving, it involves either

intending to deceive, for example, lying or withholding information when one has a duty to disclose, such as a doctor withholding information because he thinks the patient will be adversely affected by receiving that information; or saying something that one knows or ought to know to be false, to people whom one knows or ought to know will take what you say to be true.

Again, I think that the difference in our formulation depends on our view about what justifies accepting a self-evident rule or middle axiom. As noted above, Audi regards coming to accept the middle axioms on the basis of a proper understanding or grasping of their content as adequate for being justified in accepting them. I explain why the rule "Do not deceive" is self-evident because no rational person wants to be deceived, and so wants to minimize, without undue limitations on freedom, being deceived by others. Thus I do not interpret the rule "Do not deceive" as prohibiting dyeing your hair, wearing clothes that make you look younger, or so forth. And as is the case with all of the moral rules, everyone knows others do not want the rule to be unjustifiably violated with regard to them, for example, to be deceived, either. Audi seems to think his formulation receives additional support from Kant's categorical imperative. Kant explicitly claims that lying is something that no rational person could put forward as a universal maxim. However, Kant has problems dealing with justified cases of lying, so his support is a two-edged sword. My explanation why rational people would regard this middle axiom and the previous one, or the equivalent rules, as self-evident, can explain why they would sometimes want them not to be followed.

3. *PROMISSORY FIDELITY*. WE SHOULD KEEP OUR PROMISES.

Although Audi's formulation of this middle axiom and my rule "Keep your promises" are virtually identical, I think that there may be a problem for Audi as he formulates this middle axiom. This problem is related to the previous middle axiom. In arguing against "Do not deceive" as the proper formulation, Audi correctly claimed that the fact that deception could be unintentional was his reason for formulating the previous middle axiom as "We should not lie" rather than as "We should not deceive." However, this middle axiom, as formulated, can certainly be broken unintentionally. Audi, correctly, does not claim that unintentionally breaking a promise is not doing something morally wrong, so it is not clear why he does not accept that in some circumstances, unintentionally deceiving may also be doing something wrong, even if it not exploitive or involves a failure to treat someone respectfully. He could hold that all unexcused unintentional breaking of promises is either exploitive or a failure to treat someone respectfully, but I cannot see any reason to distinguish this middle axiom or rule from the previous middle axiom or rule. I recognize that John Rawls, in "Two Concepts of Rules,"[2] did claim that the rule concerning promises was fundamentally different from the rule concerning deceiving, but I do not see that he offered any argument for this view. As with the previous middle axioms and rules, Audi simply accepts the self-evidence of this middle axiom, whereas I require that the corresponding rule be one that all rational persons would put forward as a public rule to govern the behavior of all rational persons. I explain the self-evidence of this rule by noting that no one wants promises to him broken and knows that everyone

feels the same. Audi uses Kant to provide additional support for this middle axiom, because breaking it violates the categorical imperative.

4. *JUSTICE.* WE SHOULD NOT TREAT PEOPLE UNJUSTLY AND
   SHOULD CONTRIBUTE TO RECTIFYING INJUSTICE AND
   TO PREVENTING FUTURE INJUSTICE.

For some reason, Audi explicitly says in the first sentence explaining this middle axiom, "It is important to add something not clearly implicit in what was said about justice earlier: that deprivations of liberty and certain deprivations of pleasure—all of which are factually specifiable—count as injustices." I do not see the motivation for not including what I regard as the fourth and fifth moral rules, "Do not deprive of freedom" and "Do not deprive of pleasure," in the first middle axiom prohibiting injuring and harming, but this is a minor matter. What is far more significant is that Audi, similar to Rawls, includes in a single axiom or principle, a genuine moral requirement, that we not treat people unjustly, and what is only morally encouraged, that we should seek to rectify and prevent injustice. As Audi has formulated this middle axiom, I do not find it self-evident at all. I do not see that a monk or a scholar has even a prima facie obligation to leave his monastery or library in order to rectify or prevent injustice. Of course, he should not treat people unjustly, but he can follow this middle axiom and stay in the monastery or library.

Audi recognizes that "other things being equal, the first demand expressed in the principle has priority over the second and over the third. Perhaps the second duty also has priority over the third: but this need not be specified." However, this does not do away with the problem, but actually compounds it. For Audi has now explicitly claimed that all three elements of this principle are demands, and suggests that if we do not follow the second and third elements when there is no conflict with the first, or with any other middle-level axiom, we are doing something wrong. Even though Audi recognizes the distinction between moral rules and moral ideals he does not seem to appreciate its force. He does not realize that when talking about moral ideals, which cannot possibly be followed all of the time, people must have a choice about when, or even if, to follow them. There are many competing moral ideals and Audi presents no argument for claiming that any particular ideal need be followed by every person. It is clearly impossible for any person to follow every moral ideal, so I cannot see that this principle, as phrased by Audi, is self-evident. No doubt it is morally good to justifiably rectify injustice and to prevent future injustice, just as it is morally good to follow any other moral ideal, but to claim that we have an obligation to follow this particular ideal, and to make it part of the very same principle that includes the genuine moral obligation not to treat people unjustly, is to invite confusion, and to cast doubt on the self-evidence of the principle.

5. *REPARATION.* WE SHOULD MAKE AMENDS FOR OUR
   WRONGDOING.

This is an interesting principle and I have nothing in my list of rules, or even in my list of moral ideals, that corresponds to it. It certainly seems, at first sight, like a self-evident

principle, and it is commonly listed as a duty by many. There is no question that if someone unjustly injures another person, or by fraud causes him to lose a significant amount of money, it seems only decent to make amends, or at least to apologize sincerely. Someone who did not feel any obligation to make it up to the person or persons harmed by his wrongdoing does seem to be lacking an essential moral quality. However, Audi claims that making amends for our wrongdoing requires doing something for the person wronged, and I have doubts about whether this is morally required. Suppose that the person you injured, either physically or financially, is now, after your time in jail, more or less recovered. I cannot see that it is better to make amends to that person rather than to devote oneself to helping those in far greater need of help. Audi also probably does not see it as required if that means it takes precedence over helping those in greater need. Actually, I regard the virtue associated with this principle to be a social virtue rather than a moral one. It is the kind of trait of character that makes for a more harmonious and livable society, and I endorse it, but I see it neither as morally required nor as justifying breaking a moral rule.

6. *BENEFICENCE.* WE SHOULD CONTRIBUTE TO THE GOOD
   (ROUGHLY, THE WELL-BEING) OF OTHER PEOPLE.

Here the difference between Audi's concept of morality and mine are quite dramatic. Audi should recognize that this cannot be a moral requirement. Are hermits violating this principle? He should also recognize that although there might be a "theorem" that we should be beneficent toward animals and other beings capable of pain and pleasure, this creates problem if, when there is a conflict, someone follows the theorem rather than the principle. And what of the monk and scholar, do they have to go out and do good for others? Although, if one follows this principle intelligently, it may be morally good to follow it, I cannot see that anyone who does not have a duty stemming from some social role is morally obliged to follow it. Further, unless a distinction is made between relieving pain or promoting pleasure for those who have too little, and promoting pleasure for those who already have enough, this principle does not even seem to me to be a moral principle. I do not see that pastry chefs in gourmet restaurants, or comedians who play in expensive clubs, are doing anything morally good by contributing to the good of other people. I do not claim that they are doing anything bad or wrong, only that their behavior does not, except on a crude utilitarian view, have anything to do with morality.

7. *GRATITUDE.* WE SHOULD EXPRESS GRATITUDE, IN DEED,
   OR AT LEAST IN WORDS OF THANKS, IN A WAY THAT BEFITS
   GOOD THINGS DONE FOR US BY OTHER PEOPLE, WHERE,
   OTHER THINGS BEING EQUAL, OUR OBLIGATION IS STRONGER
   IF WHAT WAS DONE FOR US WAS NOT OWED TO US.

This principle, even more than the principle of reparation, seems to me to be a social virtue rather than a moral one. I am hesitant to say this, because Hobbes, whose views I hold in the highest esteem, places gratitude right after justice in the

list of natural laws. However, I have an interpretation of Hobbes that can explain this prominence, so I am willing to claim that gratitude is not a moral principle, but a social one. It is good to show gratitude, and failure to show it, at least by words of thanks, seems boorish and uncivil. But expressing gratitude does not justify breaking a genuine moral obligation, so not only is it not a moral requirement, it is not even a moral ideal. However, it is a social ideal, and one that it is worth inculcating in our children. I am not against it at all; it is just that I do not think it belongs as one of the prima facie self-evident principles of morality.

8. *SELF-IMPROVEMENT.* WE SHOULD DEVELOP OR AT LEAST
   SUSTAIN OUR DISTINCTIVELY HUMAN CAPACITIES.

This principle illustrates most clearly the different concepts of morality that Audi and I have. On the Kantian view, morality requires treating yourself as an end, and not merely as a means, as well as treating others in that way. Aristotle and Ross agree that there are duties to self, and that morality involves all of one's behavior even when that behavior does not affect anyone else, even indirectly. On my view, and that of Mill and Hobbes, morality is concerned only with the way you treat others. A rational person need not care how others treat themselves; he needs only to care about how they treat him and those for whom he is concerned. And he knows that this is what all other rational persons care about. Of course, he is likely to think more highly of those who do develop their distinctively human capacities, and may not think much of those who do not even sustain their distinctively human capacities, but I would prefer not to call this a moral evaluation.

Audi explicitly claims that, except possibly late in life, or in other unusual circumstances, "morality requires" such development, and even late in life "morality requires" maintaining the level we have. This is certainly not a requirement of any morality that I would endorse. If someone late in life wants to let himself go, and no one is depending on him, then he can let himself go. I see no reason to claim that he is required to sustain his distinctively human capacities, although as someone who values that sort of thing, I would think more highly of a person who at least tried to sustain his distinctively human capacities. Of course, by that I mean his good distinctively human capacities, for there are bad distinctively human capacities that I would not want to see any person develop or sustain, for example, talking nonsense, deceiving oneself, and so on.

More important, listing this principle as one of the ten midlevel axioms or principles of morality illustrates quite clearly Audi's Kantian and Aristotelian account of morality. For him, the last person on earth can still behave morally or immorally. For me, morality does not apply to the last person on earth, although I would certainly think more highly of someone who did not let himself go to pot, even if no one else were affected. However, others thinks differently; Omar Khayyám favors a life of pleasure without worrying about developing one's distinctively human capacities. It is not my preference, but that is because I am a philosopher. One must be careful in taking agreement among philosophers as showing anything but that philosophers agree in some of their preferences. And even those philosophers who prefer developing one's distinctively human capacities do not usually regard this as a moral matter.

9. *ENHANCEMENT AND PRESERVATION OF FREEDOM.*
   WE SHOULD CONTRIBUTE TO INCREASING OR AT LEAST
   PRESERVING THE FREEDOM OF PERSONS, GIVING
   PRIORITY TO REMOVING RESTRAINTS OVER ENHANCING
   OPPORTUNITIES.

Audi realizes that much of what this principle requires is already required by the principles of justice and beneficence. Therefore there is little more to say about this principle. I certainly agree with Audi that removing restraints has moral priority over enhancing opportunities. Removing restraints is what I regard as a moral ideal, as is enhancing opportunities for those who are deprived, that is, who have less freedom or opportunity than anyone living in a given society should have. Enhancing opportunities for those who already have enough or too much is following a utilitarian ideal, like being a gourmet chef for those who already eat well. There is nothing wrong with it, but it is not a matter of morality.

10. *RESPECTFULNESS.* WE SHOULD, IN THE MANNER OF
    OUR RELATIONS WITH OTHER PEOPLE, TREAT THEM
    RESPECTFULLY.

I am inclined to accept this as a genuine moral requirement, but that may be because I think that showing disrespect to others causes, or risks causing, them to feel bad, or encourages others to do things to them that will hurt them or make them feel bad. Showing respect to a person, on my understanding, is treating him or her as a person who has the full protection of morality and is competent to make moral decisions. Morality requires that the moral rules be obeyed impartially with regard to all persons. Failing to show respect is more than regarding a person as not competent to make moral decisions, it is treating a person in a way that shows that you do not care if you do hurt his feelings, that you do not regard him as a moral agent who is impartially protected by morality. If it is not yet a violation of a moral rule, it is taking an attitude that makes unjustifiably breaking a rule toward that person much more likely. I do not think that this principle involves our dealings with dead people, at least not those who have been dead for some time. I think that we do not have to worry about our criticisms of dead philosophers being disrespectful.

Conclusion

I think that Kant did a disservice to the German people and language when he supported the religious view of morality while supposedly taking religion out of morality. Now, many young Germans are struggling to develop a concept of morality that more closely resembles the secular morality of England and America. I do not think it is a good thing to promote this religiously derived German account of morality as the account of morality that should be accepted by English-speaking secularists. Ross, following Aristotle, is not a representative of this English tradition; rather, he embodies the Greek tradition, which is not a moral tradition at all, but of a more

inclusive way of life that includes morality but goes way beyond it. For Aristotle (Ross), morality is a way of acting that all rational persons would adopt in order to live a good life. For Hobbes and Mill, morality is a system that all rational persons would put forward to govern the behavior of all rational persons. These different accounts naturally result in morality having a different content. Hobbes and Mill include in morality only that behavior that has some effect on others, Aristotle (Ross) and Kant include in morality all behavior that they consider to be how rational persons ought to want to behave.

Audi is certainly correct that this Aristotelian tradition can be given a Kantian foundation, and I think that he does an admirable job of doing so. The joining of these two traditions, as Audi has done, certainly strengthens both of them. If one thought that these traditions should be strengthened, he would applaud what Audi has done. However, these traditions do not provide the account of morality that I think appropriate. The way of life that Aristotle (Ross), Kant, and Audi put forward ought to be carefully distinguished from morality. I do not regard combining these two traditions as providing a better account of morality, but as putting forward a way of life that rational persons may or may not want all rational persons to follow. What is taken to be morality is not merely a theoretical matter. A misunderstanding of morality, though not quite the same misunderstanding that Audi is developing, played a large and, I think, unfortunate role in the recent American elections. It is important to understand morality as being limited to that behavior of ours that affects others in ways that either increase their suffering, or risk of suffering, of harms or that decreases that suffering or risk of suffering. Any attempt to make morality into more than that, although the intention may be good, opens the way to serious problems.[3]

Notes

1. See my books, *Common Morality: Deciding What to Do* (New York: Oxford University Press, 2004, ppb. ed. 2007), and *Morality: Its Nature and Justification*, rev. ed. (New York: Oxford University Press, 2005), for a fuller account of moral rules, moral ideals, social ideals, and the relationship between them.

2. *The Philosophical Review* 64, no. 1 (January 1955): 3–32.

3. I am grateful to Mark Timmons for pointing out some mistakes in my original paper and for suggesting ways that I could eliminate these mistakes. I have accepted all of his suggestions, although perhaps not completely in the way that he intended.

Audi's Marriage of Ross and Kant

THOMAS HURKA

As its title suggests, Robert Audi's *The Good in the Right*[1] defends an intuitionist moral view like W. D. Ross's in *The Right and the Good*. Ross was an intuitionist, first, in metaethics, where he held that there are self-evident moral truths that can be known by intuition. But he was also an intuitionist in the different sense used in normative ethics, since he held that there are irreducibly many such truths. Some concern the intrinsic goods, which are in turn plural, so there are prima facie duties to promote pleasure, knowledge, virtue, and just distributions. But others are deontological, requiring one, apart from any consequences, to keep promises, not lie, make reparations, express gratitude, and not injure others.

Audi embraces both these intuitionist views, but in each case with an important addition. Ross sometimes said that if a proposition does not need proof, it is incapable of proof, or cannot be justified inferentially. Audi argues persuasively that this is not so. A proposition that is self-evident, in the sense that understanding it justifies one in believing it, can also be derivable from other self-evident propositions in a way that increases its justification. And he exploits this possibility in his normative ethics. Whereas Ross held that his prima facie duties are underivative, Audi suggests that, while self-evident, they can also be grounded in a more abstract principle. More specifically, he argues in chapter 4 of his book that they can be grounded in Kant's categorical imperative, which he applies primarily in its second, or formula of humanity, version. The result is to transform what Audi calls Rossian intuitionism into Kantian intuitionism, where specific duties about promoting pleasure and keeping promises derive from a more fundamental requirement to respect rational personhood.

I will not challenge Audi's version of metaethical intuitionism, which I think is the most subtle and persuasive yet given. Nor will I question his normative starting point in Ross's theory of prima facie duties, which I find unimpeachable. But I will challenge Audi's Kantian attempt to ground Ross's duties in the categorical imperative. My objections will not reflect a blanket opposition to abstract moral theorizing like that of Philip Stratton-Lake, who has said, in defense of the "naivete" of

Ross's approach, that grounding the prima facie duties in some overarching princi-
ple requires us to think in a way we not only do not think, but cannot persuade our-
selves we *should* think.[2] In my view, this type of objection is telling against the first,
or universal law, formulation of Kant's categorical imperative. As Thomas Hill has
said, to the question 'Why is slavery wrong?', the answer that a world in which
everyone is a slaveowner is logically impossible is not the right answer.[3] But I do not
think the objection applies so clearly to the second formulation of the categorical
imperative. Claims about respect for persons, though not explicit in everyday moral
thought, are close enough to it that they could in principle give a persuasive ration-
ale. And I will later sketch a non-Kantian theory that I think does successfully unify
Ross's duties to some degree. Instead, my objection to Audi's Kantian unification
will be simply that it does not succeed. Even in its best formulation, the categorical
imperative does not yield the specific duties it is meant to, or does not unless sup-
plemented by claims that assume what it is meant to explain. It is worth noting that
this was Ross's own view. In his 1952 book *Kant's Ethical Theory*, he starts, as
Stratton-Lake might, by expressing a general skepticism about whether there can be
a "general criterion of rightness," but continues, "We must not prejudge the case.
We must study Kant's attempt on its merits, and see whether his criterion will do the
work he claims it will do."[4] His conclusion is that the attempt fails.

Audi intends his Kantian ideas to play four roles. Ross held that when two prima
facie duties conflict we can sometimes say that one is stronger, but only on the basis
of an immediate intuitive judgment "for which no logical basis can be offered."[5] Audi
thinks the categorical imperative can explain some of these comparative judgments
about the strength of duties, making for a theory that, in his terminology, is more
"normatively complete" (85), because it justifies claims that Rossian intuitionism can
only assert. Ross also acknowledged that we sometimes cannot weigh conflicting
duties or say that either is stronger; past a point our intuitions run out. Here Audi
thinks the categorical imperative can decide some of these cases and so make Ross's
theory more "epistemically complete" (86). He also finds Ross's theory inadequate
because it provides no unifying ground for its disparate duties, instead treating the
requirements to promote the good, keep promises, and so on as independent. Again
the categorical imperative can provide that unifying ground. And in doing so it can
give the individual duties a deeper rationale, so we see why, for example, it is wrong
to break promises when we see that that is one way of not respecting others.

Audi starts his defense of these Kantian ideas with the issue of normative com-
pleteness, and does so by considering the familiar case where it seems right to break
a promise to meet a friend for lunch if that is necessary to save someone else from
serious harm. Ross said the comparative judgment in this case cannot be given a
"logical basis," but Audi thinks it can be justified, first, by the first formulation of the
categorical imperative, which he takes to ask whether we could *rationally* will the
universalization of our maxim. Would breaking one's promise in the case in ques-
tion offend a reasonable promisee, he asks. His answer is that it would not: one's
friend, if rational, would accept one's choice (91–92). He also thinks the judgment
can be justified by the second formulation: keeping one's promise would fail to treat
the person facing harm as an end, since it would put her in danger "for a less than
weighty reason" (92).

But both these arguments invite the objection that they assume what they are meant to explain. If a reasonable friend would accept one's choice, that is only because he would see that in this case the moral duty to prevent harm is stronger than the duty to keep one's promise, which is just Ross's claim. And if the reason to keep the promise is "less weighty," it is for the same reason. Audi anticipates this objection and responds that in speaking of "reasonableness" and "weights" he is not using these terms in specifically moral but only in broader senses, so his arguments do not beg crucial questions (221n25). But it is hard to see how this response is persuasive. Ross himself would not have accepted it, since he did not distinguish between moral and nonmoral reasons. Staying close here to Kant, Ross held that all categorical imperatives, or all imperatives that do not refer implicitly to a person's desires, are moral imperatives. Audi seems to take the contrary view that moral reasons are a specific subset of reasons, including categorical ones, and are perhaps arrived at by constructions over the full set of reasons, as in T. M. Scanlon's contractualism.[6] Now, I happen to share Ross's view that multiplying types of reasons adds complexity but no illumination to normative ethics. But it is not necessary to argue this point here. For we can assume that whatever other reasons might bear on our case of breaking a promise to save someone from danger are neutral about it. For example, if there are personal or prudential reasons derived from one's own well-being, they are equally balanced between the two options. Then what makes one option right can be only the intrinsic weights of the reasons to save the person and to keep the promise, which I assume are categorical. And how much have we gained if, instead of saying that the moral reason to save the person simply and underivatively outweighs the moral reason to keep the promise, we say this is because a nonmoral reason to save the person simply and underivatively outweighs a nonmoral reason not to? While not strictly circular, this explanation is surely too close to what it is meant to explain.

It is equally hard to see how the Kantian ideas can help with the issue of epistemic completeness, or enable Kantian intuitionism to decide more cases of conflict between duties. How can it be any easier to make underivative judgments about the weights of nonmoral reasons than to make similar judgments about moral duties? And this second problem generalizes. The second formulation of the categorical imperative yields at least two basic duties: to treat persons as ends and not to treat them just as means, where the second or negative duty is stronger. If successfully established, this priority claim can do some explanatory work, showing why specific negative duties such as not to harm, break promises, or lie can all outweigh positive duties. But how much epistemic work can it do? Will it be easier for us to decide whether the duty to keep a specific promise outweighs a specific duty to aid if we ask how much, in general, the duty not to treat as means outweighs the duty to treat as ends? That seems dubious. If anything, the more abstract question seems harder to answer, one we can address only by deciding specific cases of conflict between negative and positive duties.

Similar difficulties arise from an observation of Audi's that, if it has not been made before in the Kantian literature, is an acute contribution: that Kant's distinction between treating people as ends and treating them only as means does not exhaust the morally relevant possibilities. Consider an example of Audi's in which I talk loudly while you are giving a lecture (103). I am not treating you as a means,

since I am not using you in any way, but I am also not just failing to treat you as an end, as I would be if I merely failed to give you some benefit. My behavior is worse than that. Or consider a related case where I do treat you at least to some extent as an end. If to save five people from drowning I drive over a bridge where you are lying immovable on the ground, I am again not treating you as a means. But I am treating you as an end if I weigh your interests seriously, would avoid driving over you if I could, and would not drive over you to prevent lesser harms. My attitude to you is therefore the same as to another person whom I decide not to save because saving the five is more important. But many would say that while it is right to let one drown in order to save five, it is not right but wrong to drive over one to save five. So again there is a ground of wrongness distinct from treating as a means.

As these examples show, Kant's talk of treating as a means is close to the double-effect distinction between intending harm as an end or means and merely foreseeing it, which many deontologists use in formulating their constraints. But many, including Ross, also use, either in addition or instead, the distinction between doing harm and merely allowing it, and it is this distinction that explains why talking during your lecture and driving over you are worse than merely allowing similar harms to come about. Recognizing this, Audi proposes to extend the categorical imperative to forbid actions that, while not treating others as means, nonetheless treat them with disrespect or in a way inconsistent with their dignity. The question, however, is whether our grasp of the concepts of respect and dignity is sufficiently independent of the relevant judgments to explain and help us reach them. Imagine that, as I prepare to drive over you on the bridge, you say that doing so would be wrong because it would be inconsistent with your dignity. If I respond that my action would be perfectly consistent with your dignity, since it would treat you as an end in the same way I treat someone whom I choose not to save in order to save five, how can you reply? Do you not have to say that actively causing you harm is in itself more disrespectful than merely allowing harm to you, and is that not too close to the judgment it is supposed to yield to be genuinely explanatory? The *explanans* again assumes something too close to the *explanandum*.

There is a familiar general thesis in this area, which has recently been stated by Richard Arneson in relation to distributive justice.[7] It says that the concept of respect for persons has no independent content and therefore cannot be used to identify or justify claims about moral duty. To treat a person with respect is to act toward her in accordance with the moral principles that are best supported by reasons, so to know what respect consists in we must first know what those principles are. I have suggested that Audi's arguments about conflicts of duty tend to support this thesis, but there are other claims of his to consider.

Audi claims, plausibly, that moral views are unreasonably demanding if they require people to work constantly at promoting others' good. Consequentialism notoriously does this, as, in fact, does Ross's theory. Ross held that once people have fulfilled their other prima facie duties, they are required to maximize good impartially.[8] But this was a surprising view for him to take, given his general aim of formulating a theory consistent with commonsense morality. It would have fit that aim better to give people options to produce outcomes that are somewhat less than the best, because of a general agent-relative permission to count their own good somewhat more. And

given the general naïveté of his approach, he would have added that permission as an underived element of his overall view, so it is a brute fact, for which no deeper explanation can be given, that people are permitted to favor their own good.

Audi, by contrast, wants to give the permission a deeper rationale, and starts by arguing that submitting to an overriding duty to maximize the good of others would involve treating oneself merely as a means, or close to it (97). But how can this be? If in response to an overseas disaster I ask Bill Gates to donate his fortune to humanitarian relief and he does, I do not treat him merely as a means, since I act only with his consent. And acting only as another consents to is normally thought sufficient for treating him as an end. But if I voluntarily contribute my own fortune to humanitarian relief, am I not implicitly consenting to do so, and does that not answer any charge of treating myself just as a means? Does consent not cancel disrespect in this case too? Audi thinks it would if contributing the fortune were merely a voluntary ideal implying no universal obligation, but not if the beneficence is a moral duty. But how can the specific content of my motive, and in particular whether I am acting as I think I ought, affect whether my act is voluntary in the way that amounts to consent? How is a free choice to do what one thinks is required any less free? And we can press this question by asking about a case involving the less demanding duty of beneficence that Audi accepts. Imagine that in Peter Singer's famous example I save a child from drowning in the pond at the cost of dirtying my suit, and do so because I think that is my duty. Audi presumably does not think I treat myself merely as a means in this case, though someone else would treat me merely as a means if he forced me to save the child regardless of my consent. But if my voluntarily fulfilling what I take to be my duty answers the charge of treating myself merely as a means in Singer's case, why not also in the case where I act on a more demanding duty to give my fortune to overseas relief? Both cases involve similarly voluntary action on what I take to be moral duty. Audi could say the difference is that in Singer's case I act on a duty that is reasonable, while in the relief case I act on one that is unreasonably demanding. But then we would be back in the familiar situation where a Kantian idea, this time about treating oneself as means, assumes what it is meant to justify.

Audi also gives a positive rationale for limiting the duty of beneficence by options, one that appeals to our dignity as persons, in particular, persons with autonomy (99). But it is hard to see how Kantian dignity can yield the particular structure that underlies options. That structure, recall, is agent-relative, allowing each person to give extra weight just to his own good. He is permitted to save his own life rather than save those of five strangers, but not to save one stranger rather than five: the permission concerns only himself. But Kantian dignity is agent-neutral, possessed equally by everyone and with equal importance in everyone. And it must be agent-neutral if, as Kant insists, such dignity is unconditional and therefore not dependent on relations to anything outside the person. But how can a dignity that I possess and can possess no more than anyone else ground a permission to care more about my good than about other people's? How can a value that I possess equally with all others permit me to treat myself unequally? There seems a fundamental mismatch between the agent-neutrality of the proposed moral ground and the agent-relativity it is meant to justify.

A similar problem arises within the account of beneficence. Ross happens to have held that beneficence requires us to promote the good of all others equally.

But this view is in some tension with his argument, which Audi cites several times, that impartial consequentialism ignores "the highly personal character of duty."[9] It would be more consistent with this argument, as well as with Ross's general aim of mirroring commonsense morality, to adopt the view C. D. Broad called "self-referential altruism," which gives each person a stronger duty to promote the good of those close to him, such as his family and friends.[10] And Ross did include the relations of child to parent and friend to friend among those consequentialism objectionably ignores.[11] So let us imagine an extended Rossian theory whose duty of beneficence embraces self-referential altruism. It is again hard to see how the resulting agent-relativity could be explained on Kantian grounds. If the basis of the duty of beneficence is a dignity my child does and can possess no more than anyone else, how can my duty to promote her good be any stronger?

And there are more fundamental difficulties with a Kantian account of beneficence. In a later chapter Audi claims, as others including Ross have, that the value-properties of goodness and evil attach primarily to states of affairs. But that was surely not Kant's view. He held that the end whose value grounds the categorical imperative is not "one to be effected," as states of affairs are, but is "independent," so the fundamental value is located in persons rather than in states of affairs. Alan Donagan took this to be a central feature of Kant's ethics, and it is defended by contemporary neo-Kantians such as Elizabeth Anderson, Scanlon, and Stephen Darwall.[12] So to examine a Kantian justification of the duty of beneficence, we must imagine it starting with a fundamental value in persons. How would such a justification go?

Audi seems to think it would be straightforward. He simply asserts, as Donagan also did, that "to treat someone as an end is above all for the relevant acts toward the person . . . to be motivated by a concern with the good, say the physical or psychological well-being, of the person for its own sake" (92). But how exactly does this follow? The value one is responding to, recall, is possessed equally by all persons, and equally by them in all circumstances. It is no more present when a person is happy than when he is miserable, or virtuous rather than vicious; even the worst sinner retains his fundamental dignity. And for this reason the property it is based on must likewise be possessed equally in all circumstances. It cannot be just rationality, which some realize to a higher degree than others. It must instead be something like the property of being rational to at least some minimal degree m, which is equally present no matter how much greater than m one's exercise of rationality is. And how can a value that is unchanging, and based on a property that is likewise unchanging, make it right to favor some changes in a person rather than others, say, ones that make him happier rather than more miserable? How can a value that is independent of any states of affairs lead us to promote some states and not others?

Ross himself was baffled on this point. After noting that Kant's fundamental value was not in states of affairs but was what his translation called a "self-subsistent" end in persons, he wrote:

> The notion of self-subsistent ends is nothing but an embarrassment to Kant. When he remembers that men are (according to his use of words) self-subsistent ends, he interprets "treating them as ends" merely as not interfering with them. . . . At other times, e.g., in dealing with [his third and fourth examples] he interprets "treating men as ends" as the positive "advancement of humanity"—a stimulating ideal, but

> one that would be unmeaning if humanity were a self-subsisting end. The plain
> fact is that in strictness man is not an end at all, and the description of him as an
> objective and at the same time self-subsistent end can be understood only if we take
> this as a way of expressing the fact that there is something which can be realized in
> any man and is worthy of being an object of desire to every man.[13]

In other words, if values are ultimately in states of affairs, there can be duties to promote some states rather than others. But if the fundamental value is in persons and is independent of states of affairs, it cannot intelligibly ground a duty of beneficence.

Even apart from this difficulty, it is unclear how the Kantian value of persons could ground the full content of Ross's duty of beneficence, which, recall, concerns a plurality of intrinsic goods. How could that one value ground duties to promote states as diverse as pleasure, knowledge, and virtue? To cite just one difficulty, virtue or the morally good will, as Kant and Ross understand it, is good independent of relations to anything outside it, so it can "sparkle like a jewel in its own right." But knowledge essentially involves a relation of correspondence between one's beliefs and reality. How can one fundamental value yield both an essentially nonrelational and an essentially relational good?

But enough criticism. While I have questioned Audi's attempt to ground Ross's various duties in the categorical imperative, I have no objection to his general project of justifying self-evident moral principles in more abstract ones that deepen their justification, and believe there is a rival view that achieves this aim better, which has been principally defended by Germain Grisez, John Finnis, and Joseph Boyle, who call it "Thomist."[14] I think this label is unfortunate, for a variety of reasons; Grisez et al. also build into the view certain traditional Catholic assumptions that are not essential to it. While retaining the name "Thomist," I will separate it from those assumptions.

The Thomist view starts by holding, with consequentialists and Ross but against Kant, that intrinsic values always reside in states of affairs. It also holds, again with Ross, that there are a plurality of such goods, including pleasure, knowledge, and virtue. Grisez et al. include among the ultimate goods physical life, which they use to underwrite the traditional Catholic bans on suicide and euthanasia. But many will deny that life as such has value, and the view can easily abandon that claim; it can also, following a suggestion of Audi's for Ross (178–79), add freedom, or choice from a wide range of options, to the list of goods. But whatever its exact content, the Thomist starting point is a plural list of good states of affairs.

The view then claims that for each good there are two moral duties: a positive duty to promote it, and a negative one not to destroy or choose directly against it, where the negative duty is stronger than the positive. Grisez et al. echo Kant in holding that the two duties flow from a single attitude of valuing the good, with the second or negative one required because different goods are incommensurable, so the idea of maximizing overall value is incoherent. I see no special difficulty about commensurating different goods and will also drop this feature of the view; instead, I will see the duties to promote and not destroy the good as irreducibly separate. And I will take the greater strength of the negative duty to imply that destroying or choosing against a good can be wrong even though, what incommensurability would seem to preclude, its overall outcome is best.

The view then applies its two duties to each of the different goods. If pleasure is good and pain evil, there is a positive duty to promote pleasure and prevent pain but an even stronger duty not to cause pain directly, so that even if torturing one person would stop two others from being tortured, the torturing is wrong. If knowledge is good, there is a positive duty to promote knowledge in others, but an even stronger duty not to act against knowledge, as one does if one lies to or, more generally, intentionally deceives them, since in that case one's aim is false belief. And if freedom is good, there is a duty to increase others' range of options, but an even stronger duty not to reduce their options by coercing them. And so on.

This Thomist view adds several elements of unity to a Rossian deontology. First, it relates all moral duties to intrinsic goods and evils, understood as being located in states of affairs. This should be agreeable to Audi, since he too locates values in states of affairs. Second, it takes these duties to come in pairs, a positive and a negative one for each good, with the negative duty in each pair being stronger. So there is a unifying pattern across the various duties, which can be divided either by positive or negative, into two categories, or by the goods they concern, into as many pairs as there are fundamental values.

But the view has nothing like the grandiose unifying ambitions of Kant's theory; on the contrary, it leaves several irreducible pluralities. The first is between the various intrinsic goods, such as pleasure, knowledge, and virtue. They are irreducibly distinct rather than somehow unified in a single value of rational personhood. The second is between the positive duty to promote goods and the negative one not to destroy them; in my presentation these are also irreducibly distinct rather than expressing some single attitude to value. Nor does the Thomist view claim to increase the epistemic completeness of a Rossian theory. Imagine that we are trying to decide whether we may cause one innocent person pain as a means of saving other people from pain. General Thomist ideas tell us that causing the pain is not permitted just to secure a small decrease in overall pain, but beyond that these ideas do not help. We cannot solve our problem by reflecting in the abstract about the relative strengths of positive and negative duties; we must look at particular cases of causing pain in order to prevent it and try to reach intuitive judgments about them.

But these incompletenesses do not mean the Thomist view has no point. On the contrary, it both unifies the Rossian duties to a considerable degree and illuminates aspects of these duties that a Kantian approach cannot. Let me mention one such aspect. I have said that a Kantian view, which starts from an agent-neutral value of personhood, has difficulty allowing what Broad called self-referential altruism, the view that we have stronger duties to promote the good of people close to us. The Thomist view can justify this view if it says that from each person's point of view, his wife's or child's pleasure is a greater good than the similar pleasure of a stranger, yielding a stronger duty to promote it. But common sense also builds a self-referential element into its deontological constraints: while it is wrong to lie or break promises to anyone, it is especially wrong to do so to an intimate. If you lie to your wife, she can protest not only "How could you lie?" but "How could you lie to *me*?" And the Thomist view can give a unifying explanation of these two self-referentialities. If constraints are always duties not to destroy goods, they are presumably stronger when the goods are greater; thus, the duty not to cause intense pain is stronger than the duty not to cause mild

pain. But then if my wife's knowledge is a greater good from my point of view, the constraint against acting against it, as I do if I lie to her, is also stronger. With a common foundation in an agent-relative value, my positive and negative duties concerning my wife are both more stringent than comparable duties to strangers.

Let me conclude. In the perhaps unfortunate manner of philosophers, I have said least about the part of Audi's *The Good in the Right* that I find most persuasive: its account of self-evident moral truths and how we know them. Instead, I have concentrated on claims I find less persuasive and that happen to be closer to my interests. As I see it, in trying to ground Ross's prima facie duties in Kant's categorical imperative, Audi is proposing a marriage between one of the greatest moral philosophers ever to write about the subject and another whose approach to normative questions is essentially hopeless. The great moral philosopher is Ross; the one with the hopeless project is Kant. And I have protested against this marriage both because Kant is not good enough for Ross—he brings nothing useful to the partnership—and because someone else, namely Aquinas, is better. Ross's prima facie duties can indeed be given a deeper grounding, as Audi's account of self-evidence allows, but it is the more modest Thomist account rather than the grandiose fantasies found in Kant.

Notes

1. Robert Audi, *The Good in the Right: A Theory of Intuition and Intrinsic Value* (Princeton, N.J.: Princeton University Press, 2004). Parenthetical page references are to this book.

2. Philip Stratton-Lake, "Introduction," in *Ethical Intuitionism: Re-evaluations*, ed. Philip Stratton-Lake (Oxford: Clarendon, 2002), 25–26.

3. Thomas E. Hill Jr., "Kantian Normative Ethics," in *Oxford Handbook of Ethical Theory*, ed. David Copp (New York: Oxford University Press, 2006), 488.

4. Sir David Ross, *Kant's Ethical Theory* (Oxford: Clarendon, 1952), 28–29.

5. W. D. Ross, *The Right and the Good* (Oxford: Clarendon, 1930), 23, 30–32.

6. T. M. Scanlon, *What We Owe to Each Other* (Cambridge, Mass.: Harvard University Press, 1998).

7. Richard J. Arneson, "Luck Egalitarianism and Prioritarianism," *Ethics* 110 (2000): 339–49, esp. 344.

8. Ross, *The Right and the Good*, 39.

9. Ibid., 19, 22.

10. C. D. Broad, "Self and Others," in *Broad's Critical Essays in Moral Philosophy*, ed. David Cheney (London: George Allen and Unwin, 1971), 262–82.

11. Ross, *The Right and the Good*, 19.

12. Alan Donagan, *The Theory of Morality* (Chicago: University of Chicago Press, 1977); Elizabeth Anderson, *Value in Ethics and Economics* (Cambridge, Mass.: Harvard University Press, 1993); Scanlon, *What We Owe to Each Other*; Stephen Darwall, *Welfare and Rational Care* (Princeton, N.J.: Princeton University Press, 2002).

13. Ross, *Kant's Ethical Theory*, 51–52.

14. See, e.g., John Finnis, *Natural Law and Natural Rights* (Oxford: Clarendon, 1980).

Accounting for Duties

A Critical Assessment of Robert Audi's
The Good in the Right

CANDACE VOGLER

Introduction

Robert Audi's *The Good in the Right* undertakes the magisterial work of reviving the intuitionism of W. D. Ross, rescuing Ross from the overlapping shadows of Henry Sidgwick, G. E. Moore, and, to a lesser extent, H. A. Prichard, marrying Ross to Kant, and so working to produce "a full-scale moral philosophy providing both an account of moral principles and judgments—a metaethical account—and a set of basic moral standards" that might be employed in moral reasoning.[1] The book is magnificent in ambition and impressive in detail.

Nevertheless, readers may think that something has gone missing somewhere. Those who have struggled with Kant may find Kant missing from Audi's discussion of Kantian ethics. And Ross qua great English translator of Aristotle is not entirely apparent in Audi's efforts to produce a systematic intuitionism either.[2] Explaining his decision to join Kant to Ross, Audi remarks:

> The task of integration is challenging. For whereas Ross stressed intuitive induction as our route from understanding concrete instances of duty to apprehending abstract principles, Kant resoundingly asserted that one could not do morality a worse disservice than to derive it from examples.[3]

I agree that Kant and Ross make strange bedfellows, but *not* because Kant rejected example-based ethics and Ross drew from examples. I take it that, whatever their differences, it is false to characterize Kant as embracing "a top-down conception of the determination of moral obligation,"[4] whereas Ross gives us a "bottom-up theory."[5] Rather, powerful Kant is best read as a "bottom-up" man, the chief weakness in Ross traces to the "top-down" cast of Anglophone moral philosophy in the longer twentieth century (running from Sidgwick to the present), and this has nothing to do with anyone's attitudes toward examples. It has, rather, to do with the place of principles in the accounts.

In the relevant sense of principle, Audi is a man of principle, Ross is drawn that way, and Kant is not. While Audi is not setting out to do history of moral philosophy, missing

one strand of Ross and the thrust of Kant prevents Audi from giving due attention to the ground of moral judgment as that ground shows itself in Kant and hovers in Ross. Since grounding moral judgment *is* Audi's central concern, a little history is not irrelevant.

I will begin by discussing principles, setting up a contrast between Audi and Kant. In the relevant sense, the categorical imperative is not a principle. But then, prima facie duties are not necessarily best captured by principles either. In this spirit, I will turn to Ross on the social ground of duty. For some kinds of moral judgments, grounding might best be sought between Ross's social world and Kant's account of the will—the bits missing in Audi.[6]

Axiological Integration

Audi describes his deployment of Kant as an axiological integration. The axioms in question are Rossian principles. Unlike Kant's formulations of the categorical imperative, Rossian principles are intuitive and self-evident, and Audi has done a tremendous service in giving an account of self-evidence and of intuitions that is clear, precise, and cogent. My quarrel concerns the place of principles in the account. Principles are what bring together Kant and Ross for Audi.

Principles, on Audi's account, have at least the following features:

1. The content of principles can be expressed in sentences such as "*Pacta sunt servanda*" (equivalently, "Promises are to be kept").

2. Principles can be deployed in practical reasoning as premises, can emerge as inductive conclusions from experience grounded in intuitive grasp of prima facie duties, or can express a summary judgment of reflection, where reflection draws upon evidence and inference, but is not helpfully assimilated to either (reflection, as Audi argues persuasively, is neither strictly evidential nor plainly inferential).

3. All principles are general in several senses—indefinitely many agents may produce the same principle when asked why they are doing/mean to do/have done such-and-such; every agent who produces, in all sincerity and in clear understanding of the content, the relevant principle under the relevant prompt is giving exactly the same reason for acting as every other agent; the single agent adept at the business of, say, promising can utter the "promises are to be kept" sentence to give her reason for acting every time she does her word, just because one keeps promises.

Audi also takes it to mean that principles can serve as sources of reasons for acting. Here, we come in for trouble, for our next feature is this:

4. Some principles, however, are *differently* general. While "Securing the judicial condemnation of the innocent is wrong" expresses a basic general principle, "Act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end" and "Act only on that maxim through which you can at the same time will that it should become a universal law" express higher-level principles (the metaphoric height of formulations of the categorical imperative motivates Audi's insistence that Kant works from the top down, while Ross works from the bottom up).

Notice that the categorical imperative formulae are unlike the other principles in at least two closely related respects. First, it is not clear *what* they enjoin. Second, partly because it is not clear what they enjoin, and partly because they will, ideally, underwrite everything that a finite, dependent rational being does, they cannot operate as sources of reasons for acting in any ordinary sense.

This is clearest for the universal law formula. Imagine yourself trying to operate as though you are legislating and enacting universal law every time you lick a stamp in order to mail a payment to a creditor, having determined that this has first dibs on your practical attention today. Imagine, as Kant insists, that you are bound to consider yourself a source of universal law because you are a finite, dependent rational being (i.e., the kind of thing that cannot will universal law into effect and the only kind of thing that can be the addressee of an imperative).[7] In these terms, it looks as though reasoning from the most general of principles necessarily requires something worse than delusion, that the principle is no source of reasons at all.

Turning to the second formula, notice that "end" can't suggest a goal or an aim, unless we're thinking that we could bring humanity into being, or make it possible every time we act well (only marginally less strange than thinking ourselves authors of universal law, but not obviously applicable beyond specifically procreative practical reasoning). Will it help to say we have in mind ends-in-themselves rather than ends-in-the-sense-stopping-places, goals, or targets? No. Ends-in-themselves include friendship, health, pleasure, and justice—things that we have it in our power to make last, to make happen, or to make possible. Humanity is not an end in that sense. For Kant, rational actions are expressive of humanity without being productive of it. The idiom of ends, and of ends-in-themselves, is instead productive. So is ordinary practical reasoning, suggesting that whatever we are supposed to be doing with this formula, it isn't normal-person practical reasoning.

Worse, we are required to treat humanity never simply as a means, but always at the same time as an end, suggesting that, ideally, humanity will be both end and means in everything we do.[8] What could possibly count as means and at the same time end of moving a piano, making a paper hat, manning the barricades, ordering a bagel, cracking jokes with a friend, voting, formulating a policy, gathering actuarial data, piloting the blind man across the busy street, or your favorite intentional act-type. On the face of it, these have neither ends nor means in common. Nevertheless, humanity somehow names the red thread running through all of them alike, endwise and meanswise.

Kant links the formulas to maxims. Maxims give the content of actions-in-prospect in general terms, and part of the difficulty in making out his use of "ends" and "ends-in-themselves" has to do with determining how these special, non-endlike ends are to be understood with respect to actions-in-prospect. Audi rightly distances his account from the notorious maxim-specification problem. For Audi, the formulas find their proper home in evaluating lower-level principles. The formula of humanity has direct practical importance: it provides a test for permissibility and sets ends for action (if not specific ends, then at least "directions in which to seek guidance").[9] Audi offers the following gloss on treating persons as ends (having quietly shifted our focus from humanity to its bearers):

First, to treat someone as an end is above all for the relevant acts toward the person (the 'treatment') to be motivated by a concern with the good, say the physical or psychological well-being, of the person, for its own sake. Second, to treat someone merely as a means is for the relevant acts toward the person to be motivated only by instrumental concerns and accompanied by an indisposition to acquire any non-instrumental motivation toward the person.[10]

By Audi's lights I discern how to reconcile apparently conflicting duties with reference to Kant's formulas. This will involve reminding myself that a person is never simply a means, but always at the same time an end—where "end" and "means" are, it seems, homonyms for our terms *end* and *means*, and what is actually required of me is, among other things, that I be moved to act from my concern for persons' good.[11] This can't be what Kant was after, even though turning to formulations of the categorical imperative in order to find order among duties looks to be right. Kant was trying to articulate the common element across perfect and imperfect duties to self and others.[12] If the common element is the moral law, and if the delicate business of articulating its content in the formulas went well, then turning to the formulations might help to adjudicate apparent conflict. But the common element that Kant struggled to articulate is not a further, higher-order principle at all, in Audi's sense of the term. It certainly isn't a principle in the way that *Pacta sunt servanda* might be a principle. Rather, Kant hoped that the formulas bore the kind of relation to "Do not lie," "Help those in need," and "It is never permissible to take murderous means to your end" that *if* p, *then* q bears to "Where there's smoke, there's fire," or, more to the point, that a mathematical formula describing the right triangle bears to right triangles. In a footnote to the preface of the *Second Critique* Kant wrote:

> A critic who wished to say something against [the *Groundwork*] really did better than he intended when he said that there was no new principle of morality in it but only a new formula. Who would wish to introduce a new principle of morality and, as it were, be its inventor, as if the world had hitherto been ignorant of what duty is or had been thoroughly wrong about it? Those who know what a formula means to a mathematician, in determining what is to be done in solving a problem without letting him go astray, will not regard a formula which will do this for all duties as something insignificant and unnecessary.[13]

The sentence $a^2 + b^2 = c^2$ is not a *more general* right triangle, or a *higher-order* right triangle. It is not a right triangle at all. It expresses a common element in the many right triangles. Working out a problem about rectangles, one might deploy right triangle formulas directly. But the formulas are different from the figures, and that difference is a difference in kind.

What Kant meant by grounding morals was articulating that common element of all duties that at the same time reveals the constitutive principle of the finite, dependent, rational will—the one kind of thing in all of creation capable of standing as the addressee of an imperative. It is very hard to understand Kant's project. But notice that it ought to come as a surprise if it turns out that understanding what an imperfect duty to myself has in common with a perfect duty to a stranger will answer to the Sidgwickian demand that moral theory give complete normative guidance and provide for a

complete epistemic grasp on duties. Suppose a complete articulation of the nature of the right triangle in three formulas. This will help our work with rectangles, will prevent us from going astray, and could direct our attention to likely problems. It will not provide everything we need for engineering purposes and, actually, Kant's mature work in substantive ethics leaves a lot of room for casuistry. If he thought he had discovered a permissibility/requirement test in the *Groundwork of the Metaphysic of Morals* one would expect him to use it in the *Metaphysics of Morals* proper. He doesn't. Since Kant was not shy about displaying his accomplishments, this should tell us something.

Namely, it should tell us that Kant was more interested in discerning the nature of the finite, dependent, rational will than in what Audi calls "principles." The formulas are meant to articulate the moral law, which turns out to be key to rational nature, since, Kant argues, the possibility of being the addressee of any imperative—hypothetical or categorical—rests in the possibility of acting from and for the sake of the moral law. This determines the nature of the finite, dependent, rational being.

The structure of Kant's moral theory is not like Audi's. In Kant, when we turn our attention to the addressee of imperatives, struggle to articulate their content, and to discuss the two broad classifications of kinds of reasons for acting—motives (the two are duty and self-love)—our topic is the nature of the finite, dependent rational will, its capacities, and its proclivities. When we turn our attention to specific duties or principles, we are instead concerned with how this nature expresses itself in action. But since everything that the thing does on purpose expresses its nature in action, our account of specific duties or principles in complex cases, if we have any (Kant stops well short of giving us such a thing), will come to rest in the account of the will. The will is at the bottom of all the activity.

In Audi, by contrast, we have a modified Sidgwickian two-tiered model, where formulations of the categorical imperative operate as higher-order principles that can be used to orient reflection about what is to be done, and to set tests for permissibility. The tests take Rossian lower-level axioms and subsidiary rules as inputs and give concrete guidance. Given the strangeness of the formulas, however, treating the categorical imperative as of a piece with other principles involves trying to use the formulas in a way that they ought not to be used, and, at the same time, risks losing sight of the very different kind of metaethical project that informs Kant's work.

For Kant, there can be no question why quick-witted, healthy, adult humans take an interest in ethical matters. The human is one of the species of creature essentially characterized by its finite, dependent, rational will. Kant operates as though the finite, dependent rational being named a species in roughly the way that the horse or the man does in Aristotle. For Kant, it is *in* us, actually, to be tuned to ethics. We are some among the creatures for whom the ethical is at once, essentially, a challenge and a possibility. For Kant, this is a more basic fact about us than that we have arms, lungs, and mothers. For Kant, strictly, it is cosmically impossible that there should be a finite, dependent, rational being that was not accountable to the moral law by its own lights. The mind boggles. But that is, I take it, the view.

In Audi's system, by contrast, interest in ethics might be statistically more common than interest in lyric poetry or physics, but it is hard to see how the ethical is anything other than one among the many fields of engagement that might attract our interest.

His theory aims at systematizing this field. The field itself is popular, but it is unclear how its popularity is anything but contingent. Apart from our presumably equally contingent interest in having the most comprehensive treatment of a field concern at our disposal—in order to solve puzzles, or to have more satisfying modes of reflection, or because we are raised to feel somehow guilty or inadequate if we don't worry about ethics and our worry will be less acute if we have a system—it is unclear what kind of gain this is for us. In this sense, Audi's principles are fundamentally external to the creatures whose lives can be regulated by principles. Kantian imperatives are not. Kant may have been wrong, but I think that the character of his engagement with moral theory was not. And, for me, Ross is closer to Kant than to Audi on this score.

Ross on Social Relations

Ross was explicitly worried about it seeming that his theory turned on things that were external to us. Accordingly, when he introduced the term *prima facie duty*, he hedged it with caveats. *Claim* (the term suggested by Prichard) was preferable because, for example, in the case of breaking a trivial promise in order to help the victim of an accident, the accident victim and the promisee each presents a claim on me.[14] This "is an objective fact in the nature of the situation," rather than some appearance "which may turn out to be illusory."[15] The risk of losing the actuality of prima facie duties was the chief disadvantage of prima facie. But Ross needed a term that did two other jobs. First, he wanted to capture the fact that we find ourselves subject to claims, and, second, *claim* doesn't cover duties to cultivate character.[16] My character makes no claim on me. *Prima facie* is meant to capture the ethical actuality that circumstances show to quick-witted, mature humans as such, straight off.

Ross pictures a social world in which persons are bound together after the fashion of claim and subject-to-claim by duties of fidelity, reparation, gratitude, justice, beneficence, and nonmaleficence, a world in which the quick-witted must cultivate intelligence and virtue. The facts of the situation that fix my duties include my previous acts, other people's acts, unjust circumstances, the existence of others I might injure or assist, and my own imperfection.[17] Ross's is also a world where the fact that one being has a right implies that another has a duty (and this is so even if the bearer of the right is a child, or a nonhuman animal incapable of owing anything in return).[18] In short, it is a world in which "the nature and relations" of quick-witted, able-bodied, adult human beings set the terms for ethics.

One might have expected that any simplification here could have been handled by treating duties of self-improvement and duties to the subrational as derivative. If, qua quick-witted and able-bodied human being, I find myself in a network of reciprocal claims, if the fact that we are here is enough to set all this in motion, then self-improvement might be explained by noticing that I need to be capable of doing my part. That the human is rational and political might explain the social ground of adult reciprocal duties. Other aspects of human nature—that preparing the young to enter into relations of this kind requires social labor—might be handled from this point of view as well. Perhaps our relations with members of nonhuman species are haunted by the fact that we can so easily ride roughshod over them (as are relations between quick-witted,

able-bodied, adult humans and their relatively defenseless fellows), and these duties belong to the sense in which we are essentially ethically aware and ethically challenged.

In this spirit, I wish that Ross had stuck with Prichard's idea and set about illuminating the relations that ground corresponding and reciprocal claims among adult persons qua persons, working from there to develop some account of what we owe ourselves, our own young, our cognitively disabled adults, and members of other species. Ross didn't; nor did he emphasize an Aristotelian backdrop in the ways I am urging. But it seems pretty clear that Ross's view was consonant with a vast Aristotelian background.

For example, criticizing the thought that I might routinely handle the challenge of ethical conduct by moving from general principles to particular duties, Ross wrote:

> [I]t will not do to make our perception of particular duties essentially inference from general principles. For it may . . . be taken for granted that man was a practical being before he became a theoretical one, and that in particular he answered . . . the question how he ought to behave in particular circumstances, before he engaged in general speculation on the principles of duty.[19]

I think Ross adopts a chronological reading of *before* too quickly. He makes his remark in the context of considering this example, after all:

> I am walking along the street, and I see a blind man at a loss to get across the street through the stream of traffic. I . . . do not ask myself what I ought to do, but more or less instinctively take him by the arm and pilot him across.[20]

Ross's response is not an artifact of immaturity. Neither is it a sign that Ross was more a man of action than a reflective, theory-minded sort of person. Instead, the principle that one ought to assist those in need operates as a summary of a general element that indefinitely many actions by indefinitely many persons share.[21] In short, the practical grasp is nonchronologically prior to the theoretical formulation as well. The practical knowledge grounds the summary judgment. Ross remarks, "We see disinterested help being given by men to one another every day, without any thought of duty."[22] Thought about duty—reckoning that turns upon explicit formulation and consideration of general elements in ethical conduct—belongs to three sorts of situations for Ross, only two of which involve apprehending a particular duty on the basis of a general principle.[23]

Frankly, Ross's account of thinking things through is tortured.[24] Reckoning across different duties expressive of different regions of human good, informing the interdependent lives of members of the social body differently, is apparently private, difficult, and unlikely to yield sound general conclusions. For Ross, the hard part is accounting for the human who steps back and tries to work things out in theory. This is the starting point for Audi.

Nowhere in Ross's account do general principles take center stage as the basic stuff of ethical life. Upbringing, habituation, character, practical wisdom, interpersonal relations, relative power and vulnerability, merit, differential welfare—these inform conduct essentially. Intuitions are judgments that fasten onto ethically salient aspects of circumstances, and the capacity to have insight of the relevant sort—sound intuitions—is in turn a matter of being an adult already adept at decent

conduct. Or so Ross urges repeatedly. The awkwardness in Ross's system may be an artifact of the difficulty of accounting for practical knowledge in the theoretical idiom of moral philosophy in the longer twentieth century. On this reading, Ross's insistence that good conduct depends upon motive, that motive is different from intention, that a right act is not a good act if done in the wrong manner or from an ulterior motive, and so forth, becomes an attempt to get at the thought that the virtuous person is our best guide to conduct. The talk about ideally moral persons becomes a way of getting at the ancient insistence on the unity of the virtues. And it becomes impossible to miss echoes of Aristotle in passages like the following:

> . . . as might be expected, goodness of character is the only condition that with even the slightest degree of probability tends to make for the doing of right acts. If a man is not morally good, it is only by the merest accident that he ever does what he ought. The act to which he is attracted by one feature of it, itself morally indifferent or bad, may be the act towards which a good man would be attracted by its whole system of morally significant features, but if this is so, the coincidence is accidental.[25]

There is little space here for raising the question why the human as such takes an interest in ethics. The human is essentially ethically situated and active. Although the content of Rossian duties is communicated in simple sentences, the status of these sentences is not happily expressed using our words *rules* and *principles*.[26]

By Way of Conclusion

I have given a selective reading of Ross and a general take on Kant in which each understands the challenge and possibility of ethical conduct as being central to our natures, and in which neither sees the task of moral theory as providing comprehensive normative guidance. If either has an epistemological agenda, that agenda consists in articulating general aspects of practical knowledge rather than giving a systematic treatment of theoretical principles that pertain to conduct. Audi may well respond, "So much the worse for this Ross and this Kant." Audi's project is launched miles above the kind of ethical grounding that my historical figures sought.

What I find difficult and fascinating about this Kant and this Ross, however, is that each takes the job of grounding moral judgment to turn on explaining what it is about us, and what it is about ethics, that makes our interest in ethics no accident, and makes it no accident that the ethical challenges us. Such ground has to be both actual and internal to us in ways that Audi's principles are not—internally as actual sources of conduct, and actually in our social relations as these inform our individual and collective lives. To whatever extent search for such grounding is crucial to moral theory, it will not do to start in midair.[27]

Notes

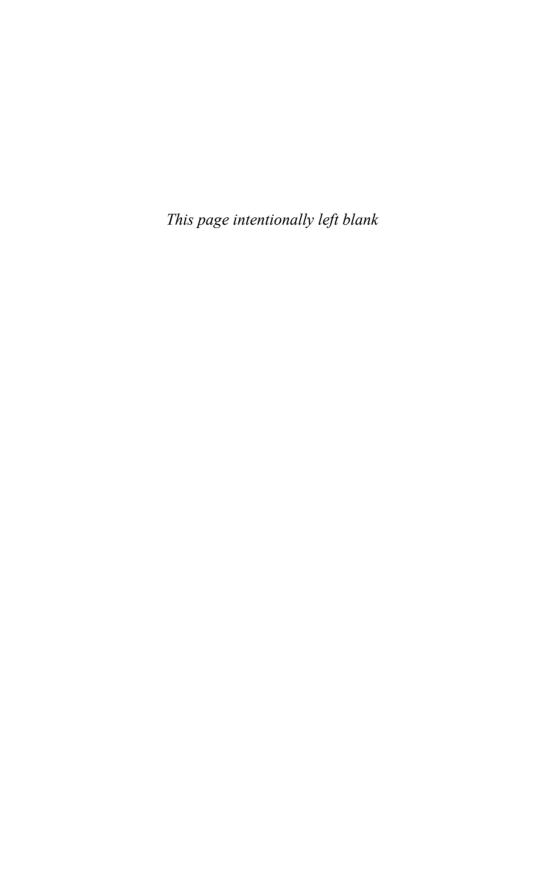1. Robert Audi, *The Good in the Right* (Princeton, N.J.: Princeton University Press, 2004) [hereafter, *GR*], 197.

2. Audi could reply that fidelity to the dead is not on his agenda. He instead aims at producing a metaethical theory that will "accommodate the Kantian, utilitarian, and virtue-ethical elements that are most needed in practical ethics" (*GR*, 197). Audi's interest in intuitionism arose through teaching professors of applied ethics in, e.g., "business, engineering, journalism, law, and medicine," who "found W. D. Ross's intuitionism at once theoretically straightforward and, in many cases, more readily applicable to practical moral problems than virtue ethics, Kantianism, or utilitarianism" (*GR*, ix).

3. Ibid., 198–99.

4. Ibid., 198.

5. Ibid., 199.

6. Of course, *having* an account of the ground as it emerges from the relevant bits of Ross, or animates Kant, may fail to provide a guide the perplexed ethicist asked for in order to determine whether the doctor should perform a cesarean now, say, or how best to cope with property in electronic text and Web-available music in an era of downloads, much less the professional charged with teaching anyone how to make these determinations in practice or in policy. But that is another matter entirely.

7. The inputs of Audi's "test" are axiomatic principles and subsidiary rules, not maxims. The problem I am gesturing toward is *prior* to the maxim-specification problem. Whatever a maxim turns out to be, however specified, it isn't clear *how* we are assessing it when we consider it from the standpoint of humanity or of universal law.

8. The ends and means of the highly abstract formulation utterly fail to distinguish lacing my shoes from watching a movie, volunteering at the halfway house, giving a lecture, and so forth, hence fail to correspond to the usual uses of those terms in practical philosophy, and so are rightly treated as very obscure.

9. *GR*, 91.

10. Ibid., 91–92.

11. Being so-directed, for Kant, is at the heart of virtue (provided that I include myself among "persons"—virtue gives me two ends: my own perfection and others' well-being); the relation of this formula to right, which sets no ends, is less clear.

12. This is why Kant gives the assortment of examples that he does—perfect duty to others; imperfect duty to others; perfect duty to self; imperfect duty to self.

13. *Critique of Practical Reason*, trans. Lewis White Beck (New York: Macmillan, 1956), 8n5 [*Ak* 5: 8].

14. *The Right and the Good* (Oxford: Oxford University Press, 1930) [hereafter, *RG*], 20.

15. Ibid.

16. Ibid.

17. Ibid.

18. Ibid., 48–50.

19. *Foundations of Ethics* (Oxford: Oxford University Press, 1939) [hereafter, *FE*], 169.

20. *FE*, 168.

21. Ibid.

22. Ibid., 169.

23. These two are found on 171–73:

a. Situations in which we take it that acts with attribute *A* have attribute *B*, we don't know why, and, we expect *B*, having satisfied ourselves that an act instancing *A* is at hand.

b. Situations we meet having arrived at some general principle that "all *A* is *B*" on sound reflection, but where we no longer recall the considerations in support of our generalization, and so must rely upon the general principle directly.

The third point of intrusion for general principles in ordinary moral life arises when we face a conflict of prima facie duties and so must work out what is required under the circumstances. Here, we may rely upon propositions that express the general elements shared across different species of good deeds in the course of thinking things through.

24. See *FE*, 173–91. It is partly tortured by concern over estimating the future in novel circumstances, over "comparing the goodness of various results," and over "balancing the duty of producing the greatest good against special obligations" (*FE*, 190). It is also tortured by the need to allow for moral insight that runs counter to common opinion, hence exceeds the practical ground established through a lifetime of moral conduct. Ross gives, as an example, "Wilberforce . . . when he devoted his life to the abolition of the slave-trade and of slavery" (*FE*, 190).

25. Ibid., 310.

26. Even though Ross uses just these terms in summarizing the results of his Gifford Lectures (*FE*, 311–15).

27. It is a further question whether Kant was right about rational nature, or whether Aristotelian human nature can do the work that Ross hoped it might.

KNOWLEDGE, JUSTIFICATION, AND ACCEPTANCE

This page intentionally left blank

Are Perceptual Beliefs Properly Foundational?

LAURENCE BONJOUR

I have known Robert Audi for most of my philosophical life, beginning with the time when we were both junior professors at the University of Texas in the late 1960s and early 1970s. We have kept in touch pretty steadily through the intervening years, and it would be difficult to overestimate how much I have learned from him. For a time, we were on opposite sides of the foundationalism-coherentism divide, and Audi's gentle but insistent probing of the weaknesses of coherentism as a general epistemological position (along with his insistence on recognizing and accommodating the genuine insights of the coherentist) was one major influence on my eventual abandonment of coherentism and shift to foundationalism. In recent years, our broad epistemological views have become increasingly similar: we share foundationalism, internalism, and rationalism (though Audi's brand of internalism is more nuanced and qualified than my own). And my respect for Audi's philosophical acumen is great enough that I cannot help but think that this convergence must mean that I am getting at least some things right!

But, of course, we do not agree on everything, and it is the somewhat curious practice of philosophers, even in the process of honoring a valued colleague and friend, to focus on disagreement more than agreement. Thus my aim in the present paper will be to explore an important issue on which Audi and I have differed, one that arises within the context of the epistemological foundationalism that we share. The issue in question has to do with the *scope* of the foundation: Audi and I agree that (some) a priori beliefs and also (some) introspective beliefs about one's own states of mind are foundational. But his view is that perceptual beliefs about material objects can also have the same foundational status as beliefs of these other two kinds;[1] whereas my own tentative albeit reluctant view has been that perceptual beliefs about material objects are never strictly foundational, with the main reason being simply that no good account of how they could have this status seemed to me to be available.[2] I remain unsatisfied with Audi's main accounts of the foundational status of perceptual beliefs, at least as I understand them. But I now *think* that I can see how to do better in this regard, and moreover in a way that it seems to me that Audi might well be willing to accept.

I will begin by examining, in the first two sections, the main accounts of the foundational status of perceptual beliefs that I find in Audi's work. The third section will then sketch and tentatively defend my own proposal.

I

It will be useful to begin with an example. Walking across an unfamiliar part of the University of Washington campus on a clear and sunny autumn day, I turn a corner and am presented with the visual experience of a large, deciduous tree standing beside the path, thirty-five or forty feet away. More or less instantaneously I form the belief that a tree of this sort is in front of me at approximately that distance. (Indeed, being something of a tree aficionado, I form also the more specific belief that the tree is a big-leaf maple, a species that is common and familiar in the Seattle area; but it will simplify the discussion to focus on the more general belief.)

Two things about this case seem initially obvious enough to be largely taken for granted in our subsequent discussion, even though neither of them is entirely immune to challenge. First, assuming that the surrounding conditions and circumstances are otherwise normal, my belief that there is a deciduous tree in front of me is *somehow* justified in the epistemic sense that is relevant to the issues of truth and knowledge.[3] Second, the visual experience that I have (or seem to have) of the tree plays in *some* way a central role in the justification of this belief. A third claim that I will assume for now but which we will have occasion to reconsider later on is that the experience plays this central justificatory role by in *some* way providing or contributing to a reason for thinking that the belief in question is *true*. The question is how all of this works and in particular whether it works in a way that gives my justified belief that there is a tree in front of me genuinely foundational status—which I take to mean at least that its justification does not depend on that of any other empirical belief or belief-like state. In the present discussion, I will also assume, even though this is also not quite beyond all controversy, that if there are foundational perceptual beliefs about material objects, the belief I am focusing on is surely one of them.

There is one more assumption that I want to make for the purposes of the present paper, even though I am well aware that there are some philosophers who will think that it is mistaken and even though I have no space here to even begin to defend it. My perceptual state as I see or seem to see the tree is complicated, and I see no reason to deny that it involves elements that are *conceptual* in character: that essentially involve or employ concepts, such as color concepts, spatial concepts, and the concept of a tree itself. But what I will refer to as the *sensory core* of such an experience, that on which its main justificatory force in relation to the belief in question rests, seems to me to be clearly *non*-conceptual in character (even though it is, like anything else, something that could at least in principle be conceptually described). I am inclined to think of this sensory core as something like the presence of a complicated spatial pattern of various shades of brown and green and yellow and other colors occupying a large part of my visual field, with the

general shape of this pattern being something very approximately like a tall, somewhat splayed rectangle with a large, fuzzy, and variegated circle resting on top of it.[4]

How then does such a non-conceptual experience of the sort indicated justify or contribute to the justification of the corresponding belief, and does it do so in a way that gives that belief foundational status? There is an approach to these questions with which Audi seems to have a good deal of sympathy, though it does not seem to capture his entire view of the matter, and which a number of other epistemologists have also advocated. It begins with the observation that a natural characterization of the experience in question, as my initial description of the case indeed suggested, is precisely as the experience of a large deciduous tree (or as being "appeared to" in a large-deciduous-tree manner, or as an experience "as of" there being a large deciduous tree there, and so on). Obviously the occurrence of such an experience does not guarantee that such a tree is present, but does it not at least yield a prima facie or defeasible reason for thinking that this is so? As Audi puts it

> . . . when, on the basis of an apparently normal visual experience . . . , one believes something of the kind the experience seems to show . . . , normally this belief is justified. Call this the *visual experience principle*. . . . [It] takes us from visual experience—conceived as apparent seeing—to justification.[5]

As stated, this principle does not quite say either that the experience alone is sufficient for justification or that the beliefs thus justified are foundational, but it is quite clear from the context that this is Audi's intent. Similar principles have been offered by others, for example by H. H. Price and by Roderick Chisholm.[6] (In what follows, I will, like Audi, adopt the fairly standard philosophical practice of focusing almost entirely on the case of vision.)

But why exactly should this principle, thus understood, be regarded as acceptable? Audi presents it as intuitively plausible, but he says relatively little in that discussion or indeed elsewhere to elucidate and display this supposed plausibility. The tenor of the passage suggests, however, that he would endorse a recent, somewhat more expansive statement of the same basic view offered by James Pryor:

> For a large class of propositions, like the proposition that there are hands, it's intuitively very natural to think that having an experience as of that proposition justifies one in believing that proposition to be true. What's more, one's justification here doesn't seem to depend on any complicated justifying argument. An experience as of there being hands seems to justify one in believing that there are hands in a perfectly straightforward and immediate way . . . *the mere fact* that one has a visual experience of that phenomenological sort is enough to make it reasonable for one to believe that there are hands. No *premises* about the character of experience—or any other sophisticated assumptions—seem to be needed.
>
>   I say, lets take these intuitive appearances at face value. Let's say that our perceptual beliefs in these propositions are indeed justified in a way that does not require any further beliefs or reflection or introspective awareness. They have a kind of justification which is *immediate*, albeit defeasible.[7]

On Pryor's view, I take it, the intuitive plausibility of such a principle is the bottom line, and no further defense or explanation is required for it to be acceptable. And

Audi's discussion of the "visual experience principle" at least suggests a similar view.[8]

That such a principle does indeed have a substantial degree of intuitive plausibility, especially from the standpoint of common sense, would be very hard to deny. There are, however, at least two reasons why it seems to me unsatisfactory to rest the case for the foundational character of perceptual beliefs on that intuitive basis alone.

The first reason is simply that our job as philosophers is, I would suggest, not only or even primarily to offer principles that are correct or even correct and intuitively justified. It is to shed light on the issues we discuss in a way that yields a deeper understanding. In other words, we want to know not just which principles are true, but also, even more importantly, *why* these principles are true. And it seems to me hard to deny that Audi's principle or the analogous principle suggested by Pryor yield by themselves very little in the way of such understanding.

One way to elaborate this point is to reflect on the fact that a perceptual experience of the sort we are concerned with is after all ontologically distinct from the worldly situation that would make the corresponding belief true.[9] This seems to make it entirely appropriate to ask *why* the occurrence of one sort of situation (my having the experience of a large deciduous tree) is a good reason for thinking that the other, ontologically distinct sort of situation (the actual existence of such a tree) genuinely obtains—assuming for the moment that the justification of the corresponding belief requires such a reason. And merely insisting that a strong degree of intuitive plausibility attaches to such a claim seems to me to be an inadequate response. If the occurrence of the one situation does in fact provide a good reason for believing in the other, there must, I suggest, be some further, more articulated account of why this is so. (Whether and to what extent ordinary believers would have to themselves have access to such a reason for their beliefs of this general sort to be justified is a further issue, to which I will briefly return below.)

Clearly the principle that Audi suggests and that is reflected in Pryor's remark relies heavily on the characterization of the perceptual experiences in question as ones that seem to show or present material objects of the sort that the corresponding beliefs claim to exist: experiences, as Pryor puts it, "as of" the believed propositions being true. It is the fact that the experiences are so naturally described in this way that makes it seem so plausible to say that the occurrence of such an experience makes the corresponding belief at least prima facie justified. What, we might ask, could provide a better reason for thinking that there is a large deciduous tree in front of me than that I have an experience "as of" such a tree?

Before accepting this result, however, it seems to me important to ask just what the significance of these characterizations of experience in physical-object terms really is and what it is about the experience that makes them seem so obviously appropriate. And once questions of this sort are raised, the answers are, I suggest, not nearly as obvious as one might initially have thought. Clearly if the characterization of my experience as an experience of a large deciduous tree is to be the basis for a *foundationally* justified perceptual belief, that characterization cannot be taken to mean simply that this is the sort of experience that is always or usually *caused* by the existence of such a tree, for this is something that seemingly could only be known

empirically and in a non-foundational way. Clearly also, the non-conceptual sensory core of the experience, from which, we are assuming, its justificatory force derives, does not essentially involve the concept of a large deciduous tree, for it does not involve any concepts. And most clearly of all, the existence of such a tree is not somehow literally a constituent of the experience.[10]

Moreover, if we think of the properties that are most immediately involved in or represented by the sensory core of the experience,[11] two further things are clear. First, many of the essential properties that pertain to a large deciduous tree are not the sort that could be present in this immediate way in the sensory core of the visual experience in question: its solidity, its weight, the roughness of its surface, its temperature, its being composed of wood, and so on. And, second, even the properties of the supposed tree that could in principle be involved or represented in the sensory core of a visual experience do not correspond very closely to the ones that are actually experienced. The tree has a complicated three-dimensional shape, but many aspects of that shape are not to be found in any very clear way in the experience. The tree has a certain size that the experience reflects only imperfectly if at all. And even the color properties that appear most immediately and unproblematically in the experience correspond only very approximately to the ones that common sense would ascribe to the tree, in ways that are affected by such things as shadows, the time of day, whether I am wearing sunglasses, and so on.

So what then is the connection between the sensory core of the experience and the tree that makes the description of the experience in this way seem so obviously appropriate from a common-sense standpoint? As already suggested, the answer seems to me far from obvious. It is clear that there is a complicated correlation between kinds of physical objects and situations, on the one hand, and various sorts of perceptual experiences, on the other—one that we confidently apply or at least rely on, and, in most cases, with little hesitation or reflection. Perhaps this correlation is learned, perhaps it is innate, or perhaps it has some aspects of each. But whatever its provenance, the nature of this correlation is neither straightforward nor obvious. And this means, I suggest, that accepting beliefs about the world on the basis of it is more problematic than it initially may have seemed. One way to put this point is to say that if we set aside the characterization of my original experience as an experience of a large, deciduous tree and focus simply on the phenomenological character of the experience itself and especially on its sensory core, it is not immediately obvious why such an experience is appropriately described as an experience of such a tree or, still less, just why it counts as a reason for thinking that such a tree is present. And this seems to me enough to at least raise doubts about the correctness of the sort of principle advocated by Audi and Pryor—which is my second reason for thinking that it is a serious mistake to rest the case for the foundational status of perceptual beliefs on that principle's initial intuitive plausibility.

I want to be clear that I am not saying, nor do I believe, that the correlation between experiences and material objects that is the basis for our commonsense descriptions of experiences in material objects is somehow merely arbitrary or conventional. On the contrary, as will emerge, I think that there are good reasons for this correlation, ones that do explain why the occurrence of such an experience should be taken to be an appearance of the corresponding sort of object and also

perhaps support the conclusion that the occurrence of the experience is after all a good reason for the correlated material object claim. But I also think that the full account of why this is so is sufficiently complicated and difficult to raise doubts about whether the justification that results can count as *foundational*. I will return to this issue in the final section of the paper. In the meantime, however, I want to look at a quite different and much more complicated account of the relevance of experience to material object claims that is offered in another place by Audi.

II

What is perhaps Audi's most extended discussion of the issues that are the focus of the present paper occurs in his paper "Justification, Truth, and Reliability."[12] The argument of this paper is remarkably rich and complicated, and it will be impossible to do full justice to it here. But the aspect that is most relevant to our present discussion may be thought of as involving five closely related theses pertaining to the visual experience principle cited earlier (here formulated in a somewhat different but essentially equivalent way) (308).[13]

The first thesis is that the visual experience principle is (like the others) known and justified a priori, via reflection on the principle itself and on relevant hypothetical cases (312). Such a principle is not observationally testable (without presupposing either that principle or others like it), and challenges to it can, Audi claims, be "properly" met in the ways appropriate to a conceptual truth: by questioning whether the challenger really understands what he is saying and is serious about it (313). Thus it is plausible to regard this principle (and the others) as "conceptually necessary" (314).

This immediately poses a problem, however, for the intuitive connection between justification and truth suggests that such a principle could be knowable a priori only if the reliability of accepting beliefs on the basis of it were also knowable a priori: only if there were some a priori reason for thinking that beliefs like my belief about the large deciduous tree are likely to be true when accepted on that sort of basis. Audi's view, however (the second of the theses we are concerned with), is that the *reliability* of this principle is "a wholly empirical matter" (314–15, 320), where the main reason cited is the possibility of something like a Cartesian demon who causes it to be the case that most of the beliefs accepted by following the visual experience principle are false (314–15).

Audi briefly discusses and rejects three familiar historical views that attempt to resolve this apparent conflict by arguing that it is, after all, necessary that most of the beliefs satisfying the visual experience principle are true. Descartes attempts to secure this result by invoking the existence and benevolence of God, phenomenalism by construing the material objects that are the objects of the beliefs in question as logical constructions from the corresponding sense experiences, and Kant by construing such objects as mind-dependent in a more complicated way that still supposedly guarantees that perceptual beliefs about them are mostly true (315). Audi rejects the idealism reflected in the latter two views; his reason for rejecting Descartes's theistic appeal is not made explicit, but perhaps he would say simply that it is implausibly indirect as a justification of the visual experience principle.

Such views, or at least the latter two of them, attempt to establish what Audi describes as an *ontological* connection between justification (as prescribed by the visual experience principle) and truth. Having rejected them, he suggests a different possible way of achieving this same result, one that is compatible with a realist view about material objects: rather than viewing objects as somehow *constituted* by the experiences in question, we might instead think of them as genuinely external to the mind but as nonetheless partly *conceived* as things that cause (and explain) the relevant experiences. Thus, for example, a large, deciduous tree is on this view essentially conceived, in part, as the sort of thing that causes the kinds of experiences that we construe, according to the experience-object correlation mentioned earlier, as experiences of such a tree (only in part, because the tree is also conceived as an object that has the full panoply of other properties that are ordinarily ascribed to it). Audi calls this view, which he apparently accepts (and which is the third of the theses we are concerned with), *epistemic realism* (318). According to it, "real objects are epistemically constituted: necessarily such that they *tend* to produce justified beliefs . . . about them" (318)—beliefs, that is, that are justified according to the visual experience principle.

Views according to which the very concept of a material object includes the conditions under which beliefs about it are justified have sometimes been put forward as adequate to solve the problem of showing that justified beliefs about such objects are likely to be true. Audi, however, is very careful to eschew any such claim. The most that epistemic realism can show, in his view, is that the existence of a large deciduous tree of the appropriate sort is *one* way to explain my experience of seeming to see such a tree. But, alas, it remains true that other explanations (such as a Cartesian demon) are also possible, so that we cannot on this basis know a priori that the explanation that would make the belief true is more likely to be correct than one of these others (319)—the fourth of the relevant theses.

Thus, in Audi's view, as I understand it, all attempts to show that beliefs justified according to the visual experience principle are thereby likely to be true fail. Where then does this leave us? His answer is that even though we cannot know that accepting beliefs according to this principle will lead us to the truth, we can still know that beliefs accepted in this way are justified. While the visual experience principle is not, as an idealist view would suggest, "partly constitutive of truth," it is nonetheless "partly constitutive of epistemically permissible attempts to show truth" (320). This is so, because as epistemic realism suggests, our very concept of justified belief "is in part constituted by" principles like the visual experience principle that underwrite our appeal to justificatory elements like the relevant sense experiences. In this way, even though there is no ontological connection and hence no a priori guarantee of success, our concept of justification and the practice that embodies it aim *teleologically* at truth—the fifth and last of the theses with which I am concerned here.

Does all this yield, as Audi tentatively suggests, a satisfactory defense of the visual experience principle and the foundational justification that it ascribes to beliefs like my belief in the large deciduous tree? Despite the care and subtlety of Audi's discussion, I find myself unconvinced that the overall result is satisfactory. I will attempt to explain why by developing two objections, one pertaining to Audi's

shift to a teleological construal of the relation between justification and truth and the other to the underlying idea of what he calls epistemic realism.

The first and more obvious objection challenges whether merely *aiming* at truth in the way that Audi has characterized is enough to show that our epistemic practice (and the principles such as the visual experience principle that it embodies) genuinely yield justification of the relevant epistemic sort. On Audi's view, as I understand it, we have in the end no good reasons (of a non-question-begging sort) to think that accepting beliefs about the material world on the basis of visual experience is likely to yield beliefs that are true. To be sure, it is (according to epistemic realism) *part* of our very conception of such objects that they lead to experiences of that kind and also that beliefs about them accepted on that basis are justified. But such objects are of course *also* conceived as having a wide variety of physical and causal properties that are entirely independent of such epistemic claims. And Audi's view seems to me to amount to saying that while both of these aspects are involved in our conception of material objects, we have in the end no reason at all for thinking that a case in which the justification conditions are satisfied is thereby one in which the other features required for such an object to exist are likely to be present.

In this way, the very conception of material objects advanced by the epistemic realist seems threatened by incoherence. How can it be part of the conception of such objects that the occurrence of sense experiences of the relevant sort justifies beliefs in their existence in a way that genuinely aims at truth, while it is at the same time true that the occurrence of those experiences provides in the end no reason at all for thinking that the other, at least equally essential features of such objects are genuinely realized? Audi remarks (320) that the epistemic practice that embodies this principle presupposes its "de facto reliability" and truth-conduciveness, though not its necessary (and a priori knowable) reliability. But how can this practice and the related conceptions of justification and of the nature of material objects that it embodies be rationally acceptable if there is and apparently could be no non-question-begging reason for thinking that this essential "de facto reliability" genuinely obtains?

A second, not unrelated objection has to do with the specific sorts of sensory experience whose justificatory relevance is, according to epistemic realism, part of the conception of a specific sort of material object, such as a large deciduous tree. I have already spoken of the correlation between material objects and experience that we on some basis accept, asking how it should be understood. The epistemic realist view amounts to saying that it is part of the very conception of a certain sort of material object that the experiences associated with it by this correlation are caused by it and justify the claim that it exists. But is there, on Audi's view, any further, justificatorily relevant reason why one particular set of experiences rather than another is associated in this way with a given sort of object? Is the connection between tree experiences like mine and the concept of such a tree simply one that we happen, willy-nilly, to accept, or is there some further rationale available for it?

The way in which epistemic realism is presented seems to suggest that there is after all no further reason or explanation as to why the object-experience correlation takes the specific form that it does. This would mean that all we can say is that *these* experiences are justificatorily relevant to the belief that a large deciduous tree exists

in a certain location simply because *these* are the experiences whose relevance is partly constitutive of the concept of such a tree. But then any experiences at all could seemingly have in principle equally well played that role on the very same sort of basis. We associate and build in as justificatorily relevant to a given sort of object one specific set of experiences rather than any of the indefinitely many possible alternatives, but there is on this view no further reason or rationale at all for the choice (beyond, perhaps, an innate one, grounded in evolution).

Such a result seems to me to render the resulting concept of justification and the epistemic practice that allegedly rests on it at bottom arbitrary in a fundamental way. Not only would we have, as already discussed, no non-question-begging reason for thinking that accepting beliefs on the basis of the sense experiences whose justificatory relevance is allegedly built into our concepts is likely to lead us to the truth; but we would also have no reason for thinking that the choice of such experiences that our concepts happen to embody has anything further to be said for it than just that it reflects the correlation between experience and material objects that we happen, for all we know on an essentially arbitrary basis, to accept. I am not sure that this is what Audi intends, but I am unable to see how his view has the resources to avoid it.

Such a view seems epistemically unsatisfactory, but it also seems intuitively wrong: surely there is some reason why this particular set of experiences rather than all of the other possibilities seems justificatorily relevant in a distinctive way to the existence of a large deciduous tree. But if there is such a reason, then the connection between experiences and material objects would not have to be viewed as simply built into our material-object concepts in the way that epistemic realism suggests. And, more importantly, if we could say what that reason is, then *perhaps* we could also find a reason for thinking that the beliefs we accept on the basis of the visual experience principle are after all likely to be true.

III

I have now examined two different (but still essentially compatible) ways of attempting to explain how and why perceptual beliefs can be justified by sensory experience in a way that makes them foundational: first, the direct appeal to the intuitive plausibility of this claim, suggested by both Audi and Pryor; and, second, the much subtler attempt by Audi to explain how such beliefs can be justified by the correlative experience in a way that is appropriately relevant to truth. Neither of these accounts seems to me satisfactory, and while there is more than one reason for this, one central problem is that they both take for granted our intuitive or instinctive correlation between material objects and sensory experience, rather than trying to understand and explain it.

But then how *does* non-conceptual experience contribute to the justification of perceptual beliefs? And does it do so in a way that gives foundational status to the resulting beliefs about the material world? My answer to the first of these questions is one that Audi mentions and even seems to endorse in a few places, including the paper just discussed,[14] but never discusses very extensively: as Locke was the first to

suggest clearly, our sensory experiences give us good reasons for our correlative beliefs about material objects because the existence of the objects in question provides the best *explanation* for the existence of such experiences. Defending this claim in detail is notoriously fraught with difficulties. But such problems not withstanding, it is very hard to see how it can fail to be true if accepting beliefs about the material world on the basis of that experience is to be a reasonable way to seek the truth at all. If there are other explanations of why we have the specific experiences we do that are in fact either better than or just as good as the material-object explanation, then it is hard to see why the rational course is not either to accept beliefs reflecting those other explanations, if they are better, or to simply suspend judgment (or accept a disjunctive claim), if they are equally good. And if we simply cannot justifiably say which explanation is best, then we seemingly have no good reason for accepting any of them. Such an assessment would be true for scientific investigation or for the work of a detective, and I can see no reason for thinking that it does not apply just as well in this more fundamental area.

I have no space here to develop anything like a full defense for this explanatory thesis.[15] What I want to do instead is to say something about the general shape of such a view and the difficulties that face it, focusing here especially on the question of whether an explanatory thesis of this kind can plausibly yield *foundational* justification for perceptual beliefs.

The place to start is with a crucial point already suggested by the foregoing discussion: the defense of such an explanatory thesis cannot appeal in any essential way to our intuitive correlation between experiences and material objects—in particular, cannot argue that the truth of a certain material object belief would be the best explanation for the intuitively correlated experience simply because we commonsensically take that experience to be an experience "as of" such an object or objects. Because the correctness and reasonableness of that correlation is a large part of what is at issue, any appeal of this sort amounts, in my view, simply to begging the question. What is needed instead is a more articulated account of *why* the intrinsic features of the experience, or more specifically of its sensory core, are in fact best explained by the existence of such an object—thereby attempting to *justify* the accepted correlation, rather than merely presupposing it.

What then are the intrinsic features of non-conceptual sensory experience and visual experience in particular that might seem to require an explanation in terms of material objects? Perhaps not surprisingly, it is the *spatial* properties presented or reflected in such experience that are, in my view, primary here. If I am right, however, the case for the sort of explanation in question requires an appeal not just to relatively momentary spatial properties, such as those to be found in my initial experience of the tree, but also to temporally extended patterns of spatial properties, especially but not only those that reflect the apparent movement of the subject through the world. Thus, to stick with the example of the tree, the full experiential basis for supposing that a three-dimensional object of that general shape is to be found in the world would, at least in principle, include the shapes reflected in all of the various experiences I might have in (apparently) approaching the tree and moving around it at various distances, (apparently) moving away from it and coming back from a different direction, (apparently) viewing it from the perspective of

adjacent buildings, and so forth. And the initial thesis would be that the existence of an external object with the three-dimensional spatial properties of the tree provides the best explanation for the systematic and connected availability of all of those experiences. Making a case for the non-spatial properties of the tree, including especially its causal properties, would be even more complicated and would require appeal to other senses, especially touch. Spelling out all of this in detail even for the case of the tree, let alone for our beliefs about the material world at large, would be tremendously complicated and difficult, but the approximate idea is perhaps clear enough for present purposes.[16] (It is perhaps worth adding, however, that I am not suggesting that the justification of a particular belief on a particular occasion—such as my belief that there is a tree before me now—requires having all of these varied possible experiences or even more than the few that I actually have. It may well be enough in a particular case that there are good reasons, based on a great deal of prior experience, for thinking that further experiences of these kinds are very probably available in relation to the one that I am actually having.)

While much more could and should be said about all this, the issue I want to focus on here is the issue of foundational status: Could such an explanatory view of the justificatory relevance of non-conceptual experience to perceptual beliefs, even if it could be successfully elaborated and defended, possibly yield the result that the beliefs in question are *foundationally* justified? There are three main hurdles to be surmounted here, each of them quite formidable. In the space remaining, I will only be able to suggest in a very tentative way how I think they might be dealt with.

First. If foundational justification is to result from explanatory considerations of the sort in question, it is obvious that the central explanatory thesis—namely that the best explanation of our non-conceptual experiences is in terms of the material objects with which they are correlated by the standard common-sense correlation—must not itself depend essentially for its justification on further empirical beliefs. While this thesis, once established, may perhaps be further refined and developed on an empirical basis, the basic justification must be a priori. But that a thesis of this sort can be justified a priori is a claim that many are likely to find extremely implausible. In considering how such a claim might be defended, I find it useful to divide the issue into two parts.

The first part of the issue is whether or not there are other, equally good or better explanations of the experience in question that are of the same basic sort as the proposed material-object explanation: that is, very roughly, explanations that propose a realm of objects, presumably spatial objects, whose properties both cause and are reflected in the properties of our experiences, but in a systematically distorted way, with the result that those objects and properties are substantially different from those postulated by the explanation that follows the common-sense correlation. This is the sort of explanation that I have elsewhere referred to as an *analog* explanation, with the contrast being with those explanations, such as that involved in the familiar brain-in-a-vat hypothesis, that work by emulating the experience that would be produced by a world of material objects (call these latter explanations *digital* explanations).[17] In these terms, my basic suggestion is that it is far from obvious that there are any alternative analog explanations of our experience that are as good as or better than what I will hereafter refer to as the common-sense

explanation. I can think of none that has ever been suggested in any detail, nor is at all obvious how the distorting mechanism that such an explanation would have to involve could work, once the full range of relevant experiences are brought into the picture.[18]

What then about digital explanations? Here it is clear that there are as many possibilities as human imagination can devise, and that if carefully formulated, they can be made to fit experience perfectly: brain-in-vat hypotheses, Cartesian demon hypotheses, and so on. My main suggestion here is that such explanations are less good than the common-sense hypothesis precisely because of their digital character: by introducing an extra layer of complexity, they are less likely to be true than the analog hypothesis that they emulate.[19] For these reasons, I believe that a case can be made, even though I have surely not made it here, that the common-sense explanation is indeed superior to the apparent alternatives on a purely a priori basis.

Second. But even if the basic explanatory thesis is justified a priori, it will still seem obvious to many—as indeed it until recently did to me—that such an approach still cannot possibly yield the result that perceptual beliefs are foundationally justified. For, it will be argued, non-conceptual experience is incapable in principle of providing in itself the basis for such a justification. Instead, such experience would have to first be *conceptually* described, with only propositions or propositional beliefs about the character of experience being capable of playing any sort of direct argumentative or justificatory role.[20] While some have taken this to mean that non-conceptual experience can play no epistemic role at all, my own previous view was that its role is limited to justifying conceptual and propositional descriptions of its content.[21] The upshot would be that it is only beliefs about that experiential content that are genuinely foundational, with the justification of beliefs about material objects depending on inference from these and so not being itself foundational. Perhaps this is already implausible enough. But a further obvious and even more serious problem is that once descriptions of experience in terms of our intuitive experience-object correlation are set aside as essentially question-begging, as I believe they should be, it becomes abundantly clear that neither ordinary people nor even philosophers have either the conceptual resources or the developed abilities required to actually give descriptions of non-conceptual experience in enough detail to bring out the specific features that, on the view sketched above, demand specific material object explanations. And this would mean in turn that the envisaged explanatory justification for beliefs about material objects, even if possible in principle, is actually available to no one—and certainly not to ordinary believers.

As already noted, I myself accepted more or less this argument in the recent past, concluding in effect that an explanatory justification for material object beliefs is at best available in principle.[22] But on further reflection, I am now inclined to think, albeit still fairly tentatively, that the fundamental assumption lying behind this objection is incorrect, one more instance of the familiar tendency of philosophers to over-intellectualize ordinary modes of thought and thereby falsify them. Is it really so obvious, I want to ask, that descriptions of experience in conceptual terms are required in order for beliefs about material objects to be justified as the best explanation of that experience? What is ultimately supposed to be explained by

these material object claims is, after all, not such conceptual and propositional claims about experience, but rather the experience itself. So why then is it supposed to be impossible to judge directly that such-and-such an experience or extended pattern of experience is best explained by the existence of such-and-such a material object, without interposing a conceptual description of that experience? Perhaps being able to indicate the relevant features of the experience in conceptual terms would make the explanatory thesis clearer and more explicit, but is it really essential to its basic justificatory force?

It is clear, to be sure, that in many, perhaps most cases, the justificatory force of explanatory reasoning does depend on a conceptual formulation of the facts to be explained, simply because there is no other way in which those facts could be accessible to the believer in question. Thus, for example, in order to be justified in believing that a suspect was at the scene of a crime because that is the best explanation of his fingerprints being found there, I must in some way be aware that fingerprints of that specific sort were found; and the only apparent way to do that is to have a (justified) propositional belief to that effect. Similarly, to be justified in believing that radioactivity is present because that is the best explanation for a Geiger counter emitting a distinctive crackling sound, I must in some way be aware that a Geiger counter is indeed doing that; and again the only way to have such an awareness is to have a (justified) propositional belief that it is. But while conceptual and propositional descriptions may indeed be required in these other sorts of cases in order to give us access to the facts to be explained, the crucial point about *conscious* sensory experience is that we are aware of it and of its specific character just by virtue of having it, with no essential need for a further conceptual description. It is still necessary, of course, to notice the relevant experiential features, but that also does not require a conceptual description—as shown simply by the fact that we are in general able to do so without (for the most part) having such a description.

Thus my suggestion is that the detailed features of our perceptual experience can constitute the basis for an explanatory justification of the sort suggested without any need for those features to be described or formulated in conceptual terms. If the various features of which we are aware are indeed best explained by supposing that various sorts of material objects exist in the vicinity and if we are able to see, even in a rough and ready way that this is so, then our perceptual experiences can constitute *in themselves* good reasons for thinking that the perceptual beliefs about such objects that we accept on the basis of them are true.[23]

Third. The final issue that I want to discuss briefly is whether and to what extent it is plausible that the sort of justification for perceptual beliefs that I have described is available to ordinary people. It is clear, of course, that they do not reason explicitly in the way suggested. But there still are many ordinary examples that seem to me to suggest that such an explanatory outlook is not foreign to ordinary thought. Think here especially of someone whose experience is puzzling or anomalous in some way, as when there is an apparent object that is hard to make out or features of the experience that do not seem to fit. It is very natural, I would suggest, to describe such a person as struggling to figure out why his experience has just the specific character or features that it does—a struggle that is often resolved by a slight change of perspective or a movement by the object in question, after which the

explanatory question is answered and things fall into place. Moreover, if an ordinary person could somehow be brought to take seriously the skeptical suggestion that the objects that he seems to himself to perceive are not really there, it is not at all unreasonable to think that his first question would be *why* he is having just that sort of experience if they are not.

My conclusion, which I still want to be regarded as a tentative one, is that I was mistaken in my earlier view about the foundational status of perceptual beliefs, and that Audi was right, albeit perhaps not for quite the reasons that he supposed. But I will leave it to him to address the extent to which the view suggested here is compatible with his.

Notes

1. See his book *Epistemology*, 2nd ed. (London: Routledge, 1998).

2. For discussion of this point, see my part of BonJour and Sosa, *Epistemic Justification* (Oxford: Blackwell, 2003). It is worth noting that Audi also counts memory beliefs as foundational. (See *Epistemology*, chapter 2.) Perhaps fortunately, there is no room in this paper for a consideration of the difficult issue of memory. I will only note that even though some of the issues I will be raising about perceptual beliefs arise for memory beliefs as well, I do not think that the analogy between two areas is close enough to create any very strong presumption that the same sort of account should be given for both.

3. Whether it is justified to the somewhat mysterious degree that is required for knowledge is a further question with which I will not be concerned here.

4. In describing it in this way, I do not mean to be taking any stand on a familiar issue concerning the ontology of sense-perception: perhaps this pattern of color is a pattern of sense-data as classically understood, perhaps it is just my being appeared to in a certain way, or perhaps it is even in *some* sense to be identified with the tree itself (if there really is one there). No view on this last issue has, as far as I can see, any bearing at all on the character of the experience qua experience—which is, of course, one of the things that makes the ontological issue so difficult to resolve.

5. Audi, *Epistemology*, 28.

6. See H. H. Price, *Perception*, 2nd ed. (London: Methuen, 1950), chap. 7; and Roderick Chisholm, *Theory of Knowledge*, 3rd ed. (Englewood Cliffs, N.J.: Prentice-Hall, 1989), 46–54, 64–68, 71–74.

7. James Pryor, "The Skeptic and the Dogmatist," *Noûs* 34 (2000): 536.

8. See also his reply to Walter Sinnott-Armstrong in a book symposium on Audi's *The Architecture of Reason* in *Philosophy and Phenomenological Research* 67 (2003).

9. One might wonder whether this is true on a "direct realist" view of sense perception. I can only say that I can find no otherwise plausible way of understanding such a view that calls this ontological distinctness into question. See my paper "In Search of Direct Realism," *Philosophy and Phenomenological Research* 69 (2004).

10. Some recent British philosophers have defended, or seemed to defend, views according to which material objects are in some way literally constituents of perceptual experience. See, for example, Bill Brewer, *Perception and Reason* (Oxford: Oxford University Press, 1999). I have no space here to consider such views, but can only say that such a claim seems to me very hard to take seriously.

11. I am being deliberately vague about just how these properties figure in or relate to the experience.

12. First published in *Philosophy and Phenomenological Research* 49 (1988): 1–29; and reprinted in Audi's book *The Structure of Justification* (Cambridge: Cambridge University Press, 1993), 299–331. References in the text in this section are to the pages of this reprint.

13. And also to analogous principles governing introspection, memory, and the justification of simple a priori claims, though I will not be concerned here with the application of the theses in question to these other principles.

14. See also *Epistemology*, 315; and *The Architecture of Reason* (Oxford: Oxford University Press, 2001), 18.

15. For a somewhat fuller, but still extremely sketchy discussion, see chap. 5 of BonJour and Sosa, *Epistemic Justification*—though, as will be discussed below, the account there now seems to me to be mistaken in an important respect.

16. See ibid. for a more detailed, but still extremely schematic account, which also includes some discussion of the extension of the argument to non-spatial properties, mainly causal ones.

17. See ibid.

18. For some relevant discussion, see Jonathan Bennett's discussion of "size blindness" in his "Substance, Reality, and Primary Qualities," *American Philosophical Quarterly* 2 (1965).

19. For some elaboration, see BonJour and Sosa, *Epistemic Justification*, chap. 5.

20. For a recent expression of a similar view, see Timothy Williamson, *Knowledge and Its Limits* (Oxford: Oxford University Press, 2000), 194–98.

21. For elaboration see BonJour and Sosa, *Epistemic Justification*, chaps. 4–5.

22. Ibid., chap. 5.

23. Such a view was suggested by Paul Moser in *Knowledge and Evidence* (London: Cambridge University Press, 1989)—though, in my judgment, without saying enough about what it is about the experience that requires this sort of explanation.

Audi on Testimony

ELIZABETH FRICKER

1. Testimony: Audi's Two Principles

Robert Audi has contributed massively to contemporary epistemology. As part of that, he has written insightfully, interestingly, and mainly persuasively about testimony. Testimony has been the main focus of my own research output in recent years. In this commentary I shall sketch and criticize some of the key elements in Audi's view, contrasting them with my own. I shall focus on the account of how knowledge and justification are gained from testimony given in chapter 7 of his 1998 *Epistemology* (henceforth *E*), and refined in his 2006 "Testimony, Credulity and Veracity" (henceforth TCV). In these two discussions of testimony, Audi clearly commits to two principles concerning the epistemology of testimony.

First, Audi maintains that "if I do not know that the speaker at yesterday's conference lost his temper, then you cannot come to know it on the basis of my attesting to it. . . . What I do not have, I cannot give to you." Further, he adds, "Testimonially-based knowledge is received by *transmission* and so is not at all independent of whether the attester knows the truth of the proposition in question" (*E* 135–36). Thus he clearly holds:

> K-Nec: A hearer can acquire testimony-based knowledge from an attester[1] only if the attester herself knows whereof she speaks.[2]

Second, summarizing his discussion in *Epistemology*, Audi concludes that " . . . at least normally, a belief based on testimony thereby constitutes knowledge provided that the attester knows the proposition in question and the believer has no reason to doubt either this proposition or the attester's credibility regarding it" (*E*, 138). Thus (ignoring the qualification about 'normally', whose significance is not clear to me), we see that Audi holds:

> K-plus-no defeaters-Suff (KPNDS): Given that the attester who tells that P knows whereof she speaks, then so long as the believer has no defeaters for trusting the testimony, if she forms belief based on the testimony, this will be knowledge.

Lackey (1999) observes that we need to clarify the kind of defeaters in question: these may be normative or doxastic defeaters. Moreover, they may defeat the proposition asserted itself, or they may defeat the supposition that the attester is trustworthy or, as Audi prefers to say, credible. I shall gloss over these matters of detail that are not crucial to my argument today.

I shall argue below that K-Nec is a correct principle regarding knowledge gained from *trust in testimony*, given my own preferred quite narrow and specific explication of what it is for a belief to be based on trust in testimony. However Audi's general conception of knowledge, and of what it is for a belief to be based on testimony, does not allow him to hold K-Nec: it is not entailed or rendered plausible by his own arguments and general stance about conditions for knowledge.

Moreover, there is relatedly a tension amounting more or less to incompatibility between holding both K-Nec and KPNDS. My own favored conception of knowledge, and account of how a belief formed from trust in testimony can be knowledge, which vindicates K-Nec, is incompatible with KPNDS. K-Nec can be held together with KPNDS only if asserted by arbitrary fiat; it cannot be given an underlying rationale, compatibly with holding KPNDS.

2. Discussion of K-Nec

Many writers about testimony find K-Nec intuitively plausible, and maintain it. One reason it is so intuitive is that without it we cannot articulate the very strong intuition that testimony is not an original source of knowledge, but only a means of its transmission, spreading within a linguistic and cultural community. As memory conserves knowledge in a subject, and so preserves it within her from an earlier time to a later one, so testimony transfers knowledge—originated from some other epistemic source—from one individual to another.

In Fricker (2006), I show that knowledge based on trust in testimony is necessarily knowledge at secondhand. Consequently, there cannot be a proposition that is known of only from trust in testimony: if P is known at all, there must be, or have been, someone somewhere who knows or once knew P not just from trusting another's testimony. I obtain this epistemologically significant result, because my account of what it is for knowledge to be based on trust in testimony—the core case, as I argue—entails K-Nec.

What rationale does Audi provide for holding K-Nec? Audi seems to waver between announcing it as an independently plausible epistemic first principle versus grounding it as entailed by a reliabilist general conception of knowledge.

I do not think that announcing domain-specific epistemic principles as independent data is methodologically convincing. As theorists of positive epistemology, we do much better if we can exhibit our favored epistemology of a specific area—perception, introspective knowledge, testimony, and so forth—as instancing a persuasive overarching general conception of both knowledge and justification.

So we need to examine Audi's attempts to ground K-Nec in a reliabilist conception of knowledge. Audi argues that if an attester lacks knowledge of what she asserts, then, even if her belief is true, it will be so only accidentally; hence a hearer

who believes what this nonknowing attester tells her does not form belief reliably in the way required for knowledge: "Suppose I make a lucky guess and am right. Then I give you correct information which I do not know; but you are also lucky to be correct and also do not know. . . . It is a fluke that I get it right; it is even more of a fluke that you get it right . . . " (*E*, 135–36). However, Jennifer Lackey has argued persuasively that this attempt to ground K-Nec in reliabilist considerations does not work (Lackey op.cit.). The key suppositions of Audi's attempted grounding of K-Nec are this:

(i) If an attester A does not know what she asserts, then her assertion does not express a belief of hers that is reliably true (this is why she lacks such knowledge).
(ii) If A's assertion does not express a reliably true belief, then that assertion itself is not reliably true.

Therefore:

(iii) Forming belief on the basis of such an assertion is not doing so reliably, or via a reliable method, in the fashion required for knowledge.

Lackey has argued that this attempt to ground K-Nec in reliabilist considerations fails, because ii is false. She describes several scenarios in which, as she maintains, someone may be a reliable testifier about some subject matter, although she is not a knower, since not a reliable believer, about that subject matter. Lackey's examples effectively undermine the reliabilist argument for K-Nec sketched above by demonstrating the falsity of (ii).

Lackey herself regards her scenarios and rebuttal of (ii) as refuting K-Nec itself, since she herself (apparently) espouses a reliabilist general conception of knowledge. I reject this general conception, and with it her overall conclusion. In 4 below I briefly sketch my own alternative general conception of knowledge, and of how belief is based on testimony, of which K-Nec is a consequence. This account is incompatible with KPNDS. First I examine Audi's conception of what it is for a belief to be 'based on' testimony, and show how this fails to provide a rationale for K-Nec.

3. Audi on Testimony-Based Belief

Audi writes that "testimony-based belief . . . arises naturally, non-inferentially, and usually unselfconsciously in response to what someone says to us. I ask you the time; you tell me it is nine o'clock, straightaway I believe this on the basis of your saying it" (TCV, 1). Later he adds that "testimony-based belief, as I construe it . . . is never inferential" (TCV, 3). In saying this, Audi contrasts belief that is *based on* testimony in his sense with belief that is based on testimony plus beliefs about the speaker's credibility: "if, as a ground for believing what you say, I must infer your credibility from background information about you, my belief of what you attest, though acquired through your testimony, may not be said without qualification to be based on it" (TCV, 3). This recaps an idea expressed earlier in TCV:

Suppose I believe something on the basis of premises about your testimony, as where the content seems implausible by itself, but I judge you to be both highly competent and unassailably sincere, and for that reason I believe what you say. . . . This is not belief or knowledge on the basis of your testimony. My basis is a combination of your testimony and my beliefs about you. (TCV, 2)

Audi presents this conception of basing as a matter of the psychology of our normal reception of testimony. However, it seems clear that he takes it, or anyway uses it, as also having normative epistemic force. From his subscription to KPNDS, it is clear he holds that belief which is thus psychologically based on testimony alone (not on testimony-plus-supporting-beliefs-about-the-speaker) is epistemically proper, and apt to be knowledge. And this has surely to do with the fact that, if I form belief that P on the basis of hearing someone assert that P, the content of my belief matches that of the assertion, rendering the latter an appropriate, internally accessible ground for the former.

The conception of knowledge invoked here thus seems to be one that is moderately internalist, in that Audi-style testimony-based knowledge has a ground that is accessible to the speaker—her perception of the speech act—and this is essential to its epistemically entitled status. But it is also reliabilist, since in supporting this type of basing as being sufficient, together with K-Nec, Audi appeals to reliabilist considerations.

Notice that Audi conceives knowledge based on testimony to be exactly parallel in its internal grounding, both normatively and psychologically, to perceptual knowledge. In his view, it often happens that you tell me something and "I straightway believe you," and this yields entitled belief in just the manner that, for instance "I may just believe that a bat flew by if I see one zigzag across the evening sky" (*E*, 131).

To summarize: Audi thinks that we typically acquire beliefs from testimony that are based on it alone, that is, are neither psychologically mediated in their formation, nor sustained by apt beliefs about the speaker's credibility, in addition to the ground provided by the perceived testimonial act itself. And he thinks that such beliefs 'based on' testimony in his sense have in it a sufficient ground, and are apt to be knowledge.

But Audi's denial of any role, in grounding a testimony-based belief, for beliefs about the speaker's credibility, leaves him open to Lackey's arguments against K-Nec. Since justified belief or knowledge that the speaker knows whereof she speaks is not part of the hearer's basis for her belief, on Audi's account, it then becomes unclear, and open to challenge, why it is necessary that the attester speak from knowledge, in order for knowledge to be acquired.

We have just seen how it is difficult to find a rationale for K-Nec, at the same time as holding KPNDS—since KPNDS requires a conception of testimony-based belief like Audi's, on which no normative commitment to the fact that the attester speaks from knowledge on the part of the hearer is involved This leaves him vulnerable to Lackey's counterexamples. In contrast Fricker (2006) develops an account of testimony in which K-Nec is vindicated, but KPNDS is false. I finish by sketching this account, to show how K-Nec is given a rationale.

4. Knowledge from Trust in Testimony

On my own account of how testimony properly works, to enable the spreading of knowledge within a linguistic and cultural community, the core case is that of acquisition of epistemically entitled belief, apt to be knowledge, from trust in testimony.

When and why it is epistemically proper to believe what an attester tells one flows from the nature of her act. Telling someone that P is a case of asserting that P. The basic force of asserting that P to an audience is that the asserter vouches for the truth of P, she offers to her audience her word that P. Correlative with this (conventionally constituted and mutually known) import of the speech act of assertion is the fact that it is governed by the norm: one should assert that P only if one knows that P.[3] Assertion thus shares with promising a performative aspect. In promising, the promiser undertakes by her linguistic act to perform a certain action: she binds herself to do so, so that the promisee is entitled to rely on that undertaking, and may properly complain if the promiser subsequently does not keep her promise. Similarly, in asserting that P, the asserter gives her word that P entitles her audience to believe that P on the strength of her say-so, so that her audience may complain if P subsequently turns out to be false, or the asserter not to have known it to be true. This being so, to take an act of assertion, a telling, at face value, as being indeed what it purports to be, is to take it to be an expression of her knowledge by the attester.

But not all purported expressions of knowledge are in fact such. Folk psychology shows each of us how human nature is susceptible to the many motives driving deception, and to honest error. Given this way in which actual value may not be face value, I do not think that we are entitled, as Audi suggests and KPNDS encapsulates, to accept testimony at face value in the absence of empirical warrant for believing the speaker to be both sincere and competent about her topic—to be 'credible', as Audi says. On the contrary, if we are to trust an attester, take her assertion that P at face value, we must have empirical warrant to take her to have roughly this property: 'Not easily would she assert that P, unless she knew that P'. If one knows that antecedently about someone, and knows that she has asserted that P, then one has ground to conclude that she knows that P. From this it follows that P. This is the justifying ground available to justify a hearer's belief in what she is told, in virtue of whose availability to her it constitutes knowledge, on my account. (Roughly: 'She told me, and she wouldn't do so unless she knew.')

This is my own preferred internalist account of the needed grounds in virtue of which a belief based on trust in testimony—that is, on taking the speech act at its face value as an expression of knowledge—is entitled, and apt thereby to be knowledge. This account is developed and defended in Fricker (1994, 2002, 2004, 2006).

The brief summary just given suffices to show how I, but not Audi, have an explanation of why K-Nec holds. If, as I maintain, part of the recipient of testimony's basis for her belief is justified belief that the attester speaks from knowledge, then if that is false her own belief is based on a false premise, and so, even if true, is not itself knowledge. It follows that knowledge gained from trust in testimony, this core mechanism, is such that K-Nec holds for it. Knowledge from trust in testimony is necessarily knowledge at secondhand.

Moreover Fricker (2006) argues that this result is not achieved through a misleading definitional stop: knowledge from trust in testimony really is the core case of the speech act of telling working as it should.

What of Lackey's scenarios, supposed counterexamples to K-Nec? My response to each of them will be one of two options: either to deny that knowledge is gained, in such a case; or, alternatively, to agree that it is gained, but to offer an explanation of how this is so, which reveals it as being not a case of knowledge from trust in testimony, but instead some other, noncore mechanism in which the fact of the testimony features, but not in the central way characteristic of knowledge from trust in testimony, for which K-Nec holds.

Notes

1. This is Audi's preferred term, and I think a good one.

2. And, I think we should add for the best formulation of K-Nec, that the attester expresses that knowledge in her speech act. I am not sure whether counterexample-mongers could find a gap between a person's having knowledge that P while asserting that P, versus her expressing her knowledge that P in her speech act. This will depend, inter alia, on whether one can know that P without being conscious that one knows it. On my own view, sketched below, of how knowledge is often gained through testimony, the nature of the attester's speech act as an *expression* of her knowledge is crucial to this epistemic process.

3. It is a further question that is the direction of priority in explanation between these two things. In my own view, the most basic fact is that, in asserting that P, the speaker offers to her audience entitlement to believe that P on her say-so; the fact that one should assert only what one knows to be so follows from this.

References

Audi, Robert. 1998. *Epistemology* (London: Routledge).

——. 2006. Testimony, Credulity and Veracity. In Lackey, J. and Sosa, E. (eds.). *The Epistemology of Testimony* (Oxford: Clarendon Press).

Fricker, E. 1994. Against Gullibility, *Knowing from Words*, eds. Matilal and Chakrabarti. (Dordrecht: Kluwer) 125–61; reprinted in S. Bernecker (ed.) *Reading Epistemology: Selected Texts with Interactive Commentary* (Blackwell 2006).

——. 2002. Trusting Others in the Sciences: A Priori or Empirical Warrant?, *Studies in History and Philosophy of Science* 33: 373–83.

——. 2004. Testimony: Knowing Through Being Told. In *Handbook of Epistemology*, eds. Niiniluoto, Sintonen, and Wolenski (Dordrecht: Kluwer) 109–30.

——. 2006. Second-Hand Knowledge. *Philosophy and Phenomenological Research* 73.

Lackey, Jennifer. 1999. Testimonial Knowledge and Transmission *Philosophical Quarterly* 49:471–90.

On Being Justified in One's Head

TIMOTHY WILLIAMSON

In "An Internalist Theory of Normative Grounds" (2001), Robert Audi provides what his title promises.[1] His account is characteristically nuanced and ecumenical; it therefore constitutes an excellent basis for an appraisal that is not merely *ad hominem* of one kind of internalism.

With admirable generality, Audi treats the normative grounds for both belief and action. For simplicity, this paper concentrates on his account of the justification of belief. Its arguments, if sound, extend to the justification of action too. Some philosophers object to assimilating the justification of belief to the justification of action, because they take it to assume an implausibly voluntaristic conception of belief. I agree with Audi that we can assess beliefs normatively without assuming voluntarism (29). In condemning a belief as irrational, one does not imply that the believer can switch it on and off like a light.

Audi explains what he means by 'normative' in the case of belief:

> cognitive (epistemic) normativity is a matter of what ought to be believed, where the force of the "ought" is in part to attribute liability to criticism and negative (disapproving) attitudes toward the person(s) in question. (21)

He takes the concept of justification as a paradigmatically normative notion for belief (20).

Given this normative fix on justification, Audi defines his form of internalism about the justification of belief as "the view that what justifies a belief, i.e., the ground of its justification, is something internal to the subject" (21). He explains the 'internal' as:

> the (internally) accessible: that to which one has access by introspection or reflection, where introspection can be simply focusing on what is in consciousness and reflection can be as brief as considering a proposition. (ibid.)

He emphasizes that having access to something does not require accessing it; the former is a potential, the latter its actualization.

Audi's access internalism about justification does not obviously entail supervenience internalism about justification, according to which there is no difference in justification between internal duplicates, in a sense of 'internal' that is metaphysical rather than epistemological. Why cannot internal duplicates differ in what they can access in the special way? Differences between internal duplicates in the content of their mental states are arguably compatible with their privileged, non-observational access to their own contents.[2] I introspect that I have reason to believe one thing, while my twin introspects that he has reason to believe another. However, Audi does take himself to be committed to some form of supervenience internalism about justification. That commitment is deeply connected to his motivation for access internalism.

According to Audi:

> Internalism is motivated by at least two ideas. One is that what justifies a belief is somehow available to the subject—through consciousness or reflection—to use in justifying it; the other is that the view explains why, if our beliefs are suitably based on internal grounds, then even if they are false, say because of bad luck or even because a Cartesian demon has caused us to hallucinate in such a way that our beliefs are "imperceptibly" false, we would remain justified in holding them. (22–23)

If supervenience internalism about justification fails, internal duplicates can differ in justification; I may have a justification that my twin in the demon world lacks. But then we seem to lose the explanation that the second motivating idea requires. So Audi endorses a form of supervenience internalism about justification.

This paper is structured as follows. Sections 1 and 2 consider Audi's clarifications of internalism in response to two familiar objections, and argue that his responses are insufficient to prevent those objections from doing serious damage to its plausibility. Sections 3 and 4 appraise his positive case for internalism, and show how externalism can do justice to all his considerations.

1. As Audi sees it, externalism about the content of mental states presents a prima facie challenge to his internalism about justification. According to content externalism, two internal duplicates may differ in what they believe or have other propositional attitudes toward. In the standard example, Oscar on Earth in 1750 believes truly that there are pools of water. Oscar does not believe that there are pools of twater, where twater is a liquid on counterfactual Twin Earth with the same superficial characteristics as water but an utterly different underlying nature, for in Oscar's world there are no pools of twater, nor are there grounds for ascribing a false belief on that matter to him. He has no specific beliefs as to the underlying nature of water, except that it has one. By contrast, on counterfactual Twin Earth, Oscar's duplicate Twin Oscar believes truly that there are pools of twater. Twin Oscar does not believe that there are pools of water, even though he is in exactly the same internal states as Oscar, for in Twin Oscar's world there are no pools of water, nor are there grounds for ascribing a false belief (in his world) on that matter to him. He has no specific beliefs as to the underlying nature of twater, except that it has one.[3] Of course, some philosophers will dispute this description of the case, but Audi does not. His strategy is to argue that such externalism about content is compatible with his internalism

about justification. It is thus legitimate to assume the content-externalist description of the case in discussing Audi's claim that it is compatible with his internalism.

It is not in dispute that we can pick an example in which Oscar's belief that there are pools of water is justified. Perhaps he is swimming in one. Thus Oscar has the justified belief that there are pools of water. But Twin Oscar lacks a justified belief that there are pools of water, because he lacks the belief that there are pools of water. Thus Oscar and Twin Oscar differ in their justified beliefs, even though they are internal duplicates. Likewise, of course, Twin Oscar has the justified belief that there are pools of twater, while Oscar lacks a justified belief that there are pools of twater, because he lacks the belief that there are pools of twater: that is just another difference in justified belief between Oscar and Twin Oscar.

In response to such concerns, Audi first considers the postulation of 'narrow contents' of belief in common between Oscar and Twin Oscar (32). Such contents remain the pious hope of some internalists, but there is little evidence that we believe any such things. Moreover, it is not clear how they would meet the original challenge. For even if Oscar and Twin Oscar do have some justified beliefs in common, the externalist argument was that since they do not have *all* their justified beliefs in common, justified belief does not supervene on internal states. The postulation of additional contents would not undermine the point that Oscar has, while Twin Oscar lacks, the justified belief that there are pools of water. Merely restricting the internalist claim to the justification of belief in narrow contents threatens to be ad hoc, for no such restriction is built into Audi's positive case for internalism about justification.

Audi prefers a different response, one that does not offer a hostage to fortune by postulating narrow contents:

> Whether the internalist can find some proposition that both duplicates actually believe is not crucial; the point is that if the two are internally identical, then what they are justified *in believing* is the same. (32)

One may not believe everything that one is (or would be) justified in believing. In terminology that Audi also uses, his internalism is a doctrine about what one has justification *for* believing, not about what justified beliefs one has (43n2). That seems to fit his characterization of his theory as one about normative *grounds*.

But can that distinction bear the burden that Audi imposes on it? Since Oscar and Twin Oscar are internally identical in Audi's sense, Audi is committed to saying that what they are justified in believing is the same. Given his disclaimer about narrow contents, the phrase "what they are justified in believing" should be read as applying to ordinary belief contents, not just to postulated narrow contents. Therefore, since (by hypothesis) Oscar is justified in believing that there are pools of water, Twin Oscar is also justified in believing that there are pools of water (although he does not in fact believe that there are pools of water). Since there are no pools of water in Twin Oscar's world, there is a false proposition that Twin Oscar is justified in believing. Similarly, since Twin Oscar is justified in believing that there are pools of twater, Oscar is also justified in believing that there are pools of twater (although he does not in fact believe that there are pools of twater). Since there are no pools of twater in Oscar's world, there is a false proposition that Oscar is justified in believing.

Twin Earth scenarios can be multiplied indefinitely. Thus Audi's proposal seems to imply that for very many of our justified beliefs there are masses of closely related false propositions that we are justified in believing, although we do not in fact believe them. The problem here for Audi is not the general idea that one can be justified in believing a false proposition, for like many epistemologists he will take that as obvious. Rather, the problem is that justification for believing something false does not come anything like as cheap as Audi's internalism makes it come. If one is justified in believing a falsehood, one's epistemic situation is in some way misleading. Yet nothing that we assumed about Oscar's epistemic situation seemed to imply that it was in any way misleading at all. He just went swimming in some rock pools. Although someone from a twatery environment who was suddenly switched with Oscar without realizing it might thereby be placed in a misleading epistemic situation, that does not imply that *Oscar's* epistemic situation is misleading; whether one is in a misleading epistemic situation depends in part on one's past history. Of course, Oscar's total evidence is in many respects incomplete. For example, he has no evidence that water is H_2O (he does not even have the concept H_2O). But incomplete evidence need not be misleading, for it can justify agnosticism. He has no evidence that water is not H_2O. In any case, Oscar has quite enough ordinary evidence to justify him in believing that there are pools of water. It is deeply implausible to suggest that the very same evidence also justifies him in believing any number of false propositions of a similar form.

Could Audi retreat to the weaker internalist thesis that although internal duplicates may differ in what *gets justified*, they do not differ in what *does the justifying*? On such a view, Oscar's internal state would justify him in believing that there are pools of water but not in believing that there are pools of twater; the very same internal state would justify Twin Oscar in believing that there are pools of twater but not in believing that there are pools of water. One trouble with the envisaged fallback position for Audi's purposes is that it does not fit a major part of his original motivation for his theory, namely, that it explains the (supposed) constancy of justification in and out of skeptical scenarios. Given that you and your twin in the demon world are internal duplicates, the weakened sort of internalism just envisaged implies that you and your twin have the same justifiers, but it does *not* imply that those justifiers justify the same propositions; thus it does not imply that your twin is justified in holding the very beliefs that you are justified in holding. So it does not explain the (supposed) constancy of justification in and out of skeptical scenarios: it lacks the explanatory capacity that Audi claims as a crucial advantage of his internalism. If you and your twin in the demon world are held to be justified in believing the same propositions because you are internal duplicates, then Audi's explanandum in effect just is his internalist principle that "if the two are internally identical, then what they are justified *in believing* is the same," whose implausible consequences were marked above.

Could Audi restrict his claims to propositions that both subjects can grasp? For it might be argued that Oscar cannot grasp the proposition that there are pools of twater, and Twin Oscar cannot grasp the proposition that there are pools of water. But that feature is not essential to the problem. We can add to the example that a traveler once showed Oscar a tiny phial of twater and told him (truly) that the liquid in it was called 'twater', had similar superficial characteristics to water but an underlying nature that was utterly different (in unspecified ways), and occurred only in

minute quantities, not pools (in this world); Oscar believes what the traveller told him and has no beliefs from any other source about twater. Similarly, a traveler once showed Twin Oscar a tiny phial of water and told him (truly) that the liquid in it was called (confusingly for us) 'twater', had similar superficial characteristics to twater but an underlying nature that was utterly different (in unspecified ways), and occurred only in minute quantities, not pools (in that world); Twin Oscar believes what the traveler told him and has no beliefs from any other source about water. Thus Oscar can think of twater and Twin Oscar can think of water. Although Oscar associates twater with different descriptions from those that he associates with water, and likewise for Twin Oscar, externalists about content will typically hold that in such circumstances Oscar can grasp the proposition (false in this world) that there are pools of twater while Twin Oscar can grasp the proposition (false in that world) that there are pools of water. After all, we share the belief that there are pools of water with people whose beliefs about water are quite different from ours. Nevertheless, Oscar is justified in believing that there are pools of water but not in believing that there are pools of twater, while Twin Oscar is justified in believing that there are pools of twater but not in believing that there are pools of water.[4] Thus the difference in justification between internal duplicates can arise even for propositions that both grasp. Indeed, we can modify the example so that Oscar, in addition to his justified true belief that there are pools of water, has an irrational, unjustified, false belief that there are pools of twater, while Twin Oscar, in addition to his justified true belief that there are pools of twater, has an irrational, unjustified, false belief that there are pools of water. Thus the difference in justification between internal duplicates can arise even for propositions that both believe.[5]

We can imagine various further maneuvers on Audi's behalf in response to such cases. But they will typically be maneuvers of very much the sort to which internalists about content are forced to resort. They will involve the postulation of something like a taxonomically basic level at which content is narrow, so that the narrow content underlying Oscar's justified belief that there are pools of water is the same as the narrow content underlying Twin Oscar's justified belief that there are pools of twater, and different from the narrow content underlying Twin Oscar's unjustified belief that there are pools of water. The obstacles to that content-internalist program are forbidding.

Audi could still claim that Oscar and Twin Oscar are equally justified in their beliefs, although with respect to different beliefs: our overall normative assessment of them as believers is the same. That would be a long step back from Audi's assertion that "*what* they are justified in believing is the same" (32, italics moved). Moreover, we shall see in section 4 that it is very doubtful that our overall normative assessment of internal duplicates as believers always is the same.

Contrary to Audi's claims, his internalism about justification is in serious conflict with externalism about content. Given the strength of the arguments for externalism about content, that is bad news for his internalism about justification.

2. A different challenge to internalism about justification, forcefully pressed by Alvin Goldman (1999), arises from bare factual memory. Many of our factual memories come without any particular supporting phenomenology of memory images or feelings of familiarity. We cannot remember how we acquired the information, and

it may be relatively isolated, but we still use it when the need arises. Although few if any memories stand in total isolation from the rest of our conscious lives, very many memories are too isolated to receive impressive justification from other internal elements. One may believe that one's memory has been fairly reliable in the past; of course, its actual past reliability is irrelevant to justification on an internalist theory such as Audi's. In such cases, there is a shortage of internal justifiers for the memory-based belief.

Audi's response is to suggest that in such cases we have knowledge without justification (30–31), for he combines internalism about justification with externalism about knowledge (26).[6] For example, I know that Trincomalee is in Sri Lanka, but I am not justified in believing that Trincomalee is in Sri Lanka. Is that treatment adequate?

Audi's talk of knowledge without justification may sound acceptable if one thinks of it as knowledge that the subject cannot back up with a discursive defense; it is plausible that we have such knowledge. But Audi's primary characterization of justification is normative rather than dialectical. Thus in postulating knowledge without justification, he is saying something negative about the normative status of the knowledge. When trying to elucidate the normativity of justification, Audi suggests:

> Justified beliefs are those one is in some sense in the right in holding: holding them is normatively appropriate given one's sensory impressions, rational intuitions and other internal materials. In the language of desert, one does not deserve criticism (from the point of view of the effort to reach truth and avoid falsehood) for holding them. (29)

Therefore, since my belief that Trincomalee is in Sri Lanka is not justified, it is not one of those beliefs that I am 'in the right in holding'; that is, I am not in the right in holding it. I do 'deserve criticism (from the point of view of the effort to reach truth and avoid falsehood)' for believing that Trincomalee is in Sri Lanka, even though I know that Trincomalee is in Sri Lanka. Such claims seem quite implausible. One ought to give up a belief once one realizes that one is not in the right in holding it. But it would be silly for me, having read Audi, to *give up* my belief that Trincomalee is in Sri Lanka. That would be to give up some of my knowledge. If believing that Trincomalee is in Sri Lanka is not 'normatively appropriate given my sensory impressions, rational intuitions and other internal materials', that merely goes to show that the internal (as Audi understands it) is the wrong place to look for the justification of my belief.

With an eye to the case of young children's beliefs, Audi insists that "not justified" does not imply "unjustified" (31). But the word "unjustified" has not been used here. Rather, Audi has sketched a necessary and sufficient condition for justification in normative terms: where he denies justification, he should therefore be taken to hold that his normative condition does not obtain. The trouble is just the implausibility of the claim that the normative condition does not obtain in cases of bare factual memory knowledge.

Audi's proposal also faces a structural difficulty. Suppose that S knows p without being justified in believing p. It may easily happen that at the same time S has a justified belief in some conditional $p \rightarrow q$ without knowing $p \rightarrow q$, for Audi denies that justified belief entails knowledge.[7] Now S applies modus ponens to $p \rightarrow q$ and p, competently deduces q, and comes to believe q on that basis alone. What is the status of her belief in q? It seems to be good in some way, since it is held on the basis

of competent deduction from premises each of which was either known or believed with justification. However, S does not know q, for S does not know one of the premises on which her belief in q is essentially based. Indeed, q may even be false, for S may have a justified false belief in the conditional. Equally, her belief in q is not justified, for S lacks justification for one of the premises on which her belief in q is essentially based. After all, given Audi's internalism about justification, it is irrelevant to whether S is justified in believing q that she has knowledge where she lacks justification. Thus her belief in q constitutes neither knowledge nor justified belief. At the very least, Audi seems to be missing an epistemological category with which to classify such cases. Since he structures his epistemological discussion around the two contrasted categories of internalist justification and externalist knowledge, something is amiss with the structure of his epistemology. By separating the internal and external perspectives in this way, he makes it hard to see what is good about S's belief in q. If one takes the internal perspective, one cannot see what is good about S's belief in q because S was not justified in believing p. If one takes the external perspective, one cannot see what is good about S's belief in q because S did not know $p \rightarrow q$.

The idea of knowledge without justification should strike us as anomalous, at least when 'justification' is understood in the normative way that Audi sketches. Knowing p is the central, least problematic case of normative appropriateness in believing p. Although one may know p without being able to make explicit an elaborate discourse of justification for p, it would be naive to regard such an ability as necessary for normatively appropriately believing p, and Audi does not so regard it (27). However tempting it may be for some epistemologists to think that those who lack the capacity for epistemological reflection are not entitled to believe anything, the thought is a mere professional deformation. The primary problem of justification concerns beliefs that do not constitute knowledge.

Audi's misalignment of the concepts of knowledge and justification results from his combination of externalism about knowledge with internalism about justification. By concentrating the normative aspect of the appraisal of beliefs into his conception of justification and the truth-tracking aspect into his conception of knowledge, he arrives at an overall view on which justification and knowledge look more independent of each other than they really are. But the two aspects are not orthogonal dimensions: 'from the point of view of the effort to reach truth and avoid falsehood', knowledge is not normatively neutral.

3. We have seen that Audi's attempts to reconcile his internalism about justification with content externalism and the phenomenon of bare memories have committed him to some exceedingly implausible claims. Only very strong arguments could justify persisting with such a view. It is time to examine Audi's positive case for internalism. We start with his first motivating idea, that "what justifies a belief is somehow available to the subject—through consciousness or reflection—to use in justifying it" (22). In the section after, we assess how much changes when we add the second motivating idea, that our justified beliefs remain justified in skeptical scenarios.

Properly understood, the claim that only what is available to a subject justifies her belief is an attractive one, for we may interpret the question of justification precisely as the question of whether the belief is a good one for the subject to hold,

given what is available to her. 'Availability' here means some sort of epistemic accessibility. In a general sense, whatever the subject is in a position to be aware of is epistemically accessible to her. On that reading, the claim that only what is available to a subject justifies her belief follows from the claim that only what the subject is in a position to be aware of justifies her belief; I will not challenge the latter claim.[8]

Audi does not intend his claim in that general sense. Rather, he has in mind a special sort of epistemic accessibility, which he qualifies parenthetically as 'internal'. He explains it thus:

> To have (internal) access to something is either to have it in consciousness or to be able, through self-consciousness or at least by reflection, whether introspective or directed "outward" toward an abstract subject matter, to become aware of it in the (phenomenal) sense that it is in one's consciousness. (21–22)

I have epistemic access in the general sense to the presence of a computer screen before me, because I know by sight that this computer screen is before me. Have I internal access, in Audi's sense, to the presence of the computer screen before me? One might argue that I have, because I am conscious of the presence of this computer screen before me, and therefore have it 'in consciousness' in some sense. However, Audi's whole discussion makes it clear that he does not intend his words to be applied so widely: if they are, nothing distinctive remains of his internalism. Instead, he takes for granted that the contents of consciousness are confined to more traditional items, such as mental images. On his view, an experience as of a computer screen before me may be 'in consciousness', but not the computer screen itself. To vary the example, I have epistemic access in the general sense to Trincomalee's being in Sri Lanka, because I know by memory that Trincomalee is in Sri Lanka. On Audi's view, I lack internal access to Trincomalee's being in Sri Lanka, for otherwise such cases would not exemplify knowledge without justification. Although I consciously remember that Trincomalee is in Sri Lanka, what is in my consciousness in his sense is not that Trincomalee is in Sri Lanka, but only that it seems to me that Trincomalee is in Sri Lanka (or that some related narrow content seems to me to be the case).

It is far from obvious that even introspection is restricted in the way that Audi requires. Pretheoretically, it is natural enough to say that I can introspect that I am occurrently aware that this computer screen is before me, or that I occurrently remember that Trincomalee is in Sri Lanka. Since 'aware' and 'remember' are factive, the deliverances of introspection would not then be neutral in respect of the external environment. Audi does not consider this possibility.

Suppose for the sake of argument that Audi has somehow defined the phrase 'internal access' so that it applies to just the sorts of cases that he envisages, even if his actual definition can be read otherwise. The phrase 'epistemic access' will henceforth be used for epistemic access in the more general sense. The question immediately arises: why accept Audi's restriction of what justifies the subject's belief to what she has internal access to, when that excludes all sorts of other truths to which she has epistemic access? What is so special about internal access with respect to justification?[9] Unfortunately, Audi does little to address this question. Yet without a proper answer, his first motivating idea for internalism looks arbitrary and question begging. On the face of it, I *can* use truths to which I have epistemic but

not internal access to justify my beliefs. For example, my belief that my computer has not been stolen may be justified by inference from premises one of which is that this computer screen is before me; my belief that Trincomalee is in Asia may be justified by inference from premises one of which is that it is in Sri Lanka.

Audi might reply that what *ultimately* justifies my belief that my computer has not been stolen does not include the premise that this computer screen is before me, and that what *ultimately* justifies my belief that Trincomalee is in Asia does not include the premise that it is in Sri Lanka. To restrict his internalism to ultimate justification would already be a significant concession, since his paper advertises a theory of justification *tout court*, not merely a theory of ultimate justification. Moreover, we have as yet been given no reason to think that even ultimate justifiers must be internally accessible. Another reply on behalf of Audi in the same spirit starts from the other end: the subject has epistemic access *primarily* to internal elements of actual or potential consciousness and only secondarily to premises such as that this computer screen is before me or that Trincomalee is in Sri Lanka. He would then have to explain why we need primary access to justifiers; why is secondary access insufficient? Moreover, we have as yet been given no reason to think that we do have primary access only to internal elements of actual or potential consciousness. To invoke such a conception of ultimate justification or primary access as rooted in elements internal to actual or potential consciousness would involve Audi in explaining and defending deeply contentious epistemological premises of a kind that would obviously beg the question in an argument for internalism about justification. Since he makes no attempt to address the difficulties, it may be uncharitable to attribute such replies to him.

Internalists might try to support the talk of ultimate justification by arguing that if serious doubts arise as to whether I really am aware that a computer screen is before me or really do remember that Trincomalee is in Sri Lanka, I am likely to fall back on defensive remarks such as 'At least, it seems to me that I am aware that a computer screen is before me' or 'At least, it seems to me that I remember that Trincomalee is in Sri Lanka'. But it is a fallacy to assume that retreats in the face of doubt always reveal a preexisting structure of justification. Someone may be simultaneously disposed to retreat to premises about appearances if put under pressure by idealists about the external world and to retreat to premises about brain scans if put under pressure by eliminativists about the mind. In responding to a doubt that we are willing to take seriously, we look for ground that it does not undermine, but where that ground is depends on the doubt. That we can be made to retreat to a place does not show that it is where we started from.

We can sharpen our sense of the dialectical position by briefly considering an externalist rival to Audi's internalism about justification. On the view that I defend, what justifies beliefs is the subject's total evidence, which is the total content of the subject's knowledge (Williamson 2000). A belief that constitutes knowledge is ipso facto justified. A belief that does not constitute knowledge is justified, to whatever degree it is justified, by its relation to beliefs that do constitute knowledge. Thus chains of justification can terminate in knowledge of any kind; they do not all lead back to *recherché* internal elements of actual or potential consciousness. Of course, knowing something depends on various preconditions: I know that this computer screen is before me, or that Trincomalee is in Sri Lanka, only if all sorts of other things are true of me too. But those preconditions for knowledge need not themselves be justifiers in the sense in

which subjects must have access to their justifiers. In the most straightforward way, this externalist account satisfies the requirement that subjects have epistemic access to their justifiers, for their justifiers are truths that they know. Although the account does not satisfy Audi's requirement that subjects have *internal* access to their justifiers, that simply makes more pressing the question: why is epistemic access insufficient?

Someone might doubt that such an externalist theory of justification really does satisfy even the epistemic access requirement. For in a skeptical scenario we lack knowledge without being in a position to know that we lack it; less obviously, we sometimes have knowledge without being in a position to know that we have it. Consequently, given that one's total evidence is one's total knowledge, something can be excluded from one's evidence even though one is in no position to know that it is excluded from one's evidence, and something can be included in one's evidence even though one is in no position to know that it is included in one's evidence. But what these points show is that something can be a nonjustifier although one is in no position to know that it is a nonjustifier, and something can be a justifier although one is in no position to know that it is a justifier. These points do not show that something can be a justifier although one is in no position to know *it*. If one is in no position to know it then one does not know it and so, on this externalist account, it is not one of one's justifiers. One can have epistemic access to a justifier without having epistemic access to the further truth that it is a justifier.

Of course, someone might want to impose the additional requirement that one must be in a position to know of any given item whether it is one of one's justifiers. But Audi does not impose that further requirement, and wisely so. For it implies that the condition that p is a justifier and the condition that p is not a justifier are *luminous*, in the sense that whenever the condition obtains, one is in a position to know that it obtains. Elsewhere, I have argued that only trivial conditions are luminous (2000, 93–113). The condition that p is a justifier and the condition that p is not a justifier carry substantive information about one's current epistemic state on any serious view of justification; for example, on Audi's view, they may concern whether one is currently experiencing a mental image of a certain kind. Therefore those conditions are not trivial in the relevant sense. So they are not luminous. Neither internalism nor externalism can meet the further requirement. Audi provides no hint of an objection to the anti-luminosity argument. Thus the further requirement should not be imposed on his behalf.

The question remains: in motivating his internalism, what entitles Audi to invoke his requirement of internal access on justification, rather than the requirement of epistemic access that the externalist theory satisfies? At one point, he suggests an explanation for the supposed internal grounding of normativity across diverse domains: our practices of justification "must have criteria usable in the everyday contexts in which we engage in these practices. One must, given how we are built, apply them from the inside" (40). But the requirement of epistemic access already secures usability in everyday contexts. If we have knowledge then we can use it in justification. Of course, we will sometimes be mistaken as to whether something is part of our evidence, but no everyday practice requires infallibility about anything, and even Audi refuses to endorse the claim that our self-ascriptions of mental states are infallible (44n11). In fact, it is internalism, not externalism, that risks losing touch with usability in the everyday contexts in which we engage in practices of justification. For justification is typically a *social* practice: we try to justify

our beliefs to others in response to their challenges. But if what I use in justifying my belief to you is internal to me, how much use will it be to you? When George W. Bush and Tony Blair were expected to justify their belief that there were weapons of mass destruction in Iraq, what was needed was publicly verifiable evidence, not mental images or feelings of conviction. Of course, internalists will concoct a story about public justification as consisting in the causation by the justifier of suitable internal elements in the consciousness of the hearer, but no such story is motivated by the ordinary criteria for public justification. An externalist account of justification as rooted in the content of knowledge in general is far better adapted to our ordinary practice of justification.[10]

My present concern is not to develop the arguments for a particular externalist theory of justification in detail, since I have done so elsewhere. The immediate point is simply that at least one externalist theory of justification meets at least as strong an access constraint as may legitimately be assumed without serious argument. In imposing his requirement of internal access, Audi starts from an assumption that lacks the pretheoretic plausibility needed if his case for internalism is to do more than preach to the converted. We must therefore turn to Audi's second motivating idea for internalism.

4. According to Audi, we retain our justification in cases of perfect hallucination. He supports the point by appeal to normative features of our situation:

> I am in no way at fault for believing what I do, nor do I deserve any criticism (at least on the non-skeptical assumption that we may generally trust our senses in this way). Far from it. I am like a surgeon who skillfully does all that can be expected but loses the patient. There I should feel regret, but not guilt; I should explain, but need not apologize; and when we know what my evidence was, we approve of what I did. We consider it reasonable. (23)

This passage mixes various normative claims, some more plausible than others. It does scant justice to the complexity of our normative thinking.

We can readily agree with Audi that the victim of a paradigmatic skeptical scenario is not to be *blamed* for forming false beliefs under the misapprehension that they constitute knowledge. The subject has a cast-iron excuse for having formed those beliefs. But the subject in the corresponding nonskeptical scenario whose beliefs constitute knowledge needs no excuse for having formed those beliefs; there is nothing to excuse. That the two subjects are equally blameless does not imply that there is no normative difference between them. The subject who blamelessly forms false beliefs has not done as *well* as the subject who gains knowledge.

Suppose that the patient in Audi's simile would not have died if the surgeon had not operated, just as error would have been avoided in the skeptical scenario by suspension of belief. Of course, the surgeon followed the best available guidelines in deciding to intervene, and operated with state-of-the-art skill and technology. Nevertheless, it turns out that it would have been better for the patient if the operation had not been performed. To say, in Audi's words, that we *approve* of what the surgeon did hardly captures the complexity of our normative response to the case. We absolve the surgeon of all blame, while also thinking that in some sense the decision to operate turned out to be wrong. We praise the surgeon less than we would have done had the operation saved the patient's life rather than causing her death.

Audi concedes that the skeptical and nonskeptical scenarios can differ norma-tively. For the content of an obligation can be external:

> Suppose I promise to care for a child. Is it enough that I do all I can, so that if a Cartesian demon makes me hallucinate success, while the child is in fact left alone, I have fulfilled my obligation? To say so would be to parlay excusability for not fulfilling an obligation into its actual fulfilment. (37)

Excusable failure is not normatively equivalent to success. Nevertheless, Audi main-tains "the *content* of my obligation can be external while its grounds are internal, just as the content of my belief about the water in my glass can be external though the grounds of that belief are internal" (37–38). Given the results of section 1, this comparison with externalism about content will not help Audi. As for the claim that the grounds of the obligation are internal, it is at best misleadingly expressed: how can the grounds of my obligation to care for the child be independent of whether the child exists and of whether I made a promise *to someone* to protect the child? Moreover, if obligations can differ between the skeptical and nonskeptical scenarios, why shouldn't epistemic obligations do so in particular? For example, if I have an obligation to believe only what I know, then I fulfill that obligation in the nonskep-tical scenario but not in the skeptical scenario, and my beliefs differ normatively between the two scenarios as a result.

Presumably, Audi will restrict 'justification' to a special sort of normativity with respect to which the skeptical and nonskeptical scenarios are equivalent, although that is already to qualify the apparent claims of his paper. Furthermore, if he is not merely to have made an uninteresting stipulation about the use of the word "justification," some argument is needed that there really is an important normative respect in which any two such scenarios are equivalent. The normative distinctions that we make between them show that the existence of such a respect cannot simply be taken for granted.

One might regard the normative differences between the two scenarios as something like moral luck (Nagel 1976; Williams 1976). Then the challenge to the internalist would be to demarcate a luck-free zone as the realm of justification. Suggestive though this way of putting the matter may be, the unrefined notion of epistemic luck is too contested to figure in a clear statement of the issue. Consider knowledge reached by a thoroughly reliable process outside the subject's ken. Externalists will tend to say that the truth of the subject's belief is not a matter of luck, because the process is so reliable. Internalists will tend to say that the truth of the belief is a matter of luck, because in some skeptical scenario the belief is false and merely appears to have been reached by that process.[11] Let us stick to the ques-tion of whether Audi can isolate an important normative respect in which a skepti-cal scenario is always equivalent to the corresponding nonskeptical scenario.

It may seem that the internalist can easily meet the challenge by appealing to Audi's first motivating idea for internalism: access. The proposal would be that the two scenarios are alike *in all normative features accessible to the subject*: more precisely, the subject in the nonskeptical scenario and the subject in the skeptical scenario have access to exactly the same normative truths. For agents have to act on the basis of the limited information available. To assess the proposal, let us ask a prior question: do the subjects in the two scenarios have access to exactly the same truths quite generally?

If 'access' means epistemic access, then one should resist any temptation to answer in the affirmative, for that way leads to skepticism. If we are not skeptics, we will insist that the subject in a suitable nonskeptical scenario knows that the child is playing.[12] The subject in the corresponding skeptical scenario does not know that the child is playing, because the child is not playing. Thus the subject in the nonskeptical scenario has access to something to which the subject in the skeptical scenario lacks access. But once it is granted that the two scenarios differ in what features are accessible to the subject, what can stop them from differing in what *normative* features are accessible to the subject? For example, the subject in the nonskeptical scenario may know that he has fulfilled his obligations. The subject in the skeptical scenario does not know that he has fulfilled his obligations, because he has not fulfilled them.

If one's total evidence is one's knowledge, then the two scenarios differ in the subject's evidence. That the child is playing is part of the subject's evidence in the nonskeptical scenario but not in the skeptical scenario. Of course, the subjects in the two scenarios share *some* evidence. For example, that it perceptually appears that the child is playing may be part of the subject's evidence in both scenarios. On this view, the difference is that in the nonskeptical scenario the subject's total evidence *entails* that the child is playing, whereas in the skeptical scenario the subject's total evidence does not entail that the child is playing, although it does (misleadingly) make it probable that the child is playing. Consequently, the belief that the child is playing is *more justified* in the nonskeptical scenario than it is in the skeptical scenario, even though in both scenarios it has considerable justification.[13]

Of course, Audi will not accept that theory of evidence and justification. But he cannot simply assume its falsity if he is to give a noncircular motivation for internalism about justification. It is not simply obvious that there is no subtle normative difference of the kind that it postulates in the status of beliefs between the two scenarios.

Audi will say that the two scenarios are exactly alike with respect to both the normative and the nonnormative features *internally* accessible to the subject, and *perhaps* he can explain a sense of 'internal' on which that is so. But that reply merely raises again the question to which section 3 found no answer: what is so important about internal access in Audi's special sense as opposed to any other kind of epistemic access? If agents have noninternal access to some information to which they lack internal access, what is the point of assessing their beliefs and actions as if they were based only on the information to which the agents have internal access? Why not assess their beliefs and actions in relation to all the information to which they have access?

A different strategy for Audi is to try to specify which dimension of normativity he intends to be internalist about in directly normative terms. The most promising candidate is the dimension of blame, since it was granted that the subject in a paradigmatic skeptical scenario is blameless, like the subject in the ordinary case. Audi's own analogy of the surgeon suggests that there are differences between the two scenarios in praiseworthiness. Thus 'justification' in Audi's sense would reduce to blamelessness. However, it is doubtful that this idea generalizes in the way that Audi requires, for several reasons.

First, if Audi concedes externalism about content, as he seems willing to do, then the blameworthiness of belief with a given content can vary across internal duplicates. For instance, section 1 sketched a story on which Oscar blamelessly

believes that there are pools of water but blameworthily believes that there are pools of twater, while Twin Oscar blameworthily believes that there are pools of water but blamelessly believes that there are pools of twater.

Second, if justification is defined in terms of blame, then presumably to be justified is simply to be blameless. Thus Audi's classification of bare memory knowledge as unjustified belief implies a category of blameworthy beliefs that constitute knowledge, which exacerbates the puzzles of section 2.

Third, paradigmatic skeptical scenarios correspond to nonskeptical scenarios in which the subject is blameless for her beliefs. The internalist must also hold that mildly blameworthy true beliefs in a nonskeptical scenario would have been no more blameworthy if false in the corresponding skeptical scenario. In Audi's simile, that is to say that the slightly clumsy surgeon who nevertheless performs the operation successfully would have been no more blameworthy if, with the same slight clumsiness, he had killed the patient. That is hardly a plausible claim. The surgeon who causes his patient's death through negligence has more to be blamed *for* than the equally negligent surgeon who saves his patient's life.

Distinguishing between being more to blame and being to blame for more, someone might suggest that the two surgeons are equally to blame overall, even though one of them is to blame for more than the other.[14] Presumably, how intensive a feeling of guilt it is appropriate to feel varies with how much one is to blame, whereas how extensive a feeling of guilt it is appropriate to feel varies with how much one is to blame for. But then, according to the suggestion, the intensity of guilt that it is appropriate for the surgeon who causes his patient's death through negligence to feel is no greater than the intensity of guilt that it is appropriate for the equally negligent surgeon who saves his patient's life to feel. That claim seems morally insensitive. Although the issues cannot be properly pursued here, it is at any rate very unclear that Audi's simile works for rather than against his argument.

Similar problems arise if we replace the concept of blame with other familiar normative concepts. Thus Audi is driven back to characterizing the relevant normative dimension in epistemic terms. We have already seen how hard it is to motivate his approach from an epistemological starting point. To conclude, we will consider a quite general problem for that approach, no matter what normative dimension Audi associates with justification.

5. On Audi's conception, the skeptical and nonskeptical scenarios are exactly the same in the relevant normative respect (whatever that is) because the difference between them is not internally available to the subject for normative use. Let us call situations 'indiscriminable' if and only if any difference between them is not accessible in the relevant way to the subject, and 'normatively identical' if and only if they are the same in the relevant normative respect. Then Audi's approach commits him to this principle:

Normative Tolerance: indiscriminable situations are normatively identical.

Is Normative Tolerance true? Imagine a sorites series of possible situations, each indiscriminable from its immediate neighbors, where the first and last members are very different, in particular in the relevant normative respect; the details will depend on the selected sorts of access and normativity. For example, we can

define such a series by taking successive situations with a very short constant time interval over a long process of very gradual change. Successive members of the series are indiscriminable because they are too similar for the subject in them to be aware by introspection or reflection that there has been any change at all, just as the subject in the skeptical scenario is not in a position to be aware by introspection or reflection that the situation is different from the nonskeptical scenario. Therefore, according to Normative Tolerance, each situation is normatively identical with its neighbors. But normative identity, like exact sameness in any other given respect and unlike indiscriminability, is transitive. Therefore the first and last members of the series are normatively identical, contrary to hypothesis.[15]

Since Normative Tolerance is false, there can be normative differences between indiscriminable situations. Once that is conceded, it is quite illegitimate to argue for internalism from the premise that there is no normative difference between the skeptical and nonskeptical scenarios. Thus Audi's argument for internalism fails. Given the implausible consequences of his internalism noted in sections 1 and 2, it should therefore be rejected.

In rejecting Audi's internalism about justification, we need not reject his insight that what does the justifying must be accessible to the subject, provided that we do not restrict the accessible to what is internally accessible in his special sense. We can put the natural externalist alternative by saying that p is available to you as a reason if and only if you know p.[16] That alternative applies to reasons for action, not just to reasons for belief. In a given case, the right reason for helping someone may be that the person needs help, not that you are in a certain internal state (the latter anyway sounds like an unattractively self-regarding reason). You can act for that reason only if you know that the person needs help. If it only seems to you that you know that the person needs help, then it only seems to you that you can act for that reason.

On a robustly realist conception of the normative, we will expect the distinction between appearance and reality to apply to every dimension of the normative, just as it applies to everything else (including appearances themselves). That does not make the normative any more mysterious or unknowable than it already was. On every dimension of the normative, there is a difference between the facts and what the subject is in a position to know of the facts, even though the subject is often in a position to know the most important facts. Once we have acknowledged that difference, we should no longer be tempted by the futile project of trying to carve out a dimension of the normative that is perfectly accessible to the subject, either internally or externally. We acknowledge that the normative is imperfectly accessible to the subject. That motivates an externalist theory of normative grounds.

Notes

DePaul; a conference in memory of Nikola Grahek at Belgrade University; the universities of Bologna and Oxford; King's College London.

2. Ludlow and Martin 1998 provides an introduction to the debate on whether content externalism is compatible with privileged access.

3. The example from Putnam 1973 is reconceptualized here as usual on the lines of Burge 1979, in terms of thought contents rather than just linguistic meanings. For those worried about snow and other H_2O that isn't liquid water, or dirty water that isn't pure H_2O, or H_2O in Oscar's body but not Twin Oscar's, the example can easily be changed.

4. For present purposes it matters neither whether Oscar is justified in believing that there are no pools of twater nor whether Twin Oscar is justified in believing that there are no pools of water.

5. The arguments of this paragraph, unlike the earlier ones about what Oscar or Twin Oscar is justified in believing but does not in fact believe, assume that belief contents are Russellian rather than Fregean: their individuation is coarse-grained (roughly: determined by the referents of the terms with which we express the contents) rather than fine-grained (roughly: determined by senses of those terms). However, the neo-Fregean form of externalism arguably also implies differences in justification between internal duplicates, since sameness of sense and therefore the validity of arguments is sensitive to external factors; see Campbell 1987–88.

6. Audi 1995 has more extended discussion of such cases.

7. Read → as material rather than strict implication.

8. To be in a position to be aware of something is to have some sort of potential for awareness of it; how remote the potential can be is an important question that Audi discusses, but is not of immediate concern here.

9. Goldman 1999 presses a similar question.

10. The remarks in the text should not be taken as endorsing the overintellectualized view that the justification of a belief, in the sense of its normative status, is determined by the believer's capacity to respond in dialectically effective ways to challenges. Beliefs that constitute knowledge are in good standing even if the slow-witted knower is easily outmanuevered in dialectic; see section 2. In the quoted passage, Audi is proposing an etiology, not an analysis, of our concept of justification.

11. See Pritchard 2005 for more on the notion of epistemic luck.

12. If you are a contextualist, pick a suitable context of utterance for that sentence.

13. The difference in justification is explained more rigorously within a framework of evidential probability at Williamson 2004, 313–14. In addition to the degree to which the subject's evidence supports the given belief content, there is also the question of how far the subject's belief is in fact causally based on the evidence. This is an issue for externalists and internalists alike (Audi notes it at 43n2). On the externalist view in the text, the subject in the skeptical scenario does worse in this respect too, since the belief is largely based on what the subject falsely regards as evidence.

14. See Zimmerman 1988; thanks to John Greco for pointing out the relevance of Zimmerman's distinction.

15. Closely related arguments are developed in much greater detail in Williamson 2000.

16. On a more liberal conception of availability, *p* is available to you as a reason if and only if you are in a position to know *p*.

References

Audi, R. 1995. Memorial Justification. *Philosophical Topics* 23: 31–45.

——. 2001. An Internalist Theory of Normative Grounds. *Philosophical Topics* 29: 19–46.

Burge, T. 1979. Individualism and the Mental. *Midwest Studies in Philosophy* 4: 73–121.

Campbell, J. 1987–88. Is Sense Transparent? *Proceedings of the Aristotelian Society* 88: 273–92.

Goldman, A. 1999. Internalism Exposed. *Journal of Philosophy* 96: 271–93.

Ludlow, P., and N. Martin, eds. 1998. *Externalism and Self-Knowledge*. Stanford, Calif.: CSLI.

Nagel, T. 1976. Moral Luck. *Aristotelian Society*. Suppl. no. 50: 137–55.

Pritchard, D. 2005. *Epistemic Luck*. Oxford: Oxford University Press.

Putnam, H. 1973. Meaning and Reference. *Journal of Philosophy* 70: 699–711.

Williams, B. 1976. Moral Luck. *Aristotelian Society*. Suppl. no. 50: 115–35.

Williamson, T. 2000. *Knowledge and Its Limits*. Oxford: Oxford University Press.

——. 2004. Replies to Commentators. *Philosophical Books* 45: 313–23.

Zimmerman, M. J. 1988. *An Essay on Moral Responsibility*. Totowa, N.J.: Rowman & Littlefield.

Audi on Nondoxastic Faith

WILLIAM P. ALSTON

I

In several publications[1] Audi has presented and defended a conception of propositional faith, faith that so-and-so, such as faith that God exercises providence over our lives, that is different from *belief* that so-and-so, and hence is termed *nondoxastic faith* (hereinafter 'NDF'). He argues that philosophers concerned with the cognitive side of religious faith have focused too exclusively on religious belief. He does not deny that belief has a place there, but he insists that religious propositional faith also takes at least one other form, a nondoxastic form, which he is concerned to delineate and to display its important role in the religious life.

At least one powerful critique of Audi on this point has sought to strike at the heart of his position by denying both that something other than belief can play the cognitive role in religious faith that Audi assigns it and that where Audi finds NDF it is really belief that, contrary to his asseverations, is what is within his sights.[2] My critique will be less radical than this. I agree with Audi that there is a kind of positive attitude toward religious, and other, propositions, that is distinct from belief, the religious form of which can do pretty much the same job in religion that religious belief does. Where I dissent from him is in denying that he has sufficiently identified this nondoxastic propositional attitude. It's not that he has said too little about it. On the contrary, as I shall show, he says a great deal. But (1) I am at a loss to find any propositional attitude that satisfies all of his specifications; (2) Even if I take some subset of those specifications, it seems to me to fall short of supplying an adequate identification of anything that satisfies all of them. They leave it mysterious what it is, if anything, that fits all of these descriptions; and finally, (3) I do have a candidate that we can characterize enough to be sure that there is such a thing and what thing it is, and that is a nondoxastic positive propositional attitude toward religious and other propositions. This is what I call *acceptance*, following Cohen 1992 though with some differences. My contention will be that acceptance does most of what Audi is looking for with his NDF, that it provides an unquestionably genuine propositional attitude

that can play the kind of role in religious faith that Audi takes his NDF to play, and without leaving us unable to understand just what is playing that role.

In the essays cited in note 1 Audi goes into aspects and types of faith much more fully than I will be discussing here, along with other matters. My sights here are narrowly focused on nondoxastic *propositional* faith, and I make no pretensions to be treating Audi's work on faith generally.

In this essay I will first, in section II, survey the various things Audi says in attempting to pin down NDF, showing that nothing fits all of what he says, and that even if we confine ourselves to a subset of them that something could fit, it is left unclear just what Audi thinks does so. Then in section III, I will briefly set out my conception of acceptance, with many references to Alston 1996 in which this conception of acceptance is presented in greater detail. In section IV, I will make the case for acceptance as the best candidate for a positive propositional attitude for religious and other propositions that is distinct from belief. And in section V, I will make a case for the thesis that my version of NDF can do almost as well as religious belief as the cognitive core of religious faith.

II

Audi's characterizations of NDF have remained remarkably consistent over the whole range of publications cited in note 1, though in Audi 1992 and 1993 he goes into much more detail than in the others. I will now list the main items, grouping them into several categories, with references to the places in the corpus where they can be found. Audi purports to tell us something of what NDF is as well as what it is not, as well as putting forward a number of looser characterizations, some of which concern differences of degree between NDF and belief. Obviously the items in the "what it is not" category fall far short of uniquely identifying NDF, though they provide important tests for candidates. In my initial run-through I will pretty much allow Audi's characterizations to speak for themselves, and I will have more to say about them in my critical remarks.

1. What NDF is not:
    a. Flat-out belief (H57, 79), (S74), (B322–23, 330).
    b. Feeling of certitude (H59), (B334).
    c. Being subject to mistake (H60), (T217).
    d. Belief + a positive evaluation of object (H62), (T218).
    e. Has a definitely *accepted* propositional object (H80), (T223).
    f. Implies existence of object (B326–27), (S77–78).
    g. Intellectual commitment to its propositional object (B336).
    h. Tentative belief that p (H61).
    i. Weak belief that p (H79), (T223), (B334).
    j. Belief that p is probably true (H83), (T227).

You might well wonder whether, given that Audi has denied that NDF is belief, it would be necessary to specify that it is not a belief of any of the sorts just listed. But when he is most explicit about denying that NDF is belief, he qualifies the

latter as flat-out belief. And he distinguishes flat-out belief that p, for a certain proposition, p, from, inter alia, tentative belief that p, weak belief that p, and belief that it is probable that p.

2. What NDF is:
   a. Incompatible with disbelief that p (H58), (T216), (S75), (B323, 332–33).
   b. Cognitive in having a propositional object (H85–86), (T225).
   c. Sufficient to qualify one as religious where the propositional object is religious (S75).
   d. Has a positive attitudinal component (T225), (B329).
   e. Involves a disposition to believe that p (H58), (T215–16), (B331).
   f. Can be strong, steadfast (H79), (T218).
   g. Involves conviction (H82–83), (T223).
   h. Requires beliefs other than a belief that p (H81–82), (T224).
   i. Is a positive attitude to a proposition (S75).
   j. Implies a cognitive trust (H80), (T224), (B332).

Although I have put these items under the heading of "What it is," the reader will note that except for i they all consist of properties, implications, involvements, and the like, rather than an identification of the attitude itself. This is shown grammatically by the fact that except for i all the items are formulated as predicates rather than as identity statements. They don't purport to say what this *is* that is incompatible with disbelief that p, is cognitive, involves a disposition to believe that p, and so on. And even i, though in the form of an identity statement, only tells us that it is a positive cognitive attitude to a proposition. But so is a belief that p. And so we are still in the dark as to what a nondoxastic positive attitude to a proposition *is*.

3. Weaker connections:
   a. Limits the degree of doubt that p (H59), (T216), (S75), (B323–24, 333).
   b. Is compatible with more doubt than belief of p (H60), (T217).
   c. Requires less for justification or rationality than belief (H61, 64, 79, 80–84), (T219, 222–23), (S74), (B336).
   d. A weaker disposition than with belief to avow the proposition, make inferences from it, and act on its basis (H80), (T224).
   e. If p turns out to be false, there is more tendency to be surprised with belief than with NDF (H60), (T217), (B333).
   f. The closer we are to belief, the less natural it is to use 'faith' (H58–59), (T216).

With these items it is even more obvious than with 2 that they are not telling us what NDF is, but rather specifying certain properties it, whatever it is, has.

4. Similarities between NDF and belief:
   a. Cognitive in having a propositional object (H61).
   b. More or less rational (H61), (T217).
   c. Influences behavior (H61).
   d. Varies in strength and in centrality (H61).

I take these to be the main clues Audi offers us as to what NDF is as he conceives it. Though this is far from my main point, I will indicate that it seems to be impossible for any one propositional attitude to fit all these specifications. The main tension concerns the relation of 2i (NDF is a positive attitude to a proposition) to 1e (NDF is not a definite acceptance of its propositional object) and 1g (NDF does not involve an intellectual commitment to its propositional object). The difficulty is to see how something can be a positive cognitive attitude to a proposition without involving an acceptance of the proposition and an intellectual commitment to it. In further explication of 1e Audi says of NDF that its propositional object is "projected." It is not clear to me either from Audi's account or otherwise just how we are to understand projection here. Clearly NDF is not just the entertainment of a proposition. That could not sensibly be called a positive attitude to the proposition. As I shall point out shortly, there is a variety of what could be called positive attitudes to a proposition p that are different from flat-out belief that p. And some of them do seem to be what 1e and 1g say that NDF is not.

For that matter, 2a (being incompatible with disbelief that p) also seems to be in conflict with 1e and 1g. If NDF involves no intellectual commitment to p and no definite acceptance of p, then why should it be incompatible with disbelief that p? What would prevent it from being conjoined with disbelief? But one reason for my not putting a lot of weight on this point is that it is not clear how 1e and 1g are to be construed. So let us suppose that they are to be understood as to be compatible with 2a, giving the latter the controlling voice here. Or alternatively, let's abandon 1e and 1g and work with what remains. Where does that leave us?

That leaves us with my chief dissatisfaction with Audi's purported identification of NDF, namely, that he has not succeeded in telling us *what it is*. As briefly set out above, he has told us quite a lot about its properties—how it is similar and dissimilar to belief, what it is and is not compatible or incompatible with, some of the dispositions it does and does not involve, some of what it is associated with, and the relative strengths of various properties. But we are still left in the dark as to what all this is true of. The only substantial hint we get on this is that NDF is a positive cognitive attitude toward a proposition. One may well think that Audi's specifications contain other (partial) hints as to what NDF is. Consider 2d (has a positive attitudinal component). This is certainly part of what distinguishes faith, religious and otherwise, from (mere) belief. If I felt that a certain outcome would be a bad thing, it would be, at best, highly misleading to say that I had faith that it would occur. If I thought it would be a good thing that Susan should finish her dissertation during this academic year (but had no flat-out belief that she would but also no flat-out belief that she wouldn't and some considerable reason for thinking she would), I could properly say that I had faith that she would. But if the same epistemic conditions held with respect to my cognitive attitude toward her *failing* to finish this year, then it would be out of order for me to say that I have *faith* that she won't finish; and out of order precisely because this is an outcome I take to be a bad thing. So 2d provides a condition that is necessary for faith, whether religious or not and whether nondoxastic or not. And so, like all the items on the above lists, it puts a constraint on what can count as a NDF. But whether it gives us even a partial specification of what NDF *is* depends on a thorny question as to the individuation of cognitive

attitudes. In particular the question is as to whether the favorable attitude toward the realization of the propositional object is a part of what the NDF is, or rather something that must also obtain in order that the propositional attitude in question count as a case of *faith that*. In either case the favorable attitude could be a necessary condition for the attitude in question being a case of faith. And so long as the latter alternative is not excluded, it is not clear that 2d tells us something of what NDF *is*. There will be a further discussion of the individuation issue later.

To return to the main topic of discussion, 2i could be satisfied by a variety of nonidentical propositional attitudes, all of which could be termed (more or less) positive and that are distinct from flat-out belief. Here are some examples, assuming that Audi intends 2i to be construed in terms of being an intellectually or epistemically positive attitude to a proposition, p, not an otherwise evaluatively or affectively positive attitude.

   i. Thinking it likely that p.
  ii. Tentatively believing that p.
 iii. Weakly believing that p.
  iv. Being inclined to suppose that p.
   v. Taking p to be worth considering/investigating.
  vi. Taking p as a serious candidate for acceptance.

To be sure, Audi has already excluded i–iii from consideration by virtue of listing them in the "What it is not" category (1j, 1h, and 1i, respectively). But the reason this does not suffice to uniquely identify the positive propositional attitude that constitutes NDF is that no matter how many candidates are explicitly excluded, it is not clear that they exhaust the alternatives to NDF. It seems that there are always more candidates lurking just around the corner. For example, consider the following possible addition to the above list.

 vii. Taking p as a possible explanation for event e.

And so it goes.

Of course, there are many other constraints on what, by Audi's lights, counts as an NDF, and these may exclude all unwanted attitudes that are distinct from flat-out belief. Take, for example, the previously discussed 2d (has a positive attitudinal component). None of the items iv–vii from the above list satisfy that constraint, whether we take the constraint as requiring that the candidate satisfy the constraint as part of what it is, or only that the constraint also be satisfied along with the candidate. For example, being favorably disposed to the truth of p is in no way a necessary condition for taking p seriously as a candidate for acceptance. But again the suspicion still remains that there are other candidates that would satisfy that constraint in one way or another.

Return to the question of whether requirements for an attitude to count as faith would have to be satisfied by an aspect of what the attitude is, or whether it is sufficient that the attitude be accompanied by something else (typically some other attitude) that satisfies the requirement. I am taking the latter alternative, primarily because it puts lesser demands on the candidate for faith, and there seems to be no sufficient reason for imposing the more severe requirement of the former alternative. But the latter alternative is itself subject to a more or less demanding construal.

The less demanding reading would only require that the additional attitude be realized by the same person at the same time as the cognitive attitude. Mere accompaniment would be sufficient. The more demanding interpretation would require a stronger connection like entailment. Thinking in terms of a favorable evaluation of the object of faith, the weaker construal would require only that the subject of the faith that be also favorably disposed to the actual occurence of what one has faith that, while on the more demanding interpretation some stronger connection is required, for example, entailment. In this connection it is interesting that when Audi rejects the view that propositional faith that p is a combination of "belief that p and a positive attitude toward p's being the case," he does so because of the absence of an entailment of the latter by the former. He writes:

> Adding such an attitude to belief is still not sufficient for propositional faith. We do not, for example, have faith that something will occur simply because we believe that it will and we have a positive attitude toward its occurrence. . . . This move shows that in addition to finding an appropriate belief component, the reductionist would have to show this belief to imply an appropriate attitude. (H62)

Audi is quite right in denying that merely believing that p and having a positive attitude toward its occurence is sufficient for having faith that p. But I dissent from his idea that the trouble with this is that the belief that p does not imply the positive evaluation of its occurrence. As I suggested above, that implication need not be required. What is missing here is rather some less than ideal support for the belief that p. More on that later in the essay.

Although Audi falls short in failing to tell us what it is that satisfies his requirements (or some suitable subset thereof) for NDF, not a few of those requirements are on target with respect to appropriate requirements for propositional faith. I have already signed on with respect to the requirement of a positive evaluation of the object of the faith. Audi does not, so far as I know, forthrightly assert that to be properly said to have faith that p, one must be taking p to enjoy less than ideal support, but his assertion that NDF requires less for justification or rationality than belief (3c) could be read as a backhanded acknowledgment of this requirement. So if he would just tell us what cognitive propositional attitude it is that satisfies all (or most) of his requirements for NDF, he would be well positioned to fill out the account of what makes it faith.

The other comment I will make in this section concerns how close to belief, even flat-out belief, he portrays NDF as being. There are the similarities he explicitly sets out (4a–4d). And in addition we find the following: NDF is subject to doubt, within limits (3a and 3b). It is subject to evaluation as either justified or rational, or both (3c). It involves a disposition to avow the proposition, make inferences from it, and act on its basis, even though these dispositions are said to be weaker than the corresponding dispositions for belief (3d). It can be strong and steadfast (2f). And so, as Audi presents NDF, it would seem to be much more similar to belief than he acknowledges. This is grist for Radcliffe's mill in his 1995, and he makes full use of it in arguing that the functions Audi ascribes to NDF can be better carried out by belief. I find many of these arguments convincing, but as I said above, I am carrying out a different critique in this essay. In accordance with that I take all the similarities

I have been presenting to be quite compatible with the nondoxastic character of NDF. But the account of just how they are thus compatible awaits the depiction of a propositional attitude that could be the basic cognitive factor of a nondoxastic propositional faith. It is this that I will undertake in the rest of this essay. I shall seek to explain just how a form of propositional faith can be very similar to belief but still different on a crucial point that makes it distinctively nondoxastic.

III

What positive propositional attitude is both distinct from belief and can, with suitable accompaniments, satisfy as many of Audi's specifications of NDF as can comfortably hang together and intuitively fit the notion of faith that p? My candidate for this office is what I shall call *acceptance*. The concept of acceptance in question was mostly inspired by Jonathan Cohen in his 1992. My version is different in some respects. My first task is to give an identifying characterization of this propositional attitude and explain how it is related to belief.

I find it most useful to approach this by being as explicit as possible as to how I am thinking of belief, not idiosyncratically but rather how I lay out what I take to be a familiar, ordinary concept of belief. I make no claim that this is the only such concept, but it is one that is dominant in philosophy, including discussions of religious belief and its connection with religious faith. Since this whole discussion is focused on Audi's attempt to identify an NDF, it would simplify things if he and I were not at odds on how to construe belief. And fortunately such is the case.

In his 1972 Audi sets out what he calls

> a theoretical construct analysis, since it construes the concept of believing as similar in some very important respects to the concepts philosophers of science call theoretical, e.g., the concepts of an ion, of a radioactive substance, and of a molecular lattice. The only important assumption I shall make about theoretical concepts is that, roughly speaking, they derive their 'meaning' from the main lawlike propositions in which they figure and are thus to be explicated by making clear their role, especially their explanatory and predictive role, in these propositions.(43)

I am in complete sympathy with this kind of approach. Here is the way Audi applies it to the case of belief. He lists a number of propositions that he takes to provide the core of a philosophical account of the meaning of 'x believes that p' (xBp), where 'x' ranges over cognitive subjects and 'p' over propositions.

1. If, under favorable conditions, x presents arguments, or appears to make inferences, in which p is an obvious step, or of which p is the conclusion, but does not mention p, then xBp.

2. If, under favorable conditions, x says things that obviously presuppose or imply that p, or the saying of which obviously presupposes or implies that p, then xBp.

3. If x is in a situation in which a normal person would perceive (or would think he perceived) that p, then, under favorable conditions, xBp.

4. If xBp, then if x discovers or suddenly comes to believe that p is false, or if x discovers or suddenly comes to believe either (a) that someone whose judgment

concerning p (or related subjects) he respects, has thoughtfully denied p, or (b) that there is (what x takes to be) substantial evidence against p, x tends to be upset or worried or surprised, or led to reconsider p or related propositions, or some combination of these conditions tends to occur.

5. If xBp, then x tends to believe (a) propositions he believes are conceptually necessary consequences of p, and (b) to a lesser extent, propositions that he believes follow from p with fairly high (greater than .50) probability.

6. xBp if and only if: under favorable conditions, x tends (a) to invoke p injustifying his beliefs, his reasoning, and his actions, and (b) to reason in accordance with p in his planning of strategies and activities.

7. xBp if and only if: under favorable conditions, x tends to say, assert, insist, affirm, avow, or the like, that p.

8. xBp if and only if: under favorable conditions, x tends to perform actions such that if an explanation of one or more of these actions were requested, it would be reasonable to invoke x's believing that p as an essential part of the explanation.

There are several things to be noted about these propositions: (1) Audi claims that whoever employs the concept will believe all of them, at least implicitly; and (2) They can claim to be lawlike only because of the qualifications "under favorable conditions" or "tends to." This is because what actions, or cognitive or affective outcomes, are forthcoming depends on various other psychological and other factors, as well as x's belief that p. As Audi points out, this assumes that it is possible to specify, at least to a considerable extent, what other factors would be "unfavorable" or would oppose the tendency in question.

My use of this approach differs somewhat from Audi's. First, although I agree that all of the items on the above list are true and lawlike, with proper qualifications, I find that on the whole Audi is more "external" and indirect in his choice of items than he needs to be. Thus 1–3 all have to do with situations in which p does not enter into what x is doing, thinking, or feeling, but in which a third party would take the situation as reason to attribute the belief that p to x. I have no objection to the truth or lawlike status of these, but I prefer to highlight relationships that seem to me more directly connected with the belief that p, as will appear below.

Second, I feel no need to stipulate that my lawlike connections must be believed, even implicitly, by x in order for them to function to pin down the concept of belief. It is sufficient that they be true and lawlike. My reason for shying away from the stronger claim is that it seems to me to require too much for possession and use of the concept. A user of the language could have acquired the ability to apply and withhold the concept in a way that tallies with these lawlike connections without being credited with *beliefs* that they hold. Correct and fluent language usage need not be the object of higher-level recognition of the principles governing it to that extent.

For that matter, if the concept of belief is like many other familiar concepts, it will be impossible to make the truth of any set of propositions setting out conditions that are either necessary, sufficient, or both for its application constitutive of the concept. The connection may be looser than that. It may be that the conditions in question are best thought of as "belief-making conditions," so understood that so

long as a sufficient number of them are satisfied in a given case the concept applies, but without our being able to say exactly how many is sufficient, much less draw a sharp line between those each of which must be satisfied for the concept to apply and those that are not individually necessary for this. So-called "family resemblance" concepts are of this sort. But since my main concern in this essay is elsewhere, I will not go into detail about ways in which concepts can fail to conform to a straightforward necessary and sufficient condition model, and I will assume such a straightforward model for this discussion.

With these caveats in mind, here is my favored list of lawlike conditionals, the truth of which is sufficient to identify the concept of belief in which we are interested ('xBp').

A. If xBp, then if someone asks x whether p, x will have a tendency to respond in the affirmative.

B. If xBp, then if x considers whether it is the case that p, x will tend to feel it to be the case that p, with one or another degree of confidence.

C. If xBp, then x will tend to believe propositions that she takes to follow from p.

D. If xBp, then x will tend to use p as a premise in theoretical and practical reasoning where this is appropriate.

E. If xBp, then if x learns that not-p, x will tend to be surprised.

F. If xBp, then x will tend to act in ways that would be appropriate if it were the case that p, given x's goals, aversions, and other beliefs.

This list is very similar to Audi's. My E is substantially identical to a part of Audi's 4. The same kind of approximate identity holds between my C and Audi's 5, my A and Audi's 7, my D and Audi's 6 and, to a lesser extent, my F and Audi's 8. With this degree of overlap it is not surprising that the comments I made concerning Audi's list about the lawlike propositions being of a tendency or "under normal conditions" form apply equally here, as do the hints that precise necessary and sufficient conditions, and a sharp line between those constitutive of meaning and those that are not may not be possible. Aside from different preferences as to the best way to formulate the basic point in each case, the main difference between these five propositions in my list and their correlates in Audi's has to do with the greater simplicity and directness of mine. The three items on Audi's list that have no correlates on mine—1, 2, and 3—are the ones I singled out earlier as being unnnecessarily indirect in their connection with believing that p. As I pointed out then, I accept them as true of believing, but I don't feel that they add much to a list of lawlike conditionals that are central to the nature of believing.

The reader will note that there is a still unmentioned item on my list, B, that has no correlate on Audi's and that gives rise to an important difference between the two accounts. I have no reason to suppose that Audi would be inclined to deny my B. Perhaps his reason for not including anything like it is that he was concerned at that early stage of his career to stress the behavioral and other external manifestations of belief. Or perhaps it was something else. In any event, I take it to be at the heart of belief and, as we shall see, it makes one of the crucial differences between belief and acceptance. Cohen in his 1992 gives something like it pride of place in his account of belief. His definition of 'xBp' is in terms of it alone, leaving all other characteristic dispositions in another status. His formulation runs as follows: "the

disposition, when one is attending to issues raised, or items referred to, by the proposition that p, normally to feel it true that p and false that not-p." Note the same tendency qualification with "normally." My main reservation about this formulation, apart from the claim that it is not enough by itself to tell us what belief is, is that it restricts believing to subjects that have the concept of truth and falsity. Even if we restrict ourselves to human subjects and ignore lower animals, it seems clear to me that humans who are at too early a stage of cognitive development to be in possession of the concepts of propositional truth and falsity have beliefs. Hence my formulation in terms of "feeling it to be the case." Note too that my formulation allows for degrees of belief, though I am not completely happy about thinking of belief in this way.

I follow Cohen in using the term 'feel' in order to stress the *immediacy* with which one realizes a conscious belief, that it is something one *experiences*, that it is a matter of being *struck* by how something or other is. It is something one *finds* oneself with, rather than something one has thought out. This is one of the ways in which acceptance is different from belief, despite the fact that analogues of all the other conditionals on my list are true of it as well. This is how acceptance can be very like belief though completely distinct from it. The other crucial difference, which is closely connected with this one, is that an acceptance that p is, at least in its first appearance, a deliberate, voluntary act, rather than a state in which one finds oneself.[3] But belief is not something one can voluntarily come to possess. I cannot by an act of will bring it about that I believe that World War II is still going on or that the Holocaust never happened, or anything else that I believe to be obviously false. But this doesn't clearly distinguish belief from acceptance, because anything I take to be obviously false I am unable to accept as well. To be able to (really) accept p I must find p to be somewhat credible, enjoying some support though not enough to make it something I find myself believing. And indeed the involuntary character of belief stretches over what we have some reasons for and what we have some reasons against as well as the obviously false. If the reasons in question are insufficient to automatically bring me to find myself with the relevant dispositions, I cannot produce them by an act of will. This will be illustrated shortly.

I find the voluntary character of the act of acceptance to be the best way of giving an initial idea of it. The act of acceptance, unlike a state of belief, is the adoption, the *taking on* of a positive attitude toward a proposition. More specifically it is a mental act, as distinguished from the linguistic act of assenting to p, which can be an overt expression of acceptance, though 'assent' is sometimes used for what I am calling 'acceptance'. But when we come to say just what positive attitude to a proposition is adopted when one accepts it, we are back to the pervasive similarity of acceptance and belief. The other five conditionals in my account of belief have close analogues for acceptance. Analogues of the dispositions that one simply finds oneself with when one believes that p are voluntarily taken on when one accepts that p.

Thus accepting that p is both a complex dispositional state markedly similar to believing that p, but distinguished from it by the fact that it issues from acceptance in the other sense, a mental act that is voluntarily engaged in. It is, no doubt, a strain to use 'acceptance', which by rights is a term for an act that is performed at a

certain time, for a more or less long-lasting dispositional state. But since the act and the resultant dispositional state are so intimately connected, I will continue to follow Cohen in using 'acceptance' for both.

Having voluntarily accepted p and maintained this situation for a longer or shorter period of time, one may eventually find oneself believing that p. The state of acceptance can develop into belief. But is it possible to simultaneously believe and accept that p? I think that this is not a live possibility. If one believes that p, one is already in the relevant dispositional state, and it would be redundant to adopt it as well. One adopts the state only if one doesn't already simply find oneself with it.

I will seek to make the notion of propositional acceptance more concrete by giving a couple of examples, leaving the connection with faith to the next section. Consider an army general whose forces are facing enemy forces with a battle imminent. He needs to proceed on some assumption as to the disposition of those forces. His scouts give some information about this but not nearly enough to make any such assumption obviously true, or even overwhelmingly probably true. So what does he do? He *accepts* the hypothesis that seems to him the most likely of the alternatives, though he realizes that he is far from knowing that this or any other such hypothesis is true. He uses this as a basis for disposing his forces in the way that seems most likely to be effective, even though he is far from believing that this is the case.

There are several points to be made about a case like this in which the impetus for an acceptance is a practical one, the need for guidance of action. First, we have what William James called a forced option. The general must proceed on the basis of some hypothesis about the disposition of enemy forces; remaining in suspension about it is not an option. Second, accepting one of the alternative hypotheses, in the sense of 'accept' with which I am concerned, is different from adopting a "working assumption" in the sense of proceeding as if it were true, in contrast to *taking* it to be true. Acceptance in my sense is the latter. Though the general neither knows nor believes that the hypothesis is true, he *takes* it to be true, he *commits* himself to its truth. He reasons from it, acts on the basis of it, and will be disposed to defend it if his subordinates question it. It may well be difficult in this sort of case to determine whether he accepts the hypothesis in this sense or simply chooses to act as if it were true. But acceptance that p in the sense under discussion is a voluntary taking it to be the case that p.

Now for a theoretical case in which there is no such practical pressure to make a choice between alternatives. I investigate different positions on the free will issue. On the basis of careful consideration I conclude that they support most strongly the libertarian position. Do I flat-out believe that this is the correct position? Do I find myself spontaneously feeling confident of its truth? No. Having performed an initial act of accepting it as true, I may require supplementary injections of such an act to keep myself in the relevant dispositional state vis-à-vis this position. But so long as I am in the dispositional state I adopted, I announce libertarianism as my position, I defend it against objections, and I draw various consequences from it.

I have been stimulated by some comments by, and conversation with, Dana Radcliffe to emphasize the point that cases in which there are considerations that support each of several incompatible alternatives might, depending on the relative

perceived strength of these considerations, as well as on the motives and tendencies of the subject, give rise to belief, acceptance, or adopting a working assumption. These and others are all possible outcomes of such a situation, though depending on which other factors are present, the live possibilities might be reduced, even reduced to one. I make this point in order to avoid the impression that I think that all cases meeting the conditions specified at the beginning of this paragraph will lead to acceptance or to any other particular outcome.

IV

The time has come to consider whether acceptance can be the cognitive core of *faith that*, both religious and otherwise, and more specifically to determine whether it can play this role for something like Audi's NDF. Before getting to Audi I will consider what in my view must be added to acceptance to get propositional faith.

1. The accepter of p must consider the truth or realization of p to be a good thing. It would be sufficient for the general's acceptance of the hypothesis that the enemy forces are disposed in a certain way not to be a case of propositional faith that the general does not consider such a disposition of forces not to be a good thing. If he doesn't positively evaluate it as a good thing, it would not be correct to credit him with *faith* that the dispositions are that way even if he should accept it. Again, I might (or might not) positively evaluate people's having libertarian free will. And if I do not, it would not be correct to say that I have *faith* that humans have libertarian free will when I accept the libertarian position.

2. To count as faith that p, S's acceptance that p has to be less than ideally or fully supported by reason, evidence, or experience, at least in S's view of the matter. One might think that this requirement is automatically satisfied by every case of acceptance just by virtue of its not being belief that p. For if it were ideally satisfied, S's attitude to p would be one of belief rather than acceptance. But this is too quick. It all depends on how far below the maximum the support for p must be in order that the acceptance count as faith. It may be that there is a gray area in which the support for the proposition is less than what would automatically result in belief but more than the maximum that would be consistent with a propositional faith status. It is unrealistic to suppose that there are precise cutoff points in this matter. But at least we can say that there is rough degree of support for p such that more support would prevent our calling the acceptance of p a case of faith that p.

The paradigmatic cases of propositional faith satisfy both of these conditions. This fits nicely with the classic formulation in the Epistle to the Hebrews 11:1: "Faith is the assurance of things hoped for, the conviction of things not seen." We have the pro-attitude toward the realization of the object in the first phrase and the weak epistemic situation in the second. But we can still ask what happens when only one of the conditions is satisfied. Suppose I very much want you to get the job but also have strong reasons (strong enough to produce belief normally) that you will. And contrast this with the case in which I don't want you to get the job (am either opposed to this or indifferent) but lack strong reasons either that you will or that you will not get the job. It seems much more plausible to credit me with faith that you will get

the job in the first case than in the second. Indeed, that description seems clearly out of order in the second case.

Now let's turn to Audi's characterizations of NDF that were presented at the beginning of section II. In section II I cut down this list somewhat. For reasons given there, either 2i, or 1e and 1g must be eliminated, since they are incompatible. The list can be further trimmed down by eliminating redundancies. Thus 1b, 1d, and 3f follow from the denial that NDF is flat-out belief. And 2b follows from its being a positive attitude to a proposition. There are others that are either mistaken about faith or at least not clearly true of it. These include 1c, 2g, and 3d. As for 2j, I take it that this applies to *faith in* rather than *faith that*. With what does that leave us?

First and most important, NDF is not flat-out belief (1a). Since Audi makes explicit that it is not tentative or weak belief and is not belief that it is probable that p (1h, 1i, and 1j), it would be in the spirit of his characterization to drop the 'flat-out' qualification in 1a and simply deny that it is belief of any sort. The other anchor to the cognitive part of NDF is that it is a positive attitude to a proposition (2i). We can see the nondoxastic character of the positive propositional attitude that constitutes NDF reflected in two other specifications: that it requires less for justification or rationality than belief (3c) and that if p turns out to be false there is less tendency to be surprised with NDF than with belief (3e). But though NDF is not any kind of belief and contrasts with belief in the ways just mentioned, it also exhibits important similarities to belief in addition to being a positive propositional attitude. Audi explicitly mentions the ones I grouped under 4. In addition to those, it exhibits the dispositions mentioned in 3d, it is subject to evaluation as justified or rational (3c), and it is subject to doubt (3a and 3b). So far this gives us a cognitive core that consists of a positive propositional attitude that is other than belief but similar to it in important ways. But this still isn't sufficient for propositional faith. In addition we need the condition that it is accompanied by a pro-attitude to the proposition's being true (2d). I suggested above that there is also the condition that the support for p is relatively weak. The closest Audi comes to that on his list is 3c, that NDF requires less for justification and rationality than belief. That doesn't quite do it, but I will assume that Audi would not object to adding this condition. I will ignore 2c for the moment because I am concerned with propositional faith in general rather than the specifically religious form, and I will continue to ignore it because I don't see that 2c makes any distinctive contribution to getting clear about what religous NDF is. The only other items of Audi's I haven't mentioned are that it can be strong and steadfast (2f), that it requires certain beliefs other than the belief that p (2h), and that it involves a disposition to believe that p (2e).

Now my claim is that acceptance that p is ideally suited to satisfy Audi's specifications that I have not eliminated and that have to do with the cognitive core of NDF, and that when we add the condition of a pro-attitude toward p's being the case and a relatively weak support for p, it satisfies all the specifications in my slimmed-down list and so provides us with an identification of what NDF is.

To wit, (1) acceptance is distinct from belief of any kind, though very similar to it in the ways Audi mentions, as well as involving analogues of the dispositions I claimed in section III to be characteristic of belief except for the one that distinguishes belief from acceptance. In addition, acceptance is subject to doubt though

within limits, and being less of an unqualified commitment to the truth of p than belief it requires less for justification or rationality. As for 2e, involving a disposition to believe that p, acceptance satisfies something of the sort in that, as we have seen, it can develop into a belief that p. As for the possibility of being strong and steadfast, I see no reason that an acceptance can't enjoy these features to any given degree. And as for requiring certain beliefs other than p in order that a positive attitude to p be possible, it seems that this holds of acceptance for just the same reasons that it holds for belief. For example, in both cases one can so much as entertain the proposition that p only if one has certain beliefs.

Hence I submit that acceptance of p as I have depicted it, when supplemented by a pro-attitude toward the truth of p and relatively weak epistemic support, is ideally suited to fill the gap in Audi's characterization of NDF by giving a satisfactory identification of it. To be sure, Audi might complain that I have arrived at this result by deleting some of his characterizations of NDF. Indeed, one such characterization (1e) seems to explicitly deny that it is an *acceptance* of p, and acceptance obviously doesn't satisfy that![4] But remember that, with the possible exception of 1e, all of the deletions were for reasons other than what is satisfied by acceptance. And if those reasons are sound, then the list has to be trimmed down in the ways I did, whatever the candidate for playing the role of NDF. Moreover the absence of any uniquely identifying indication of what NDF is in Audi's account is a very serious lacuna, and if it can be filled out only by trimming down his set of characterizations as I did, then that change is well worth it.

V

The last topic on my agenda is the role of nondoxastic propositional faith, in my acceptance + form, in religion and, more specifically, in Christianity. In the rest of this essay I will be supporting Audi rather than criticizing him. He argues at some length that NDF can and does serve adequately as the positive cognitive attitude toward religious propositions that is crucial for a religious form of life. In this section I will be supporting a parallel position for my acceptance + form of nondoxastic propositional faith. In doing so I will be presenting in brief form some points made much more fully in Alston 1996, to which the reader is referred. This section will also be contrary to Radcliffe 1995 in which the author argues that nothing other than belief can successfully play the role belief does in the religious life.

Before making a case for the religious importance of acceptance, let me be clear that this by no means implies a derogation of religious *belief*. On the contrary, there are many respects in which a firm belief in the major doctrines of the Christian faith is a superior position to a nonbelieving acceptance of them. For example, the accepter will receive less comfort from her faith than the unquestioning believer, and she will be more troubled by doubts and waverings. But I will be contending that the accepter is not necessarily inferior to the believer in commitment to the Christian life or in the seriousness or intensity with which she pursues it.

I am inclined to think that a sizable proportion of contemporary, sincere, devout Christians are accepters rather than believers. I have no statistical evidence

for this, but in Alston 1996 I cite clear cases of both groups from two collections of faith stories of philosophers—Clark 1993 and Morris 1994. Here are two samples from the latter volume. First from the belief side: "I was sure that I had met God and that He had granted me salvation" (103). . . . "I had no doubt that it was God encountering me, speaking to me, forgiving me, and so on" (106). And from the acceptance side: "I long thought that even if Christian theism isn't more probable than not, it is still reasonable to *embrace* it" (80). "My attitude is in many ways similar to T. S. Eliot's . . . When asked why he *accepted* Christianity, he said he did so because it was the least false of the options open to him" (85, my emphases).

If acceptance as well as belief is common among committed Christians, how is it that I can present my claim as a startling discovery. I think it is because 'believe' has been allowed to spread over all positive propositional attitudes. This is reflected in dictionaries, even in one so sophisticated as the Oxford English Dictionary. There we find among the entries for 'believe'—"To give credence to, to accept (a statement) as true"; "mental acceptance of a proposition . . . assent of the mind to a statement." There are many indications in the philosophical literature of this inflated use of 'believe'. Audi cites some contemporary philosophers as defining religious faith as consisting in belief. Again, in discussions of whether belief can be voluntary, philosophers are cited as proponents of a positive answer where on careful reading it is clear that it is, rather, acceptance of which they are speaking. In *Summa Theologiae* II, II, 2, 9, Aquinas characterizes faith as "an act of the intellect assenting to divine truth at the command of the will moved by the grace of God." This is standardly taken as a view that belief is subject to voluntary control; but 'assent', being an *act* as Aquinas makes explicit, is a term on the acceptance side of my distinction.

Suppose I am right in my view that acceptance of some or all fundamental Christian doctrines, rather than belief in them, is common among committed Christians, and in my view that they may be as serious and committed to the Christian life as those who believe those doctrines. It could still be that they fall short of a full participation in the Christian form of life. This would be the case if Christianity requires its adherents to believe the articles of faith and not just to accept them. (After all 'believer' is a common term for such adherents.) To consider this issue we can examine some basic formulations of the articles of faith to see whether propositional belief of them is required of professed Christians. We will see that the evidence for such a requirement is far from overwhelming.

First look at a couple of New Testament passages that are often taken as requiring belief: "For whoever would draw near to God must believe that he exists and that he rewards those who seek him" (Hebrews 11:6); "He who does not believe will be condemned" (Mark 16:16). But the Greek verb translated 'believe' in these and many other passages is 'pisteuo', the verb cognate of the noun 'pistis'—faith. In English we lack a verb cognate of 'faith', and this leads translators to settle on 'believe'. But once we realize that it is not always belief that is the cognitive core of propositional faith, we can see that a better translation would be "have faith that he exists" for the first passage and "he who does not have faith" for the second. And we have seen that propositional faith can be realized by either acceptance or belief, in which case the requirement can be satisfied by either.[5]

The extrabiblical formulations most widely accepted by Christians are the Nicene and the Apostle's creeds, in which 'believe' occurs only in the 'believe in' form. In the latter we have "I believe *in* God, the Father Almighty . . . *in* Jesus Christ his only Son our Lord . . . *in* the Holy Spirit, the holy catholic Church." It is the same story with the Nicene Creed. Of course, believing in various items presupposes propositions about those items, and hence some positive attitude toward those propositions on the part of the 'believer'. But that could be satisfied either by believing those propositions or accepting them.

For another example consider the Baptismal Service in the Episcopal Book of Common Prayer, which in its essentials is rooted in early Christian tradition. No propositional beliefs are required of candidates, or of the sponsors of infants. They are called on to *renounce* "Satan and all the spiritual forces of wickedness that rebel against God," to affirm that they *accept* Jesus Christ as their savior, to *promise* to follow Jesus Christ and obey him as their Lord, and to *affirm* the Apostle's Creed. But nowhere are they called on to make a statement of the form "I believe that ___." Lest you dismiss this as what one would expect from a bunch of liberals like the Episcopalians, consider the major Protestant confessions from the period of the Reformation, such as the Augsburg Confession (1530) or the Westminster Confession (1647). These consist of doctrinal statements without any explicit injunction to believe them. Of course the implication is that a member of the communion in question is to accept them (in a wide sense of 'accept'), but that could consist in either accepting them in my sense or in believing them.

This, in brief, is my case for the sufficiency for the cognitive aspect of the Christian faith of acceptance of the relevant doctrines. Though it does not give the adherent everything that firm belief does, it provides an ample basis for a full-blooded participation in the Christian form of life. The case is set out more fully in Alston 1996.

To sum up the burden of this paper, I find that Audi's characterizations of NDF, though they contain many appropriate specifications, are deficient in failing to identify sufficiently just what NDF is. I have presented a candidate for such an identification—a positive attitude to a proposition p I call 'acceptance', which, when accompanied by a pro-attitude toward the realization of p, is the best candidate I can find for Audi's NDF. Though it is not satisfied by all of Audi's specifications, some of which are disqualified by independent considerations, it does, I believe, carry out the spirit of Audi's project in the most satisfactory way. In addition, I have sought to show that my version of NDF can play the role Audi assigns NDF of providing an alternative to belief as the cognitive core of religious faith, and more specifically of Christian faith.

Notes

1. See Audi 1991 (hereinafter "T"), 1992 (H), 1993 (S), and 2003 (B). If you wonder why these designations were chosen, they represent the editor in each case, James Tomberlin (*Philosophical Perspectives*), Marcus Hester, Eleonore Stump, and Rainer Berndt.

2. See Radcliffe 1995.

3. In calling it a voluntary act I do not mean to imply that it is a exercise of free will in a libertarian sense of the term, or that it is not. Either way, it is voluntary in that it is an exercise of will, or, if you prefer, of choice.

4. This, of course, depends on what concept of acceptance Audi is employing here, and it is not clear that he is employing the concept I have presented.

5. Consider the following statistics from Smith 1977, chap. 3. Words in the 'pistis' family occur 603 times in the New Testament. In only 4 percent of these occurrences is the word followed by a proposition. The noun 'pistis' occurs 246 times, and in 217 of these there is no object at all.

Bibliography

Alston, William P. 1996. Belief, Acceptance, and Religious Faith. In *Faith, Freedom, and Rationality*, ed. Jeff Jordan and Daniel Howard-Snyder. Lanham, Md.: Rowman & Littlefield.

Audi, Robert. 1972. The Concept of Believing. *The Personalist* 53: 1.

——. 1991. Faith, Belief, and Rationality. *Philosophical Perspectives* 5.

——. 1992. Rationality and Religious Commitment. In *Faith, Reason, and Skepticism*, ed. Marcus Hester. Philadelphia: Temple University Press.

——. 1993. The Dimensions of Faith and the Demands of Reason. In *Reasoned Faith*, ed. Eleonore Stump. Ithaca, N.Y.: Cornell University Press.

——. 2003. The Theology of Faith and the Ethics of Love: A Philosophical Perspective. In *Religion in der Moderne*, ed. Rainer Berndt. Wurzburg, Germany: Echter.

Clark, Kelly James, ed. 1993. *Philosophers Who Believe*. Downers Grove, Ill.: Intervarsity.

Cohen, L. Jonathan. 1992. *An Essay on Belief and Acceptance*. Oxford: Clarendon.

Morris, Thomas V., ed. 1994. *God and the Philosophers*. New York: Oxford University Press.

Radcliffe, Dana M. 1995. Nondoxastic Faith: Audi on Religious Commitment. *International Journal for Philosophy of Religion* 37, no. 2.

Smith, Wilfred Cantwell. 1977. *Belief and History*. Charlottesville: University Press of Virginia.

This page intentionally left blank

INTENTION, SELF-DECEPTION, AND REASONS FOR ACTION

This page intentionally left blank

Trying with the Hope

FREDERICK ADAMS

> Now if Sally knows that she is so bad a marksman that her chances of
> hitting the fox are very slim, she might hope to hit it, but not intend to;
> and she would probably answer with something like "I'm going to try."
>
> —Audi 1993, 71

I. Introduction

Al is standing at the half-court line of the basketball court. Al is a professional philosopher, not a professional basketball player. He decides to shoot for the basket. Can Al intend to make a basket? He does not have high hopes of its going in. He does not think it is impossible that he make the shot. Can he intend that the ball go in? There is surprising disagreement among professional philosophers about the answer to this question. (I'd be surprised if there were much discussion about it by professional basketball players, though they do ask "Did he call it?" when a ball goes in off the glass.) In this paper, I will consider the relevant views and do my best to arrive at an answer. I'd like to say that Al can indeed intend to make the shot, even if he does not think his chances are very good, as long as he does not think it is impossible. This paper is about whether that view is defensible.[1]

Robert Audi's view, represented in the opening quote, has inspired many.[2] His view represents the most plausible alternative to the one I will propose here, and it is likely the most intuitive view that we will consider. To many, it just sounds wrong to announce that you intend to do something like make a basket from mid-court, if you are fairly confident that your chances are slim. It is much easier on the ear to say that you will try or that you intend to try. In what follows, I will consider Audi's view and his defense of it (along with other views—many of which his view has inspired).

II. A Plethora of Views

In this section, I consider Audi's view, along with some of the views that came before and inspired his view, as well as some of those that came after and were inspired by

his view. Since I am looking at the views on a continuum of belief strength, I will take up Audi's view in order of belief strength he requires of intention. However, as things unfold, we shall see that the main differences between a view like Audi's and the one that I wish to defend will turn on two considerations: pragmatic features of language and avowal of intentions, and the nature of trying. I will maintain that Audi's view is overly influenced by the pragmatics, and that attention to the cognitive nature of trying itself lends support to a weaker belief constraint on what one can intend to do.

Among philosophers there is a diversity of opinion about the beliefs necessary to intend to do an act A. The range runs from belief certainty that one will do A (unless prevented) to a weak negative belief that it is not impossible that one not do A—with a wide variety of moderation of strength of belief in between. So let's say the strongest view is that to intend to make the basket from half-court, Al must believe that it is certain that he will, unless he is prevented from doing do.

(1) INTENTION TO A → BELIEF THAT *IT IS CERTAIN THAT ONE WILL A (UNLESS PREVENTED)*.

This is a view briefly suggested at one time by Hampshire and Hart (1958), though they later express a weaker view.

(2) INTENTION TO A → BELIEF THAT *ONE WILL A*.

This view is held by Paul Grice (1971), Monroe Beardsley (1978), Wayne Davis (1984), Gilbert Harman (1986), and, perhaps, Michael Bratman (1990). Beardsley (anticipating the work of Bratman on the role of intention in planning) says that intending to A requires more than a belief that one will try to A because plans made by oneself and others may depend on one's A-ing. For example, making plans to meet someone at the airport may require more than the belief that one will try to make it to the airport. One must be fairly confident of getting there to intertwine the lives or livelihoods of several people at once.

Still if act A is a solitary act, it may not require the belief that one will A. Even here, Beardsley seems to require the stronger belief that one will A. He says of one said to be searching for the fountain of youth, without believing he will find it: "he may search for it, but it does not follow that he intends to find it. So, here is no case in which we have intending without believing" (1978, 178). And of one holding to a ledge as long as he can, Beardsley says: "though the man may intend to stay as long as he can, he surely believes that he will do this" (1978, 179)—describing the desperate state of a man intending to hold on for a few seconds more (and believing that he will, at least for a few seconds more).

Michael Bratman (1990) comes close to agreeing with this approach insofar as it relates to rationality constraints on planning. Bratman hedges that "normally" when one intends to A, one also believes that one will A (1990, 22), or at least has "support for the belief that one will A" (1987, 38). I take this to be the view that, when it comes to intending and planning to A, the default position is that one believes that one will A, but that default is defeasible, and Bratman lists cases where there are exceptions. He even explicitly claims to have "rejected . . . Grice's positive account"(1990, 131).

He acknowledges that one knows there are cases where one "may well forget" to act on one's plans (1987, 37), or where one may well be "agnostic" about success (1987, 38). But at the same time he notes that "examples such as these do not *prove* that an intention to A does not require a belief one will A" (1987, 38).

Wayne Davis (1984) seems to agree with the approach, saying that S intends to A only if S believes that he will A (thinks he will A or expects to A). I say "seems" because Davis places a low threshold on belief (being more certain of p than of not p). It sounds as though one may harbor doubts about doing A, on this view. In addition, Davis requires of the intention to A only that one be "inclined" to believe one will A. It is not clear whether this is a weakening of belief, or a disposition to hold a stronger belief. Davis may also detect something pragmatic going on. He appeals to the belief requirement on intention to explain why " 'I will call' is an *utterance implication* of 'I intend to call',", and he notes that " 'I intend to call but I won't', while not logically inconsistent . . . is most odd" (133). I agree that it is "most odd," but will revisit this issue later on. Still, Davis makes crystal clear that he rejects the weak view (#7 below), saying "unlike hoping, intending that *p* requires more than not being certain that not *p*" (133). In comparing the weaker view to hope, not intent, he falls in with a long line of others, as we shall see. Davis, and others, defends this by pointing out that when one knows or believes one's chances of A-ing are slim (as in Al's mid-court shot), to say "I intend to do A" would be an "exaggeration" or a "delusion." I should also point out that Davis is explicit that the conditions on *trying to A* are weaker than on *intending*. Davis maintains that we can try to do what we do not expect to do and therefore that "trying something does not entail intending it" (134).

(3) INTENTION TO A → KNOWING THAT *ONE WILL DO*,
     OR TRY TO DO, A.

This view is held by Hampshire (1960, 102), although Hampshire expresses other views as well. It is rejected by Audi (1993) on the grounds that if one intends to raise his hand and believes he has normal bodily control, one would not (merely) try to raise his hand. Hence, Audi does not accept the view that trying can be effortless and involved in things one knows one can do. Were one to accept that trying can be effortless (Armstrong 1980; Adams and Mele 1992), then Audi's objection would fall short.

(4) INTENTION TO A → BELIEF THAT *ONE'S ACTIVITIES*
     WILL BE A PROBABLE WAY TO ACHIEVE A.

As we saw in the opening quote, this is the view held by Robert Audi (1993, 1997). Audi maintains that the belief that what one is doing has a likelihood of bringing about A or is a way of doing A is too strong. For one may intend to help a stranded motorist but not definitely believe one's actions will result in help. Yet if one's beliefs about success were too low, Audi maintains that in acting one "could be properly said only to hope to help him by doing so. To distinguish intending to bring about Φ by A-ing from merely hoping to bring about Φ by A-ing, we need to require that S at least believe her A-ing will be a probable way to achieve Φ" (57). Audi goes on to clarify that by "probable" he means "more likely than not" (1993, 57).

Audi points out that we typically expect to do what we intend, but that this is defeasible. He also says that one can both intend to do A and hope to do A, but when one's belief in the probability of doing A slips below half, then a person only hopes to A and does not intend to A (1993, 58).

While I have placed Wayne Davis with the stronger view #2, I should point out that he seems to embrace view 4 at one point as well. He says: "Intending that *p* only requires being more certain that *p* than not-*p*. As Audi observed (1993, 58), I might intend to speak to someone in Boston, even though I think the chances of him being there are barely better than even" (1984, 134). I pointed out above that Davis's weak conditions on belief itself make this possible, and therefore make it difficult to distinguish Davis's view as that of #2 rather than #4.

(5) INTENTION TO A → NOT BELIEVING *ONE WILL NOT DO A*.

This view is held by Annette Baier (1970, 657) and Berent Enc (2003, chap. 6). It is rejected by Audi with an example of a hunger striker who intends to accept food tomorrow because he believes he cannot possibly maintain his hunger strike beyond tomorrow. Hence, he intends what he believes he cannot avoid doing (Audi 1993, 69). Enc defends this view as "a logical consequence of the rationalizing role of intention attributions" (191).

(6) INTENTION TO A → NOT BELIEVING (AT THE PERTINENT
      TIME) *THAT ONE (PROBABLY) WILL NOT A*.

This view is held by Al Mele (1992, 130). Mele accepts this negative belief requirement, rather than the positive one in Audi's #4, because he thinks there are actions that clearly are intentional that #4 rules out but that #6 does not. He points out that a basketball player who shoots free throws at a rate of less than 50 percent may shoot and make a free throw. Mele claims most would say he made the shot intentionally (and that he intended to make it), even though view 4 would say he only tried with the hope of making it (if he knew his true percentage for the season). On Mele's view, the player not only intends to make it but also makes it intentionally, if it goes in.

Mele defends this view against examples that suggest the weaker view #7. In an example of Myles Brand's (1984) that is very similar to Audi's example of the hunger striker, we are to imagine one who intends to take up jogging for the future. However, the jogger also knows about himself that his resolve will fade over time and that it is unlikely that he will be jogging by the same time next week. Nonetheless, Brand maintains that, even knowing this, the person intends now to jog every day next week. In another case, a trapped man's only escape is to jump a chasm. The man may have serious doubts about his ability to make the jump successfully, but with no other way out, the man intends to jump the chasm.

About such cases, Mele maintains that at best these actors *intend to try* to do the relevant acts (jog next week, jump the chasm). So in this he agrees with Audi and Davis, that when one's confidence slips too low, an intention to A must be replaced with a hope to do A or an intention to try.

(7) INTENTION TO A → NOT BELIEVING THAT
        IT IS IMPOSSIBLE TO DO A.

Davidson (1980), Adams (1986, 1997), and McCann (1986, 1989, 1998) hold this view. Brand (1984) gives several examples that are consistent with this view, though he does not formally endorse it. Anscombe (1957, 93) gives an example (similar to Audi's) that is consistent with this view. She says: "A man could be as certain as possible that he will break down under torture, and yet determined not to break down." Davidson says: "I do not believe anything will come up to make my eating [some candy$_{fa}$] undesirable or impossible. That belief is not part of what I intend, but an assumption without which I would not have the intention" (1980, 100).

(8) INTENTION TO A $\nrightarrow$ NOT BELIEVING THAT *IT IS*
        IMPOSSIBLE TO A.

Thalberg (1972) wins the prize for the weakest belief requirement on intention, namely, none. He seems to maintain that intending to do A does not imply that one lacks the belief that doing A is impossible. On his view, one can even intend to do what one believes to be impossible (to do what one knows to be hopeless). Ludwig (1992) thinks one can *try* to do the known impossible. I will discuss his view later when I talk more specifically about trying.

   Now I think that, in the end, Thalberg's view may collapse into view 7. I say that because Thalberg first restricts his cases to ones not known to be logically impossible (106). This makes me suspicious, to say the least, that he really holds view 7, and, second, he claims that beliefs about the impossibility of doing A are almost impossible to determine and that they may not be that strong, in the end. Although Thalberg says the type of agent he is thinking of will "possess what he considers sufficient grounds for believing he cannot execute his design" (106), he hedges by adding "our criteria for saying 'He thinks it's impossible' are impossibly vague" (107). And the examples that he gives tend to trade on cases where, to the agent, his action's success "remains conceivable," and that while the evidence that his success will be impossible is overwhelming, "future experience might disprove it" (108). For instance, Thalberg holds "if there is only one chance in ten that a platoon will overwhelm an enemy pillbox, most commanders would say, 'It's not humanly possible' " (108). I will say more later about trying and intending, but for now I will add that Thalberg sees no comfort in changing the cognitive attitude to *trying* rather than *intending* because he thinks trying to A implies intending to A (113). So he does not accept the idea that we might replace *intending* with *trying with the hope* when belief about success wanes.

III. Why So Many Views?

The above list is not meant to be exhaustive, nor are the entries meant to be mutually exclusive (maybe not for Davis, for example). Yet seeing the list displayed before us, one cannot help wondering why there are so many differences of opinion. Here

are some possible explanations. One is that philosophers have focused on different features of intention (or even that there are multiple notions of intention in play). Two is that some of what is going on is due to pragmatics of intention talk, as much as semantics. Three is that philosophers seem to disagree about the nature of trying.

III.A: *Different Aspects of Intention*

Audi (1986), Harman (1986), and others provide much help in seeing the sources of the diversity of opinion above. Audi attributes it to different aspects of intention, and Harman to different senses of "intend."

Let's start with the aspects or notions of intention in play. Harman (1986) has noted the difference between the strong belief requirement and the weak by suggesting that there may actually be "two notions here" (363). Harman uses "intending" for a stronger notion, like view 2 and "willing" for a weaker notion that does not require the belief that one will do A. Harman uses "willing" to name something akin to a propositional attitude (like belief or intention) that can initiate action. Harman goes on to discuss the differences between these notions and the relationship between them. For our purposes, it is sufficient to note that he thinks they are distinct notions and that they have different belief requirements. If true, that could explain some of the variety in the list. Philosophers are tuning into different notions of intention (or willing) as they construct their view of its belief requirements.

Another possibility, also mentioned by Harman (and by us in the case of Davis), is that the standards for belief and intending may be different. Harman's example is of a man who intends to go to New York tomorrow. On the strong view, it may be thought that he does not really intend to go, if he believes that something may come up and he may not get there (violating condition 2). But he may *believe* that he is going to New York tomorrow, realizing that something may come up and he may not get there. That realization does not violate his belief that he is going to New York tomorrow, for he cannot both believe that he will go and not believe that he will go. So his realization may not violate either his belief that he will go or his intention to go to New York tomorrow (view 2). Variation in belief strength on intending and belief itself may be accounting for some of the variety among views.

As a practical (or what Harman calls "psychological") explanation for the strong belief constraint on intending (or willing), Harman points out that considerations of maximizing expected utility argue in favor of calculating not only what one desires but also one's expected chances of success. Harman adds that, since Tversky and Kahneman's studies (1974), we have known that people are not good at juggling multiple measures of probabilities, and that it may work out for creatures like us that if we will only what we intend and intend only what we believe we will do, then we would not be faced with such calculations.

Audi (1986) gives four reasons why intending is connected with believing one probably will do what one intends. First, if intending is not connected with believing one will do what one intends, we would likely plan less to count on the assumption that we will do what we intend to do (forcing us to tend to discount the role of intending in planning). Second, since we do not discount the role of our intending in planning, it is rational to believe that one will do or probably do what one

intends. Third, we assess persons of good intent to be those who regularly do and believe they will do what they intend. And fourth, the belief condition helps clarify the notion of confidence associated with intending: "the stronger the belief that one will A, the more confidence we have regarding the intention; and when confidence is reversed, intending gives way to hoping, or if S comes to believe A-ing to be impossible, to mere wishing" (27).

Following along these lines, when one thinks of planning (Bratman 1987) and rationality constraints on means-ends reasoning that go along with planning, one can readily see the relationship between putting plans of action into play and having reasonably strong belief about one's chances for success (or at least, constraints on consistency of plans). For this one may need somewhat positive beliefs about success.

However, if one is trapped on a ledge and the only way to get off is to jump and one does not want to die, planning constraints seem to fall by the wayside. If one doesn't jump, one estimates that one will die. If one does jump, chances are good that one may die. If one doesn't jump, one will surely die, and the desire to stay alive may compel one to screw up courage and jump. One may also believe that one must be fully committed to the effort. A halfhearted effort would surely fail. The chasm is wide. Here planning constraints may seem to fade, and the motivation and initiation of a splendid and powerful leap needs to be effected, not the coordination of plans or avoidance of combinatorial probability calculations. One needs to get off the ledge and one needs an inspired leap. It may be that only a strong, motivating intention to escape can do the causal work. Can't one intend to escape by making the jump? The consensus of opinion from all but a few above is that one can *try*, and *intend to try*, but not intend to leap across the chasm (and thereby escape). But why would one not intend to leap across the chasm if this is a "one-off" act and not a matter of something that will lead to possible conflict or inconsistency of planning? It may be that when the strong belief requirement is placed on intending, it is not with such one-off acts in mind. Attending to cases such as these may lead to the view that intending requires only absence of a negative belief about impossibility of success.

Thalberg (1972, 109) notes that there is a general background of the very existence of intentions as a general kind of mental state. Generally one strives for attainable goals. Generally we rely on a person to carry out projects, when informed about the feasibility of the projects. And generally we take extensive planning of specific details as evidence that one's intentions are firm. Thalberg adds: "Such facts are woven into the pragmatic background against which we use the notion of intending. If these circumstances did not obtain, our concept of intending would change. . . . But notice that these facts and similar ones are only *general* conditions for having the concept of intention we have. Consequently it does *not* have to be true in each particular case" (109).

Still one may think, for reasons like Bratman's (1987), that most intentions are distal (i.e., intentions directed toward the more distant future). One may think of intentions as typically elements in larger coordinated plans—plans that typically involve different agents or the same agent over different times and that there are "consistency" constraints on such planning that require a stronger belief component

(Bratman 1987, 108–9). However, questions can be raised about whether even such typical intentions require stronger belief conditions.

Here is an example from Mele:

> There are three seconds left in the basketball game. The Pistons are behind by four points, and Dumars has just drawn a foul. The players have a standing plan for just this situation. Dumars will sink the first free throw. He will throw the second one hard off the front rim so that a teammate may get the rebound while Dumars runs to the three-point line. Finally, if a teammate does get the ball, he will quickly pass to Dumars, who will sink a three-point shot. (1992, 137–38)

Mele explains that Dumars and his teammates believe that their chances of executing the entire plan "are slim." Yet it is their best or perhaps only chance to win or tie the game. They have rehearsed the plan and it clearly is a coordinated effort, but Mele adds: "It's coordinative success need in no way depend upon their believing that the plan will (probably) be successful" (138). With regard to the difference of opinion on whether the belief requirement on intending is for a strong or weak belief, Mele thinks his example shows "there is nothing about coordinative capacity itself that calls for these belief constraints on intention" (138).

Mele goes on to argue (1992, 140) persuasively, to my mind, that the functional role of intentions to initiate action plans, initiate practical reasoning, or terminate deliberation can be played by intentions independently of whether there is a strong or weak belief constraint on intending. We shall return to this notion of the functional role of intentions below, when I compare intending to trying.

III.B: *Conversational Implicature*

Now for our second sort of reason as to why there may be such a wide range of views. For some, and I think this is especially true in Audi's case, the stronger belief constraint seems correct because, frankly, it sounds wrong or misleading to say "I intend to A, but I may not do A." For others this seems to be a pragmatic constraint on the use of the word *intend*, not a semantic constraint or other sort of constraint on the cognitive state of intending.

Let's consider some evidence from the literature. Those who endorse a fairly strong positive belief requirement on intending often appeal to the "oddity" or "misleading" nature of avowals of intention that are followed by claims that would undercut an implied strong belief.

Here are some typical claims.

1. "Note how odd it is to say such things as 'I intend to go to your paper, though it is not likely that I will make it'" (Audi 1993, 57).

2. "I intend to call but I won't, while not logically inconsistent . . . is most odd" (Davis 1984, 133).

3. "To say that Isabel intends to beat Judy in their next chess game would seem to be to say something stronger than Isabel merely intends to try to beat Judy" (Harman 1986, 364).

4. "Ordinary speakers of English are disinclined to attribute intentions to A to agents who estimate their chances of succeeding in A-ing as less than even. What

accounts for this, I suspect, is not just that there is something very odd about such *assertions* as 'I intend to A but I believe that I probably will not A' but also that the ordinary concept of intention incorporates a confidence condition—perhaps only a negative one" (Mele 1992, 148).

5.Saying "I intend to do it" or "I will do it" is much like, or on occasion identical with, promising to do it. If I say any of these things in the right context, I entitle a hearer to believe I believe I will do it . . . under certain conditions, then I *represent myself* as believing that I will. I may not intend that my hearer believe I will, but I have given him ground for complaint if I do not. These facts suggest that if I not only say "I intend" or "I will" in such a way as to represent myself as believing I will, but am sincere as well, then my sincerity guarantees both that I intend to do it and that I believe I will. Some such line of argument has led many philosophers to hold that intending to do something entails believing that one will (Davidson 1980, 91).

And on the other side:

6."The argument just sketched does not even show that intending implies belief. The argument proves that a man who sincerely says 'I intend to do it' or 'I will do it' under certain conditions must believe he will do it. But it may be the saying, not the intention, that implies the belief. And I think we can see this is the case" (Davidson 1980, 91).

7."Once we have distinguished the question of how belief is involved in avowals of intention from the question how belief is involved in intention,[3] we ought to be struck with how dubious the latter connection is" (Davidson 1980, 91).

8."In writing heavily on this page I may be intending to produce ten legible carbon copies. I do not know or believe with any confidence, that I am succeeding. But if I am producing ten legible carbon copies, I am certainly doing it intentionally" (Davidson 1980, 93).

9."I think X spoke correctly and accurately, but misleadingly, when he said 'I intend to go to the concert'. He could have corrected the impression while still being accurate by saying, 'I now intend to go to the concert, but since I may be put in jail, I may not go there'. A man who says, 'I intend to be there, but I may not be' does not contradict himself, he is at worst inscrutable until he says more" (Davidson 1980, 94).

10."We can now see why adding, 'if I can' never makes the statement of an intention more accurate, although it may serve to cancel an unwanted natural suggestion of the act of saying one intends to do something" (Davidson 1980, 100).

11.[About Audi's "try with the hope examples"] "I believe the difficulty here stems from the fact that 'I intend to A', said before acting, carries a note of confidence of success that admittedly is inappropriate for such cases" (McCann 1998, 101).

12.Bratman runs an example where a man who doubts that he may be able to move a log from his driveway is asked if he "intends to move it" and responds "No, but I intend to try." McCann (1998, 210) says this: "Instead of taking the 'No' as a disavowal of an intention, therefore, we should take it as a pragmatic weakening of the avowal aimed at diminishing audience expectation about . . . success."

Now I'm looking for a way to make sense of both sets of quotes. My preferred way of doing this is to maintain that "I intend to A" or "S intends to A" does not literally

(semantically) imply that I (or S) have a strong positive belief that I (or S) will (or probably will) A. Instead it implies that one has A-ing as a goal, and that one has a plan to A upon which one has settled and is determined to act.

Now if I assert "I intend to A" and no more, I am giving it to be understood that I have a certain air of confidence in success (due to the usual Gricean maxims of conversational implicature). Were I to want to temper that implicature, I would say instead "I intend to try to do A" or "I'm going to try to do A." Indeed, I think that saying one is going to try is a way of canceling the implicature of confidence in one's intention because I think trying to A contains an intention to A within the attempt. We normally only say we are "going to try," when we are unsure of success or at least low in confidence. But in my view that does not show there is not an intention to A in place causally guiding the attempt. I will discuss this more below.

It may be said that the difference between intending and trying *is* the difference between being confident and not being confident, but that would be to ignore the functional role of intending. I will say more about this in what follows. At this point, since we are considering various implicatures of intention talk, I want to register my own example of conversational oddities.

Consider Al's shot of the basketball from mid-court. According to Audi, Mele, and a host of others on our list, it is okay for Al to try to make it (or intend to try), but not to intend. However, suppose Al says, "I'm going to try to make it, but I have no intention of it going in." I trust this has as much oddity to it as his saying "I intend to make it, but I'm not very skilled at this." If this is right, and there is linguistic oddity in such a statement of denial of intention when one's confidence of success is low and one is "only trying," what explains that oddity? In what follows, I will suggest that there is an intention (to do A) within the very attempt (to do A) and this is what explains it.

III.C: *Trying*

The third source of variation is differences of view about the nature of trying itself. Enc (2003) accepts view 5 and explains the variation in views between strong and weak constraints on intending by appeal to the difference between intending and trying. He gives the contrast between Samson intending to bring down the roof of the temple (thinking his God is on his side) versus Samson trying to bring down the roof of the temple (thinking his God may have abandoned him). Enc says: "When the context is one in which the explanation aims merely at describing a means-end belief of the agent, and is not expected to reveal anything about the degree of confidence involved in such beliefs, intentions may be attributed without being sensitive to the above discrimination . . . and also this may be why [view 4] . . . has been so controversial" (198).

But what explains this "context" effect? Is it that conversational implicature kicks in for "intends" that is cancelled for "tries"? If it is the semantics of "intend" and "try" that are at work, why does "context" matter? Furthermore, Enc explicitly says: "attributing the intention to do A to an agent carries with it the information that she believed that she will do A. . . . This information may be cancelled by replacing the intention attribution with the attribution of trying" (216). As I indicated above, I would attribute this to the pragmatics of the terms, not to their semantics.

Enc gives another example (2003, 198–99) of a spur of the moment, New Year's resolution to lose weight that he may make with no intention to reduce calorie intake or increase exercise output, or in light of any other weight loss planning. He says in this case the intention is not accompanied by a belief that he will lose weight. But he goes on to say it is not a violation of view 5 because it is actually no more than a "wish," and is not really an intention. It is not an intention because it is not really an element in an action plan that does any causal work or ever gets implemented. So he denies that such cases contradict view 5.

So I take Enc to be giving two reasons to account for the variation on our list. First, the weaker belief-constraints only are given when one is ignoring attributions that may provide information about one's confidence of success. Second, attributions of intention, to be informative, must refer to states that are causally explanatory of some actual act tree the agent will implement. Attributions by the agent or others that amount to low belief of success may amount to no more than wishes, in extreme cases, but may be seen as weak intentions. Intentions, as opposed to hopes or wishes, must do some causal cognitive work.

Now I take it that Audi and Mele and many others would agree that a difference between intending and trying is that they not only have different levels of confidence about them, but also different cognitive goals. This is clearly the suggestion of Audi's claim that what one cannot intend (because one's belief about the probability of success is too low), one can try with the hope of success. It surely sounds as though, on this view, what one tries is different from what one would have intended. And Mele (1990a, b, 1994) has argued that one can intend to try (and try) what one does not even desire. This surely would not be the case for what one intends. Mele does not say one can intend what one does not desire (as far as I know).

Thalberg (1972) and Ludwig (1992) claim that one can try to do the impossible (but we have seen reasons to reinterpret Thalberg's claims). Ludwig's example is that one need not believe it is possible to start his car parked in a neighbor's drive in order to try to start it. He maintains that one can try to start it, if only to show the neighbor that it won't start. However, Ludwig does not claim that one rationally can intend the known impossible.[4]

Bratman (1987) famously argues that one rationally can try to do each of A and B, knowing that one cannot succeed in doing both, but one cannot intend to do each (knowing one cannot succeed in both). For to intend each would make one's intendings (but not one's tryings) irrational. With the exception of Thalberg, all of these views place different constraints on intending and trying, and in some cases because there is a shift in cognitive object of the action.

Of course there are also those who think differently about trying. Adams (1994a, 1994b, 1997) has argued that trying to A contains an intending to A within. Adams and Mele (1992) have argued that tryings are "intentions at work." Thalberg (1972) and Hugh McCann (1998) also hold that trying to A includes intending to A. "My final point will be that the principle 'If you try X, then you intend X' applies to a striking range of attempts to do the impossible" (Thalberg 1972, 113). "No one who gets up to open the door does so without the intention of opening the door, and neither does anyone who tries to do something act without the intention of doing it" (McCann 1998, 101). "*Trying* never names a unique species of action, but rather functions always

as a general name for the business of *going about* the intentional performance of action." (McCann 1998, 104). We find this in Hampshire, as well: " 'He is trying to do so-and-so' already states the agent's intention, with the added implication that there is some difficulty and a possibility of failure" (Hampshire 1960, 107).

As I say, some have argued that trying to A or acting with the intention of A-ing (but not really intending to A full stop) have different cognitive (intentional) objects. Bratman says:

> Let's call *what* is intended, *what* one endeavors to achieve, *what* one does inten-tionally, the *intentional* objects of the elements of the standard triad. In typical cases of intentional action we not only have all three elements of the standard triad, but we also have a *match* in their intentional objects: what is intended, what one endeavors to achieve, and what one does intentionally, all match. The point I want to make here, however, is that even when all three elements of the standard triad are present there will not always be such a match in their intentional objects.
>
>   In the video-game example, I endeavor to hit target 1, and I hit that target intentionally; yet I do not intend to hit target 1. I do intend to do something—to shoot at the target, say—but the consistency demands on intention give us reason to doubt that what I intend includes hitting the target. (1987, 134)

In regard to whether intending to A and intending to try (or trying to A) have different cognitive objects, Thalberg (1972) has an instructive example to support the contrary. Thalberg's example comes in reply to remarks by Hampshire that basi-cally when one tries (as opposed to intends) to A, one only intends to "come as near as possible" to doing A. Thalberg uses the example of a lifeguard trying to revive a drowning victim. Suppose the lifeguard has serious doubts about the revivability of the victim, yet applies artificial respiration nonetheless. "Does he intend to revive him? What would it mean to deny this, and maintain that he only aims 'to come as near as possible?' In this case there is no intermediate result, closer to reviving him than failing" (1972, 110).

I have argued (Adams 1997) that one cannot try to do what one believes to be impossible, and that when people give examples to the contrary we can see that in the examples actors are just going through the motions of one who is trying, but don't have the necessary cognitive accompaniment. My example is this: Can one try to move the St. Louis Gateway Arch by merely pushing? Surely one can exert much effort. But if one knows or believes that one cannot really move it, is one really trying? Or is one merely straining and grunting and exerting maximal effort (all of which one knows one can do). I argue for the latter position. If I'm right, the difference between intending to do A and merely trying to do A is not due to a change in the cognitive goal (intentional object). The object is still A. The only thing that changed is the confidence of success in A-ing.

For reasons such as this, it is not plausible that a lack in confidence of success changes one's cognitive content or object when trying to do something (vs. intend-ing).[5] So it is not plausible that since Al isn't confident that he will sink the shot from the mid-court line that his intention to try or his trying takes only "throwing the ball as hard as I can in the direction of the basket" not "making the shot" as its representational content. For then Al would not be trying to make the basket, just

trying to "throw the ball as hard as he could in the direction of the basket." Indeed, Al *knows* he can do that—so there is no reason why he could not flat-out *intend that*, rather than *try*. So, it is not plausible to think that the cognitive goal or object changes in an intention to try or a trying (when one's confidence of success is low).

Furthermore, given the views of Audi and Mele and others about trying, if there are weak belief constraints on trying (as most maintain who hold that there are strong belief constraints on intending), then if McCann and I are right, there would be weak constraints on the intentions within the attempts as well. So, even disagreement on the nature of trying itself can lead to differences in view about the belief constraints on intending.

So, these three sources of variation may explain the variety of views on belief constraints on intending. The question now is what to do about it.

IV. Functional Role of Intention

Let's proceed by considering the functional role of intending. I want to maintain that it is highly plausible that the functional role of the cognitive state in intending is the exact same as the functional role of the cognitive state that exists in trying—same content, same causal role. For that reason, I want to maintain that the cognitive (and motivational) state existing with trying to do A *just is* an intention to do A.[6]

First, the case of one Connie: Mele (1992) uses the example of Connie trying to make a free throw and finding out that she is not the 80 percent shot-maker that she thinks she is. This does not affect the way she shoots or her trying. Nothing relevant to the causal role of intending changes with a diminished belief (131). Before she shoots, she intends to make the shot. After she learns of her free throw record, she only tries with the hope or intends to try. Mele doesn't say such news would never affect one's shot no matter how low one's confidence goes, but in Connie's case, she learns her percentage has gone just below 50 percent (just below where he and Audi and others would say she could intend to make the shot), but this does not affect her shot. Notice that nothing relevant to the cognitive role and motivational role of the state that causally features in her taking the shot seems to change when she learns the bad news.

Second, the case of two Connies: Later Mele offers us two Connies. Connie1 has the belief that she is an above 50 percent free throw shooter. Connie2 has the belief that she is below 50 percent. Connie2 "is guided no less by her plan for sinking the shot than [Connie1] is by hers. Given that the plans are identical in all relevant respects, the behavior-guiding capacity of the plan component of the two Connies' intentions does not rest on her being on either side of the doxastic states required by views [4 and 5_{fa}]" (Mele 1992, 136). This works to my advantage because I can say the cognitive states of the two Connies are functionally equivalent (which is Mele's point)—functionally equivalent to intendings.

Third: let's consider some of the main cognitive functions of intentions. Here are some of the things that philosophers say about their functional role. I shall concentrate on the cognitive (belief-like) side of intention, not the conative (desire-like) side, but naturally intentions have both sides. I will look mainly at the cognitive side here because we are stalking the connection between intention and belief.[7]

Intentions have the function to represent goals one sets out to attain. They are cognitive representations (Brand 1984, Adams 1986, Adams and Mele 1989, Bratman 1987, Mele 1992) of actions (or states of affairs) that an agent desires to bring about and believes [insert one of seven views above here] can be accomplished. I will focus mainly on the cognitive representation for present purposes.

Intentions can initiate and sustain action (Brand 1984; Adams 1986; Adams and Mele 1989; Mele 1992, 130; Adams 1994a; McCann 1998). One starts acting when one believes the time to act is now. One's continuing to (do or attempt to do A) causally depends in the right way on one's continued intention to do A. And one ceases one's part in acting when the intention ends.[8]

Intentions can serve as guides and monitors to ongoing actions (Adams 1986; Adams and Mele 1989; Mele 1992, 136; Adams 1994a). One compares what one has (or hasn't) done with what one intends to do and makes corrections to diminish the difference.

Intentions can be elements in coordinative plans (Harman 1986, Bratman 1987, Mele 1992, McCann 1998). What I do (and intend) may depend on what you will do (and intend) in a larger coordinated plan. This places rationality constraints on consistency of intentions.

Intentions can terminate practical reasoning (Bratman 1987, Mele 1992, McCann 1998). That is, when one settles on doing A as what one intends to do, then one no longer reasons about whether to do A (except when changing one's mind), but reasons about how best to achieve A.

Intentions, when satisfied, can be extinguished, and thereby terminate action (Adams 1986, Bratman 1987, Adams and Mele 1989, Mele 1992).

Fourth: now let's reexamine trying. We will find each of these functional components within trying, lending excellent support for the view that intentions are the components within attempts or endeavors that play these cognitive roles. Let me attempt to forestall one obvious response to my strategy. Someone like Mele (1990a, 1990b, 1994) will concede that all trying incorporates[9] an intention, namely, the intention to try. So trying to A incorporates the intention to try to A, not the intention to A. But I have already indicated that this changes the intentional object from A-ing to something less than A-ing and that seems to me to have all the problems I mentioned using examples from Thalberg and one of my own. In addition, we are working toward understanding what trying to do A amounts to. One may indeed intend to try to do A, but the question is what trying is. I think we have to get clear on what trying to do A is before we can become clear on what intending to try to do A is. To simply assert that trying to A has the intention to try to A in it doesn't answer the question of what is the cognitive component of a trying to do A. That is my quest. When we know what that is, then we will know what it is when one intends to try to do A (Danto 1966, Adams 1994a, 1994b).

The same sort of claim can be made for the desire component of trying. One can say that trying or intending to try to do A requires only a goal of doing A, not an intention to do A. Like McCann, I agree that "unless 'guiding desires' play exactly the functional role usually reserved for intention, they cannot yield intentional behavior" (McCann 1998, 208). But I will not discuss desire and intention here, since I've done that at length elsewhere (Adams 1994a, 1994b).

Let's look again at trying. Adams and Mele (1989) say this about trying. Trying does not involve making a special effort or overcoming an uncommon difficulty.

Rather, trying is involved in every case in which an intention to A issues in intentional A-ing. Trying is the agent's contribution to what he succeeds in doing when he executes his intention. Any linguistic oddness to saying things like "Benny tried to raise his arm" when this is something Benny easily can do, is due to "linguistic oddness" (i.e., implicature) not logical or semantic oddness about "trying."

Trying is not a mediator between intentions and intentional actions.[10] Tryings are not ballistic. Tryings depend upon the continuation of the relevant intention one is trying to achieve. Tryings result from the normal functioning of appropriate intentions. Trying to A is an event, in a chain of events that begins in the brain and ends, when successful, in A-ing. As we say, tryings to A begin in the brain and their initiation is the immediate effect of the formation or acquisition of a proximal intention (a here-and-now intention) to A.

Hence, proximal intentions, while they are not themselves tryings, initiate tryings and they sustain tryings. Complete tryings are actions, but trying may begin (with signals going down the efferent pathways), prior to bodily movement or other action beginning. So tryings begin when proximal intentions start playing their causal-functional role in the brain.

An agent stops trying to A when the trying brings about the relevant bodily movement that is the agent's contribution to A-ing.

On this view, tryings simply are intentions at work. On this view tryings clearly have cognitive components that play all the roles that intentions play. They have elements that represent the goal of the action. They have elements that intitiate, guide, and sustain the action (or trying). They have elements against which one can monitor the success or failure of the trying as it unfolds. Tryings can be coordinated into larger attempts of a single agent or of multiple agents. Tryings end practical reasoning about what to try—clearly one has settled on the action or state of affairs one is trying to bring about. And trying ends upon successful completion of the attempt. Of course it can end upon unsuccessful completion of an attempt too, when one decides to end the trying for reasons other than success.

The point I want to stress is that there is clearly a cognitive component to trying that plays the same functional role as intention. The most straightforward explanation of that is that in an attempt to A there exists an intention to A, an intention with which the attempt is being conducted.

McCann (1998) says this about trying. Trying is essentially intentional action. It is something we do, and something where we know what we are doing (trying to do) in a noncorrigible way. Further, McCann claims it is "self-contradictory" to assert "He is trying to A, but he is not acting with the intention of A-ing" (101). McCann holds that when a person who is trying succeeds, his trying is a part of the process that is A's performance. He adds that trying is the name of the "general business of going about the intentional performance of the action itself" (104). He maintains that we only talk about "trying" when we want to separate in thought the performance from the success conditions of the acting.

I will just add that since McCann thinks all tryings to do A contain an intention to A with which one acts, he sees tryings as inheriting all of the cognitive functions on our list above for intentions. In this, his view of the cognitive role of tryings fairly well coincides with that of Adams and Mele (1989). Any departure is due to

McCann's emphasis on volition, as opposed to our emphasis on proximal intendings, when it comes to the initiation of tryings. Other than that, there is not much difference in our views (even though we came to them from quite different directions).

So what is wrong with someone like Audi saying that since Al is unconfident of making the basket, he is trying to make it with the hope of making it, not with the intention of making it? In the end, he may be hoping to make it as well, but trying to make it is to act with an intention of making it. The intention is not held with confidence, and so Al would not want to mislead a hearer and express the intention in a way that could not be conversationally canceled. He may cancel it by saying he will "try" or that he only "intends to try." But I would maintain that Audi's *hope alone* is not strong enough to prompt an attempt. We all hope for world peace, but not that many of us are trying to bring it about. Hope may be in the hearts of those who are trying to bring it about, but, if I'm right, it takes more than a hope to initiate, guide, and sustain an attempt; it takes an intention.

V. Acting with an Intention

There is precedent in the literature for saying that expressions such as "acting with the intention of doing A" are not legitimately expressing an intention to do A. We see it in Audi who accepts that "we normally cannot explain why S A-ed (where her A-ing is intentional) by citing her intention *to do it*, though we normally can explain an intentional action by citing an intention *with which* S performs it" (Audi 1993, 70–71). This implies that the referent of "intention" is not exactly the same in both sentences.

We see it in Davidson: "If someone digs a pit with the intention of trapping a tiger, it is perhaps plausible that no entity at all, act, event, or disposition, corresponds to the noun phrase, 'the intention of trapping a tiger'" (Davidson 1980, 88). Yet even Davidson goes on to add: "But it is not likely that if a man has the intention of trapping a tiger, his intention is not a state, disposition, or attitude of some sort. Yet, if this is so, it is quite incredible that this state or attitude (and the connected event or act of *forming an intention*) should play no role in acting with an intention" (88). I would go on to add that not only would it be incredible if it were not a state or disposition or attitude of some sort, but that it is incredible if it is not *the intention to do* A. The functional role of the state *is that of intending*, even when it occurs in trying (as I have argued above).

In comparison, Harman expresses an intermediate position. "It seems marginally better to me to say he throws with the intention of making a basket [than to say he intends to make the basket _fa_]. More generally, I find I am more inclined to see a difference between intending and intending to try than to see a difference between acting with an intention and acting with an intention to try" (Harman 1986, 366).

Nonetheless, I maintain that in the light of the exact similarity of functional role of the cognitive element in trying and in intention, that there is no good reason to think that "intention" in "with an intention" does not take its usual referent. In general, "with an X" is not an intensional context. If Lois Lane was flying with Superman (in Superman's arms) then Lois Lane was flying with Clark Kent, since Superman is Kent. True, the *content of* the intention can generate an intensional context, but hardly the mere *reference to* the mental state.

Idioms or metaphorical use of language could create a context in which "intention" doesn't really mean *intention*. But it is not really plausible that "S did X with the intention to do Y," or "S tried X with the intention to Y," are metaphors or idioms. So this sort of reason to be suspicious that we have normal reference to an intention in these contexts falls short.

Bratman (1987, chap. 9) has given an argument that "acting with an intention" has multiple senses: a *strong* sense in which "acting with the intention to A" does entail having the intention to A, and a *weak* sense in which it does not. However, his defense is based upon his video game example, where he attacks the Simple View of intentional action. He argues that one can't intend to do A (hit target 1) and intend to do B (hit target 2) when one knows one cannot do both (hit both target 1 and 2), but that one can try to do each (hit each). "Trying to do each" only implies acting with an intention in Bratman's weak sense. So trying to do each does not commit one to intending to do each of A, B, in Bratman's eyes.

However, Bratman's sole reason for distinguishing two senses of "acting with the intention to A" is his argument based upon the video game example itself. If one rejects those arguments against the Simple View (as do Adams [1986] and McCann [1986]), one rejects the basis for the two senses. McCann and I believe that if one can try to do each of A, B, without irrationality, then one can try each with the intention to do each, without rationality. Our weak negative belief requirement on intention (and trying) frees an agent from irrationality for the same reasons Bratman's agent (who is only trying to hit each) is freed. Since one need not believe one will (or probably will) hit each target, one may try with the intention to hit each (without irrationally intending two things that one knows cannot be conjoined or "agglomerate," to use Bratman's word).

So McCann and I find sound the argument that Bratman thinks is equivocal (due to his two senses of "act with the intention to A").

Here is the argument:

From (1) S tries to A,

to (2) S acts (or tries) with the intention to A (of A-ing)

to (3) S intends to A (Bratman 1987, 133).

VI. Conclusion

In the end, I think Audi and I (and those he has inspired) agree that "I intend to A" is a confident expression of an intention to A. We may not agree that "I intend to try to A" or "I will try to A" is a nonconfident expression of an intention to A. But that is the view I have defended here. I think my view has the advantage of elegance and plausibility. It explains why intentional acts are intentional by placing an intention in their causal history and production, whether expressed as an intention *full stop* or as an intention to *try* or a *trying*. It explains why an intention or attempt to *do A*,

has *A* as its content (or intentional object), by placing the cognitive representational content *to A* within the intention that leads to the action (or attempt).

My support and defense also explain the kinds of intuitive evidence Audi and his followers find so compelling. My view of the weak belief constraint on trying and intending handles the implausibility of expressions of intention in situations of weak belief. It does so by reference not to the semantics of "intend" but to the pragmatics of intention talk. "I intend to, but I may not succeed" may sound strange, but no stranger than "I'm trying to, but I have no intention of succeeding." On the current view, the former is conversationally odd but true, while the latter is an explicit semantic inconsistency. Both expressions sound odd, but for different reasons. The current view can explain the oddity of both, while offering a single, consistent, core view of what makes an act or attempt intentional with respect to the intentional object or goal of doing A. Basically, it is the intention to A. I hope to have shown that, armed with the right view of trying and the difference between the semantics and pragmatics of intention talk, one can defend this view that at the heart of every intentional doing A (or trying to A) lies the intention to A. There may be variation in the strength or conviction of belief that one will A (or the likelihood of one's A-ing), but there will be an intention to A, nonetheless.[11]

Notes

1. In this paper I will not discuss matters such as luck or causal deviance that enter into whether one's action may be intentionally performed, even if intended. I have discussed these matters elsewhere (Adams 1989).

2. Audi's first paper on intention was published in 1971 and influenced most of the philosophers who have written on the topic since that time. His view has been so influential that I have chosen to work a portion of it into the title of this paper.

3. Davidson says one may complain that this is about acting "with an intention" as opposed to "pure intention," and adds: "Nevertheless, it is hard to imagine that the point does not carry over to pure intending" (92).

4. We all can intend the physically impossible, as long as we don't know it is impossible or are irrational. I'm excluding these cases for this paper.

5. Of course, I don't think there is a difference because I think an intention is embedded in a trying.

6. I won't argue this for the conative state here, but have elsewhere (Adams 1994a).

7. For the conative side see Adams 1994a, 1994b, and 1997.

8. Of course the action may continue if it takes time to unfold even after the intention ends, in some cases. I plant a seed of a perennial plant capable of living on its own. I grow the plant long after my intention ends.

9. By "incorporate" I mean nothing more than Mele accepts that trying depends on intentions. He does not think intentions are "components" of tryings, as I do.

10. For an excellent early recognition of this see Danto (1963, 440; 1965, 265; 1966, 56).

11. I have the greatest respect and admiration for the philosophical work of Robert Audi in the area of action theory and elsewhere. It has been a pleasure to know him and learn from him through the years, and it is a special privilege to be able to discuss this work and its influ-

References

Adams, F. 1986. Intention and Intentional Action: The Simple View. *Mind and Language* 1: 281–301.

——. 1989: Tertiary Waywardness Tamed. *Critica* 21: 117–25.

——. 1994a. Trying, Desire, and Desiring to Try. *Canadian Journal of Philosophy* 24: 613–26.

——. 1994b. Of Epicycles and Elegance. *Canadian Journal of Philosophy* 24: 637–42.

——. 1997. *Cognitive Trying. In* Contemporary Action Theory, *vol. 1, ed. G. Holmstrom-Hintikka and R. Tuomela, 287–316. Dordrecht: Kluwer.*

Adams, F., and A. Mele. 1989. The Role of Intention in Intentional Action. *Canadian Journal of Philosophy* 19: 511–32.

——. 1992. The Intention/Volition Debate. *Canadian Journal of Philosophy* 22: 323–38.

Armstrong, D. 1980. Acting and Trying. In *The Nature of Mind*. Ithaca, N.Y.: Cornell University Press.

Anscombe, E. 1957. *Intention*. Oxford: Blackwell.

Audi, R. 1986. Intending, Intentional Action, and Desire. In *The Ways of Desire: New Essays in Philosophical Psychology*, ed. J. Marks. Chicago: Precedent.

——. 1993. Intending. In *Action, Intention, and Reason*, ed. R. Audi. Ithaca N.Y.: Cornell University Press.

——. 1997. Intending and Its Place in the Theory of Action. In *Contemporary Action Theory*, vol. 1, ed. G. Holmstrom-Hintikka and R. Tuomela, 177–96. Dordrecht: Kluwer.

Baier, A. 1970. Act and Intent. *Journal of Philosophy*, 67: 648–58.

Beardsley, M. 1978. Intending. In *Values and Morals*, ed. A. Goldman and J. Kim. Dordrecht: D. Reidel.

Brand, M. 1984. *Intending and Acting*. Cambridge, Mass.: MIT/Bradford.

Bratman, M. 1987. *Intention, Plans, and Practical Reason*. Cambridge, Mass.: Harvard University Press.

——. 1990. What Is Intention? In *Intentions in Communication*, ed. P. Cohen, J. Morgan, and M. Pollack. Cambridge, Mass.: MIT/Bradford.

Danto, A. 1963. What We Can Do. *Journal of Philosophy* 60: 435–45.

——. 1965. Basic Actions. *American Philosophical Quarterly* 2: 141–48. Reprinted in *The Nature of Human Action*, ed. Myles Brand. Glenview, Ill.: Scott, Foresman. (Page references are to this reprint.)

——. 1966. Freedom and Forbearance. In *Freedom and Determinism*, ed. Keith Lehrer. New York: Random House.

Davidson, D. 1980. *Essays on Actions and Events*. Oxford: Clarendon Press.

Davis, W. 1984. A Causal Theory of Intending. *American Philosophical Quarterly* 21: 43–54 (pages quoted are from Mele 1997, where reprinted).

Enc, B. 2003. *How We Act*. Oxford: Clarendon.

Grice, H.P. 1971. Intention and Uncertainty. In *Proceedings of the British Academy*, 57. London: Oxford University Press.

Hampshire, S. 1960. *Thought and Action*. New York: Viking.

Hampshire, S., and H. Hart. 1958. Decision, Intention, and Certainty. *Mind*, 67.

Harman, G. 1986. Willing and Intending. In *Philosophical Grounds of Rationality*, ed. R. Grandy and R. Warner, 363–80. Oxford: Clarendon.

Ludwig, K. 1992. Impossible Doings. *Philosophical Studies* 65: 257–81.

McCann, H. 1986. Rationality and the Range of Intention. *Midwest Studies in Philosophy* 10: 191–211.

——. 1989. Intending and Planning: A Reply to Mele. *Philosophical Studies* 55: 107–10.

——. 1998. *The Works of Agency*. Ithaca, N.Y.: Cornell University Press.

Mele, A. 1990a. He Wants to Try. *Analysis* 50: 251–53.

——. 1990b. He Wants to Try Again: A Rejoinder. *Analysis* 51: 225–28.

——. 1992. *The Springs of Action*. Oxford: Oxford University Press.

——. 1994. Desiring to Try: Reply to Adams' "Trying, Desire, and Desiring to Try." *Canadian Journal of Philosophy* 24: 627–36.

——. 1997. *The Philosophy of Action*. Oxford: Oxford University Press.

Thalberg, I. 1972. *Enigmas of Agency: Studies in the Philosophy of Human Action*. London: Allen and Unwin.

Tversky, A., and D. Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science* 185: 1124–31.

Self-Deception and Three Psychiatric Delusions

On Robert Audi's Transition from Self-Deception to Delusion

ALFRED R. MELE

I have learned a lot from Robert Audi over the years. About twenty-five years ago, he even introduced me to a topic that I did not know philosophers had ever discussed: self-deception. I was invited to comment on a conference paper of his, "Self-Deception, Action, and Will." This was to be my first experience as a commentator; and in order to decrease the probability that I would look like a complete fool, I read everything that Audi had ever written on self-deception and on belief and then proceeded to read dozens of other articles on his topic. By the time I wrote my commentary, I almost had a view of my own on self-deception. I at least had some inkling of a view that I subsequently developed in several articles and a small book. I certainly am indebted to Audi for getting me started on the issue.

In this paper, I will explore a theme in Audi's "Self-Deception, Action and Will" (1982): the relationship between self-deception and delusion. He writes: "Self-deception may pass over into simple delusion, and the transition may be gradual. But we need to distinguish these things; and normally, at least, when we reach a point at which it is clear that S consciously believes p, he has, I think, passed from self-deception to genuine delusion and no longer believes that [$\sim p$]" (140). Here p is a false proposition. A distinctive feature of Audi's view is that the person who is self-deceived with respect to a false proposition, p, does not actually believe that p (147), but, instead, "sincerely" (137)—or "non-lyingly" (139)—avows it or is disposed so to avow it. I have objected to this feature—and other features—of Audi's view elsewhere (Mele 1982; 2001, 52–53, 56), and I will not dwell on our differences here.

In remarking on "a major difference" between his "account of self-deception and inconsistent-belief accounts," Audi writes: "I maintain that S [the self-deceived person] does not believe p, i.e., the proposition which, on some accounts, he consciously believes or believes to some degree" (147).[1] Now, if it were clear that no one who is "in self-deception with respect to p," p being a false proposition, actually believes that p, then there would be a clear difference between cases of self-deception and cases of delusional belief. If, as I believe, typical cases of self-deception regarding p are cases

of false belief that *p*, the question of how self-deception is related to delusional belief is more interesting. In section 1, I provide some background on self-deception and on Audi's view and my view of the phenomenon. In section 2, I turn to delusions and their relationship to self-deception.

1. Self-Deception

Self-deception is often said to be paradoxical. The main paradoxes or puzzles are generated by an understanding of self-deception as largely isomorphic with stereotypical interpersonal deception. The following assumption is common in the literature and is associated with a familiar puzzle about self-deception:

> *D*. By definition, person *A* deceives person *B* (where *B* may or may not be the same person as *A*) into believing that *p* only if *A* knows, or at least believes truly, that ~*p* and causes *B* to believe that p.

If *D* is true, then deceiving oneself into believing that *p* requires that one know, or at least believe truly, that ~*p* and cause oneself to believe that *p*. At the very least, one starts out believing that ~*p* and then somehow gets oneself to believe that *p*. Some philosophers claim this entails that, at some time, self-deceivers both believe that *p* and believe that ~*p* (Kipp 1980, 309). And, it is claimed, this is not a possible state of mind: the very nature of belief precludes one's simultaneously believing that *p* is true and believing that *p* is false.[2] Here we have a static puzzle about self-deception: self-deception, according to the view at issue, requires being in an impossible state of mind.

In fact, *D* does not entail that in all instances of deceiving, there is some time at which the deceiver believes that ~*p* and the deceived person believes that *p*. In some cases of interpersonal deception, *A* has ceased believing that ~*p* by the time he causes *B* to believe that *p*. Imagine that *A* attempts to deceive *B* by mail. *A* tries to deceive *B* into believing that *p* by lying to him in a letter: *p* is false and his assertion of *p* in the letter is a lie. When he sends the letter, *A* is confident that ~*p*, but he comes to believe that *p* by the time *B* receives the letter. If *A*'s lie is successful, *A* deceives *B* into believing that *p* in a way that provides confirmation for assumption *D*. But there is no time at which *A* believes that ~*p* and *B* believes that *p* (see Sorensen 1985).

A philosopher inclined to claim that there is a basis in "the concept of deception" for the thesis that self-deceivers simultaneously believe that *p* and believe that ~*p* need not cave in at this point. It may be true that in *stereotypical* cases of interpersonal deceiving, there is some time at which *A* believes that ~*p* and *B* believes that *p*. And it is open to a philosopher to contend that self-deception is properly understood only on the model of stereotypical interpersonal deception.

The claim that self-deception must be understood on the model just mentioned produces another puzzle about the state of self-deception. In stereotypical cases of interpersonal deceiving, there is a time at which the deceiver does *not* have a belief that *p* and the deceived person does have a belief that *p*. If self-deception is strictly analogous to stereotypical interpersonal deception, there is a time at which

the self-deceiver both has a belief that *p* and does not have a belief that *p*—a perplexing condition, indeed, given its impossibility!³

Audi's position on self-deception avoids the static puzzles in an innovative way. In his view "self-deception with respect to *p* is a state in which S unconsciously knows (or has some reason to believe, and unconsciously and truly believes) that [~*p*], sincerely avows, or is disposed so to avow, that *p*, and has at least one want which in part explains why the belief that [~*p*] is unconscious" (1982, 155; also see 1997a, 132). As Audi understands sincerely avowing that *p*, it does not entail believing that *p*; and on his view, the self-deceived person has the true belief that ~*p* while lacking the false belief that *p* (1982, 147). If he is right about what self-deception is, the static puzzles rest on a mistaken conception of the phenomenon.

Like Audi, I reject D; but I do so on different grounds. In my original commentary (1982) on his "Self-Deception, Action, and Will," I argued, among other things, that there are ordinary cases of self-deception in which the person does believe, falsely, that *p* and does not believe—unconsciously or otherwise—that ~*p*. I will not dwell on that disagreement here, but I will say something more about it. Audi maintains that a kind of tension is necessary for self-deception and that "it is ordinarily represented . . . by an avowal of *p* . . . *coexisting* with knowledge or at least true belief that ~*p*" (1997b, 104; also see 1997a, 144). I am on record as arguing that even if self-deception often involves "considerable psychic tension," such tension is not conceptually necessary for "entering self-deception in acquiring a belief that *p*" (2001, 52–53). The point that I want to make now is that whereas Audi represents the tension in terms of an unconscious true belief that ~*p* and an avowal that *p*, an alternative view represents it in terms of a conscious false belief that *p* and, for example, an awareness of significant evidence that ~*p*. This, of course, is more in line with the commonsense idea that the self-deceived person has a false *belief* that *p*.

Elsewhere, I have distinguished between what I call *straight* and *twisted* cases of self-deception (Mele 1999, 2001). In straight cases, which have dominated the literature, people are self-deceived in believing something that they want to be true—for example, that their children are not using illegal drugs. In twisted cases, people are self-deceived in believing something that they want to be false (and do not also want to be true). For example, an insecure, jealous husband may believe that his wife is having an affair despite having only thin evidence and despite his not wanting it to be the case that she is so engaged. I also have suggested that there are a variety of ways in which our desiring that *p* can contribute to our believing that *p* in instances of straight self-deception. Here are some examples: often, two or more of the phenomena I describe contribute to self-deception.⁴

1. *Negative Misinterpretation*. Our desiring that *p* may lead us to misinterpret as not counting (or not counting strongly) against *p* data that we would easily recognize to count (or count strongly) against *p* in the desire's absence. For example, Don just received a rejection notice on a journal submission. He hopes that his article was *wrongly* rejected, and he reads through the comments offered. Don decides that the referees misunderstood a certain crucial, complex point and that their objections consequently do not justify the rejection. However, the referees' criticisms are warranted; and a few days later, when Don rereads his paper and the comments in a more impartial frame of mind, it is clear to him that this is so.

2. *Positive Misinterpretation*. Our desiring that *p* may lead us to interpret as *supporting p* data that we would easily recognize to count against *p* in the desire's absence. For example, Sid is very fond of Roz, a college classmate with whom he often studies. Wanting it to be true that Roz loves him, he may interpret her refusing to date him and her reminding him that she has a steady boyfriend as an effort on her part to "play hard to get" in order to encourage Sid to continue to pursue her and prove that his love for her approximates hers for him. As Sid interprets Roz's behavior, not only does it fail to count against the hypothesis that she loves him, it is evidence *for* the truth of that hypothesis. This contributes to his believing, falsely, that Roz loves him.

3. *Selective Focusing/Attending*. Our desiring that *p* may lead us both to fail to focus attention on evidence that counts against *p* and to focus instead on evidence suggestive of *p*. Beth's father died a short time ago, not long after her twelfth birthday. Owing partly to her desire that she was her father's favorite, she finds it comforting to attend to memories and photographs that place her in the spotlight of her father's affection and unpleasant to attend to memories and photographs that place a sibling in that spotlight. Accordingly, she focuses her attention on the former and is inattentive to the latter. This contributes to Beth's coming to believe—falsely—that she was her father's favorite child. In fact, Beth's father much preferred the company of her brothers, a fact that the family photo albums amply substantiate.

4. *Selective Evidence Gathering*. Our desiring that *p* may lead us both to overlook easily obtainable evidence for ~*p* and to find evidence for *p* that is much less accessible. For example, Betty, a political campaign staffer who thinks the world of her candidate, has heard rumors from the opposition that he is sexist, but she hopes he is not. That hope motivates her to scour his voting record for evidence of political correctness on gender issues and to consult people in her own campaign office about his personal behavior. Betty may miss rather obvious and weighty evidence that her boss is sexist—which he in fact is—even though she succeeds in finding less obvious and less weighty evidence for her favored view. As a result, she may come to believe that her boss is not sexist. Selective evidencegathering may be analyzed as a combination of hypersensitivity to evidence (and sources of evidence) for the desired state of affairs and blindness—of which there are, of course, degrees— to contrary evidence (and sources thereof).

In none of the examples offered is it said that the person has—unconsciously or otherwise—the relevant true belief that ~*p*. Yet, assuming that these people acquire relevant false, unwarranted beliefs in the ways described, these are garden-variety instances of self-deception.[5] Don is self-deceived in believing that his article was wrongly rejected, Sid is self-deceived in believing certain things about Roz, and so on. So, at least, I have argued elsewhere.

Although I have never offered a conceptual analysis of self-deception, I have suggested the following proto-analysis: people enter self-deception in acquiring a belief that *p* if and only if *p* is false and they acquire the belief in *a suitably biased way* (2001, 120). The suitability at issue is a matter of kind of bias, degree of bias, and the nondeviance of causal connections between biasing processes (or events) and the acquisition of the belief that *p*. I suggest that, as self-deception is commonly

conceived, something along the following lines is a test for a level of motivational or emotional bias appropriate to a person's being self-deceived in acquiring a belief that p: given that S acquires a belief that p and D is the collection of relevant data readily available to S during the process of belief-acquisition, if D were made readily available to S's impartial cognitive peers and they were to engage in at least as much reflection on the issue as S does and at least a moderate amount of reflection, those who conclude that p is false would significantly outnumber those who conclude that p is true. Call this *the impartial observer test*.[6] It is a test for a person's satisfying the suitable bias condition on self-deception. A person's passing the test is evidence of bias suitable for self-deception.

By "cognitive peers," I mean people who are very similar to the person being tested in such things as education and intelligence. Cognitive peers who share certain relevant desires with the subject—as one's spouse may share one's desire that one's child is not using illegal drugs—may often acquire the same unwarranted belief that the subject does, given the same data. But the relevant cognitive peers, for present purposes, are *impartial* observers. At least a minimal requirement for impartiality in the present context is that one neither share the subject's desire that p nor have a desire that $\sim p$. Another plausible requirement is that one not prefer avoidance of either of the following errors over the other: falsely believing that p and falsely believing that $\sim p$. A third is that one not have an emotional stake in p's truth or falsity. The test is a test for a level of motivational or emotional bias appropriate to self-deception. I take the suitability of the impartial observer test—or something similar, at least—to be implicit in the conceptual framework that informs commonsense judgments about what is and is not plausibly counted as an instance of self-deception.[7]

2. Delusions and Self-Deception

With this background in place, I turn to the relationship between self-deception and delusion. The term "delusion" has a variety of uses. Audi's remarks on delusion leave it open how closely he means to associate his use of "delusion" with the standard psychiatric use. He identifies "being deluded as a result of self-deception" with "being deceived in believing p, as a result of having been in self-deception with respect to it" (1982, 142). And he writes: "Since one can pass from self-deception to genuine delusion—in which one is simply deceived in believing p, without the ambivalence, oscillation, anxiety, self-manipulation, or the like characteristic of self-deception—we must distinguish being self-deceived from being simply deceived as a result of being self-deceiving (or of self-deception)" (1982, 143).[8] Since the only substantial body of work I know of on delusion is on the psychiatric phenomenon, I will let DSM-IV set the tone for the remainder of this paper:

> **delusion** A false belief based on incorrect inference about external reality that is firmly sustained despite what almost everyone else believes and despite what constitutes incontrovertible and obvious proof or evidence to the contrary. The belief is not one ordinarily accepted by other members of the person's culture or subculture (e.g., it is not an article of religious faith). When a false belief involves a value judgment,

it is regarded as a delusion only when the judgment is so extreme as to defy credibil-
ity. Delusional conviction occurs on a continuum and can sometimes be inferred
from an individual's behavior. It is often difficult to distinguish between a delusion
and an overvalued idea (in which case the individual has an unreasonable belief or
idea but does not hold it as firmly as is the case with a delusion). (1994, 765)

Two points merit emphasis. As DSM-IV characterizes delusions, they are
exceptionally resistant to contrary evidence and the contrary evidence is very strong.
Both points are reinforced elsewhere in DSM-IV: "The distinction between a delu-
sion and a strongly held idea . . . depends on the degree of conviction with which
the belief is held despite clear contradictory evidence" (275). I take it that "degree
of conviction" (or firmness of belief) here is at least partly a matter of how strong the
contrary evidence would need to be to undermine the belief.

The idea that all delusions are "based on incorrect inference about external
reality" is dispensable. A person might have the delusion that he lacks certain inter-
nal organs (Davies, Langdon, and Breen 2001, 136). Presumably, such a delusion
need not be based on an inference about external reality. With this exception, I fol-
low the quoted gloss.

One way to approach the connection between self-deception and delusion fea-
tures intuitions about cases. Another approach features an investigation of the
causes, in these spheres, of the pertinent beliefs, at least if I am right that, in typical
cases of self-deception, people actually believe the false proposition with respect to
which they are self-deceived. I opt for the latter. I will consider three delusions: the
Capgras delusion, delusional jealousy (or the Othello syndrome), and the reverse
Othello syndrome.

2.1. The Capgras Delusion

Carl believes that his wife has been "replaced by an exact replica or impostor"
(Stone and Young 1997, 327). This is an instance of the Capgras delusion. Part of
the cause in Carl's case, apparently, is a brain injury that deprives him of his nor-
mal affective response to his wife's face (337). Various views have been advanced
about additional causal contributors.

Brendan Maher's model of delusions includes the following two hypotheses:

1. Delusional beliefs, like normal beliefs, arise from an attempt to explain experience.
2. The processes by which deluded persons reason from experience to belief are not sig-
    nificantly different from the processes by which non-deluded persons do. (1999, 550)

Carl has a new way of experiencing his wife's face. On Maher's view, Carl's delu-
sional belief is a product of his attempt to explain this, and his reasoning is not sig-
nificantly different from normal reasoning. The claim about Carl's reasoning
suggests that most normal people who are presented with the evidence Carl has
would come to his conclusion about it. Is it true that if Carl's new way of experi-
encing his wife's face were made clear to nondeluded people, most would infer that
she has been replaced by an impostor? Certainly, Carl's doctors do not infer this,
and it seems that normal people would find the impostor hypothesis about as far-
fetched as his doctors do. Also, there is evidence that some people with "the same

kind of experience of faces as Capgras patients" do not have the Capgras delusion (Davies, Langdon, and Breen 2001, 144).

Considerations such as these have been used to motivate a two-factor alternative to Maher's model of delusion. Davies, Langdon, and Breen suggest two possibilities for a second factor and express a preference for the following idea. Carl experiences his wife as someone who looks just like her but is not really her, and he "accepts this perceptual experience as veridical" (153). The first factor is this experience, which includes the impostor idea as part of its content, as opposed, for example, to the idea's being a hypothesis that is separate from and prompted by the experience. The second factor is a problem that accounts for Carl's accepting the experience as veridical rather than rejecting it as not veridical. The main proposal Davies, Langdon, and Breen offer about the form this problem takes is intriguing. As they observe, "Normal subjects are . . . able to suspend their unreflective acceptance of veridicality and make a more detached and critical assessment of the credentials of their perceptual experiences" (153). Their proposal is that Capgras patients have a deficit in this connection. If this is "the nature of the second factor in the etiology of delusions," then hypotheses that are included in "the patients' own perceptual experience [are] resistant to being critically assessed and recognized as implausible, but hypotheses generated by someone else [are] assessed in the normal way" (Davies, Langdon, and Breen 2001, 153).

Davies, Langdon, and Breen recognize that their proposal generates the prediction that people with this deficit will be led to have false beliefs by their visual illusions in general, and they are clearly uncomfortable about this (153). My own immediate concern is with the bearing of their proposal and Maher's proposal on the impartial observer test. I start with the latter proposal. As I mentioned, it certainly seems to suggest that most normal people who are presented with the evidence the Capgras patient has would come to the patient's conclusion about it—that a loved one has been replaced by an impostor. This suggestion seems so implausible that one wonders exactly what Maher meant. Suppose he believes that it cannot actually be made clear to normal people what the Capgras patient's pertinent experience is like. Then he may say that people who lack that experience cannot actually have the pertinent evidence. This would limit members of the panel, for the purposes of the impartial observer test, to people who have experiences of the sort characteristic of Capgras patients. Exclude all such people who do not satisfy the conditions for membership on the panel. What would the majority of the remainder conclude?

As Davies, Langdon, and Breen observe, "At least some delusional patients show considerable appreciation of the implausibility of their delusional beliefs" (149). Andrew Young writes: "Capgras delusion patients can be . . . able to appreciate that they are making an extraordinary claim. If you ask 'what would you think if I told you my wife had been replaced by an impostor', you will often get answers to the effect that it would be unbelievable, absurd, an indication that you had gone mad" (1998, 37). Even many delusional patients on the panel might judge that Carl's wife was not replaced by an impostor. And a higher percentage of panel members with similar experiences but no delusions might make that judgment. Suppose the overwhelming majority of panelists deem Carl's belief false. Would that constitute good evidence that Carl's treatment of data is motivationally or emotionally biased?

The basic question behind the impartial observer test, of course, is whether something in the motivation/emotion category biased the subject's treatment of data in the process that produced the belief at issue and whether, if this happened, the biasing was robust enough to be appropriate for self-deception. The idea is to strip away the potential motivational and emotional sources of bias while holding the evidence fixed and to see what happens. If the subject's belief is reached by the great majority in the absence of those sources, that is evidence that they did not play the biasing role at issue in the subject. If the converse belief is reached by the great majority, that is evidence that motivation or emotion did play this biasing role in the subject, and the relative size of the majority is evidence of the robustness of the role. But, of course, some nonmotivational and nonemotional factor might be present in the subject in the latter case and absent in the panel, and it might be doing a great deal of causal or explanatory work. This is exactly the situation with Capgras patients if what Davies, Langdon, and Breen propose is correct. That is, what would account for the difference in belief is a certain *cognitive* deficit that is outside the categories of motivation and emotion. And even if it were insisted that people must have that deficit in order to count as *cognitive* peers of the target person, that would make no difference; for the proposed deficit shows up only in responses to one's own experiences.[9]

If what produces the Capgras delusion is a weird experience, together with the removal or disabling of a cognitive mechanism that, in special cases, inhibits a kind of default transition from experience to corresponding belief, the delusion seems to lie well beyond the sphere of self-deception. And independent of the proposal by Davies, Langdon, and Breen, if we lack good reason to believe that motivation or emotion biases the Capgras patient's treatment of data, thereby contributing to the delusional belief, we lack good reason to believe that the delusion is an instance of self-deception. Notice that accepting that the Capgras delusion is explained partly by emotional factors does not commit one to accepting that emotion biases the person's treatment of data. For example, we apparently should accept that a major emotional change—a certain loss of affect—plays an important role in producing the delusion. But this loss is a *cause* of relevant experiential data: *causing* data is one thing; *biasing* a person's treatment of data is another.

Young reports on a "person who experienced both the Cotard and Capgras delusions in sequence" (1999, 577). People with the former delusion believe that they themselves are dead. Young writes:

> This curious association of two unusual delusions has been reported in other cases too, and the key factor seems to be the patients' moods—when in a suspicious mood, they think that other people are impostors, when depressed they think they are dead. There is an obvious parallel here to . . . findings that people with perse-cutory delusions tend to make external attributions and depressed people internal attributions as to the causes of negative events. (577)

What might Davies, Langdon, and Breen say about this? Perhaps, that just as the Capgras patient's experience includes the impostor idea as part of its content, the Cotard patient's experience includes the idea that the subject is dead as part of its content. Perhaps in people with both delusions at different times, their feelings

of suspicion are part of the cause of their having an experience that includes the impostor content, and their depression is part of the cause of their having an experience that includes the "I am dead" content. If so, affective states—depression and feelings of suspicion—would help to explain the delusions. But again they would do so by helping to cause experiential data—these experiences with strange content—rather than by biasing the person's treatment of data. My question is whether the Capgras patient's treatment of relevant data is motivationally or emotionally biased. The evidence and theorizing that I have seen does not support an affirmative answer.

2.2. *Delusional Jealousy*

Next on the agenda is delusional jealousy, one of the types of delusion identified in DSM-IV's gloss on delusion. It is defined there as "the delusion that one's sexual partner is unfaithful" (1994, 765). David Enoch asserts that it is difficult to differentiate "between normal and excessive, excessive and neurotic, and neurotic and psychotic [jealousies]. The various types overlap and the boundaries are blurred" (1991, 52). In section 1, I offered a scenario featuring a jealous husband as an illustration of twisted self-deception. Enoch's assertion suggests that, in the sphere of jealousy, one might be able to locate self-deception on a continuum that includes delusional jealousy, and that being self-deceived in believing that one's sexual partner is unfaithful might at least overlap with delusional jealousy. This suggestion is consistent with DSM-IV's description of the "jealous type" of delusional disorder: "This subtype applies when the central theme of the person's delusion is that his or her spouse or lover is unfaithful. This belief is arrived at without due cause and is based on incorrect inference supported by small bits of 'evidence' (e.g., disarrayed clothing or spots on the sheets), which are collected and used to justify the delusion" (1994, 297).

There are also grounds for pessimism about the suggestion at issue. Michael Soyka observes that "delusions of jealousy are a frequent symptom in various psychiatric disorders. . . . Most . . . patients with delusions of infidelity are schizophrenics" (1995, 118). Barbara Breitner and David Anderson report that "three large studies found 30–50% of the morbidly jealous suffered from psychosis, a similar proportion neurosis or personality disorder, 5–7% alcoholism and the remainder miscellaneous conditions, most commonly organic disorders" (1994, 703). Silva et al. assert that "delusional jealousy rarely exists as the only prominent symptom but is usually found in conjunction with other symptoms, including other delusions and psychotic symptoms" (1998, 616). In a study of twenty people with delusional jealousy, half had directly relevant auditory hallucinations (some of which were commands to attack the partner) and two had relevant visual hallucinations (Silva et al. 1998, 615–16). In a study of 133 demented patients, "All patients with delusional jealousy . . . had at least one other psychotic symptom," as compared with "70.5% of patients without delusional jealousy" (Tsai et al. 1997, 492).

One possibility is that although jealous people who are self-deceived in believing that their partners are unfaithful and people with delusional jealousy believe the same thing, the causes of that belief in the two groups are so different that the groups do not overlap. Consider people with delusional jealousy who have auditory

hallucinations informing them that their partners are unfaithful or visual halluci-
nations of their partners being unfaithful. A proposal like the one Davies, Langdon,
and Breen make about Capgras patients may be made about them. Perhaps, owing
to a cognitive deficit, they accept the "experience as veridical." In those without
such hallucinations, one needs to look elsewhere for causes. Tsai et al. found in
their study of 133 demented patients that "the frequency of delusions of theft [and]
persecutory delusions . . . was significantly higher in the delusional jealousy group"
(1997, 492).[10] When delusions show up in pairs or larger groups, one is inclined to
look for a common cause, especially when the delusions are thematically related.
Infidelity may be viewed as encompassing both theft (by the new romantic partner
or partners) and persecution. To the extent to which one is inclined to see delusions
of theft and persecution as falling outside the sphere of self-deception and as being
explained in part by a cognitive deficit, one should have the same inclination
toward delusions of infidelity in people who have one or both of the other delusions.

2.3. *Reverse Othello Syndrome*

Reverse Othello syndrome is "delusional belief in the fidelity of a romantic partner"
(Butler 2000, 85). As in ordinary, straight self-deception, the person believes some-
thing that he wants to be true. Indeed, a stock example of straight self-deception is
the person who believes that his or her spouse has been faithful despite strong evi-
dence to the contrary—evidence that would lead the great majority of impartial cog-
nitive peers to believe that the spouse has been unfaithful. Accordingly, the prospects
for an important biasing role for motivation in this syndrome might look bright.

 Peter Butler examines the case of a middle-aged man, B. X., who suffered a
severe head injury in a high-speed car accident. His romantic partner, N, ended their
relationship five months later, which B. X. acknowledged. But, despite the absence
of contact with her, he subsequently "developed an intense delusional belief that
[she] remained sexually faithful and continued as his lover and life partner" (86). He
even came to believe that he married N while he was a patient (87). Doctors tested
B. X. for other delusions and found no unrelated ones (88). After some months, "his
delusional system began to break up." A few months later he accepted the truth.

 One important difference between B. X. and his self-deceived counterpart in
the stock example I mentioned is B. X.'s belief that he married N. If there is any
experiential basis for his belief in the marriage, it is something on the order of
dreams or hallucinations. B. X. reported that the wedding "occurred at the Central
Synagogue in front of several hundred guests" (88). He might have dreamed or hal-
lucinated that. Suppose he did. And suppose the dream or hallucination—possibly
a repeated one—was a cause of his belief and was caused in part by a wish to be
married to N or some wish of that kind. Then motivation played a role in B. X.'s
belief in his marriage. But its playing this particular role would highlight a role for
a serious cognitive deficit. When a dream or hallucination is radically out of line
with obvious reality, people without a serious cognitive deficit do not regard the
experience as veridical after they awake or exit the hallucination.

 Butler reports that "when questioned about the absence of photographs of the cer-
emony or corroboration from his family [B. X.] remained adamant the marriage had

occurred and set his communicator to repeat the words just because'" (88). Seemingly, B. X. wants not to think about these absences. He may understandably be motivated to focus his attention on the imagined marriage and to ignore considerations that point to its being a fantasy. The belief that he is married to N obviously gives B. X. pleasure, and entertaining challenges to that belief is unpleasant for him. Selective focusing or attending, which is at work in some ordinary cases of self-deception, may also be at work in B. X. Even if he does not enter self-deception in acquiring the belief that he is married to N, he may be self-deceived in continuing to believe this.

What about people with the Capgras delusion? Might they be self-deceived in persisting in believing that a loved one has been replaced by an imposter? Recall the assertion by Davies, Langdon, and Breen that "normal subjects are . . . able to suspend their unreflective acceptance of veridicality and make a more detached and critical assessment of the credentials of their perceptual experiences" (153). Suppose that people with the Capgras delusion are literally *unable* to do this. Then even if, like B. X., they refuse to reflect on challenges to their beliefs raised in conversation, this is not a cause of the persistence of their delusional beliefs. For even if they were to reflect on the challenges, no change of belief would result; they are, by hypothesis, unable to shed the beliefs. Whether these people are self-deceived in retaining their delusional beliefs depends on the causes of their retention of them. If selective focusing is present here but is not a cause of belief retention, the observation that it is present does not warrant a diagnosis of self-deception.

Suppose that the pertinent cognitive deficit in some Capgras patients does not render them unable "to suspend their unreflective acceptance of veridicality" and instead makes it extremely difficult for them to do this. Then processes like selective focusing might do some work in sustaining the delusional belief. But the causal contribution may be so small that we may be disinclined to count the Capgras patient as self-deceived.

3. Conclusion

I agree with Audi that "we need to distinguish" self-deception from delusion (1982, 140), but my route to this belief differs from his. If Audi's view of self-deception is correct, then (R1) no one who is "in self-deception with respect to *p*" has a belief that *p* (1982, 147); and therefore (R2) no such person has a delusional belief that *p*. If R1 is false, one must look elsewhere for support for the thesis that "we need to distinguish" delusion from self-deception. On my own view, a true judgment about whether particular people do or do not enter self-deception in acquiring or continuing to have a delusional belief that *p* depends, among other things, on what roles motivation or emotion plays in the production or sustaining of these beliefs. Self-deception, as Audi understands it, "is not a *historical* concept. If I am self-deceived, so is my perfect replica at the very moment of his creation" (1997b, 104). I regard this as an interesting stipulation. As I see it, asking what self-deception would be like given the stipulation that the concept is not a historical one is analogous to asking what remembering that *p* would be like if "remembering" were not a historical concept. We can abstract away from the historical dimension of remembering that *p* and try to figure out what remains. On my own view, which I do not regard as stipulative,

"beings who have not deceived themselves are not self-deceived, or in a state of self-deception, no matter what else is true of them" (2001, 56). (I hasten to add that I argue in Mele 2001 and elsewhere that deceiving is not essentially intentional and that garden-variety self-deceiving is not intentional deceiving.) It is partly because I understand self-deception as an essentially historical phenomenon (like remembering that p) that I believe that whether any people with delusional beliefs that p are self-deceived about p depends on what processes produce these beliefs or cause these beliefs to persist. The relative merits of historical and ahistorical conceptions of self-deception are a topic for another day.[11]

Notes

1. Also see Audi 1985, 174–75 and 1997a, 144.

2. It is assumed here (and hereafter) that the substitution instances of both occurrences of "p" are represented in the same way. I forego discussion of Kripke's puzzle about belief (Kripke 1979).

3. For a brief review of some literature on this puzzle, see Mele 1987b, 4 and 8.

4. See Mele 2001, 26–27. Also see Mele 1983, 369–70; 1987a, 125–26. For some other routes to self-deception, see Mele 1987a, 149–51, 157–58.

5. If, in the way I described, Betty acquires or retains the false belief that her boss is not sexist, it is natural to count her as self-deceived. This is so even if, owing to her motivationally biased evidence gathering, the evidence that she actually *possesses* does not weigh more heavily in support of the proposition that her boss is sexist than against it.

6. This is a modified version of the test suggested in Mele 2003, 164. Discussion with Charles Hermes and Brian McLaughlin motivated the modifications.

7. I say "or something similar" because, for reasons that emerge in section 2, the test, as formulated, may not be reliable in unusual cases of certain kinds.

8. Audi also comments on delusion in Audi 1985, 171, 175; 1988, 96, 109; 1989, 252–53; and 1997a, 134, 144.

9. Recall my assertion in section 1 that "a person's passing the [impartial observer] test is evidence of bias suitable for self-deception." One moral of the paragraph to which this note is appended is that if a special cognitive deficit of the kind at issue is doing the causal or explanatory work, that fact undermines the evidence.

10. They also mention visual hallucinations and the Capgras syndrome in this sentence.

11. Randy Clarke read an abbreviated version of this paper at the conference, and Rich Reilly presented commentary. I am grateful to both of them for pitching in.

References

American Psychiatric Association. 1994. *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed. Washington, D.C.: American Psychiatric Association.

Audi, R. 1982. Self-Deception, Action, and Will. *Erkenntnis* 18: 133–58.

——. 1985. Self Deception and Rationality. In M. Martin, ed. *Self Deception and Self Understanding*, 169–94. Lawrence: University of Kansas Press.

——. 1988. Self Deception, Rationalization, and Reasons for Acting. In *Perspectives on Self-Deception*, ed. A. Rorty and B. McLaughlin, 92–120. Berkeley and Los Angeles: University of California Press.

——. 1989. Self-Deception and Practical Reasoning. *The Canadian Journal of Philosophy* 19: 246–66.

——. 1997a. Self-Deception, Rationalization, and the Ethics of Belief. In *Moral Knowledge and Ethical Character*, 131–56. New York: Oxford University Press.

——. 1997b. Self-Deception vs. Self-Caused Deception: A Comment on Professor Mele. *Behavioral and Brain Sciences* 20: 104.

Breitner, B., and D. Anderson. 1994. The Organic and Psychological Antecedents of Delusional Jealousy in Old Age. *International Journal of Geriatric Psychiatry* 9: 703–7.

Butler, P. 2000. Reverse Othello Syndrome Subsequent to Traumatic Brain Injury. *Psychiatry* 63: 85–92.

Davies, M., R. Langdon, and N. Breen. 2001. Monothematic Delusions: Towards a Two-Factor Account. *Philosophy, Psychiatry, and Psychology* 8: 133–58.

Enoch, D. 1991. Delusional Jealousy and Awareness of Reality. *British Journal of Psychiatry* (suppl. 14): 52–56.

Kipp, D. 1980. On Self-Deception. *Philosophical Quarterly* 30: 305–17.

Kripke, S. 1979. A Puzzle about Belief. In *Meaning and Use*, ed. A. Margalit. Dordrecht: Reidel.

Maher, B. 1999. Anomalous Experience in Everyday Life: Its Significance for Psychopathology. *The Monist* 82: 547–70.

Mele, A. 1982. Self-Deception, Action, and Will: Comments. *Erkenntnis* 18: 159–64.

——. 1983. Self-Deception. *Philosophical Quarterly* 33: 365–77.

——. 1987a. *Irrationality: An Essay on Akrasia, Self-Deception, and Self-Control*. New York: Oxford University Press.

——. 1987b. Recent Work on Self-Deception. *American Philosophical Quarterly* 24: 1–17.

——. 1999. Twisted Self-Deception. *Philosophical Psychology* 12: 117–37.

——. 2001. *Self-Deception Unmasked*. Princeton, N.J.: Princeton University Press.

——. 2003. Emotion and Desire in Self-Deception. In *Philosophy and the Emotions*, ed. A. Hatzimoysis, 163–79. Cambridge: Cambridge University Press.

Silva, J., M. Ferrari, G. Leong, and G. Penny. 1998. The Dangerousness of Persons with Delusional Jealousy. *Journal of the American Academy of Psychiatry and the Law* 26: 607–23.

Sorensen, R. 1985. Self-Deception and Scattered Events. *Mind* 94: 64–69.

Soyka, M. 1995. Prevalence of Delusional Jealousy in Schizophrenia. *Psychopathology* 28: 118–20.

Stone, T., and A. Young. 1997. Delusions and Brain Injury: The Philosophy and Psychology of Belief. *Mind and Language* 12: 327–64.

Tsai, S.-J., J.-P. Hwang, C.-H. Yang, and K.-M. Liu. 1997. Delusional Jealousy in Dementia. *Journal of Clinical Psychiatry* 58: 492-94.

Young, A. 1998. *Face and Mind*. Oxford: Oxford University Press.

——. 1999. Delusions. *The Monist* 82: 571–89.

Motivating Reasons for Action

RAIMO TUOMELA

I. Introduction

In this paper I will give a philosophical and conceptual account of motivational reasons that agents have for acting. My account covers not only the case of single-agent action and reasons for them (section II) but also the case of social reasons for action (sections IV and V). The latter topic has not been properly discussed before in the literature on reasons for action. Another aim of this paper is to relate my account to Robert Audi's theory of reasons (section III). My account is largely compatible with his, but there are still theoretical differences. For one thing, my view is more externalist and does not connect reasons to psychological states as closely as does Audi's account.[1]

To achieve a comprehensive account of reasons it is of course essential to discuss also social reasons for single-agent actions as well as joint actions. I will divide the social reasons into "private" (or "I-mode") reasons and group reasons (or "we-mode" reasons). Group reasons are institutional reasons broadly understood, as the agents here are taken to function and act as group members. Joint reasons for (we-mode) joint action represent a central case here, and I will concentrate on them. In section V of this paper I will show how to extend Audi's account to the case of social, especially joint, reasons.

Human agents are, at least largely, organisms trying to cope with their physical and social environment by means of their actions and other behaviors. The broader motivational basis of such activities is typically provided by the biological and cultural needs and desires that agents have. There are also desire-independent needs and reasons for action. Thus, for instance, promises and moral and social norms can give such reasons for action.[2] In addition, there are desires and interests that are not tightly connected to coping with the demands of external physical and social nature. Agents may be curious and try to find knowledge not only about their immediate environment but to find out things that seem to have no practical significance whatsoever (at least some philosophical thinking belongs here). Here we also have

reason-based action that, however, is not, strictly speaking, a response to the demands of the environment.

Agents can take facts (viz., obtaining states of affairs) in their environment to be reasons for performing certain kinds of actions, but they can also, alternatively, respond to potential facts, namely, states that the agent wants or intends to make actual.[3] In general, reasons for action are expressible by means of that-clauses: Wanting to stay dry, my reason for taking my umbrella with me was that it was raining, or, alternatively, that rain, according to my expectations, was forthcoming. External facts can influence an agent's intentional action only if he gets information about them "through his intentional channel" and comes to believe that the facts obtain or will obtain. In the present case, the agent must be taken to believe that it is raining or will rain. Even if the agent's belief state is involved in this "registering sense," the agent might in such cases be responding to his mental state rather than to its content, but typically this is not the case. In general, reasons are to be regarded as contents of wants and beliefs, but in some cases they can be want or belief states.[4]

In the case of false beliefs and beliefs that have no truth value yet (cf. the umbrella case) the agent is typically trying to respond to his (present or future) environment and not his belief state. Often there is something real in the environment that the agent is responding to, but even that need not be required, as we can treat the matter in analogy with what was just said about the want case. Thus, a reason is typically the content of a representational mental state rather than the mental state itself.[5] The agent has—intentionally, albeit perhaps subconsciously—taken this content as his reason in a sense to be discussed. The reason state need not objectively exist.

When the action is performed for the reason in question the reason-action relation is at least weakly normative but possibly only in a relative sense. Thus, that there is reason for me to keep the yogurt in the fridge depends on the view that unless yogurt is kept in a cool place it will spoil. There is an end (the yogurt will be unspoiled tomorrow morning) and there is the instrumental or technical normative connection based on the aforementioned natural necessity.[6] One has reason (here: one *ought*) to keep the yogurt in a cool place, "undetachably" relative to the mentioned end. In many cases one can also speak of a reason as *favoring* the action, relative to the end in question.[7] Of course, there are also categorical, nonrelative reasons (e.g., moral ones) that are fully normative. A trivial example: under normal conditions, promising gives a fully normative reason to keep the promise.

Objective facts are nonnormative things, but we predicate reasons to them: The sentence "State of affairs s is a reason for action A" then concerns a reason in a predicative sense of "is" (viz., "is a reason" is a predicate), not in the identity sense of "is." That something is a reason in this sense can be regarded as a human artifact, something depending on (explicit or implicit) human acceptance:[8] *we* accept or *I* accept that some fact qualifies as a reason for action.[9] This applies to natural reasons (like that there is a reason to warm up a cold hut to make it habitable), moral reasons (think of the promise case), and institutional reasons (a driver facing a red traffic signal has a reason to stop). If a reason in addition is a *motivational* reason, that is an

"up to the agent" type of thing. In the traffic light case, the agent therefore must believe that the red light is lit. A reason for an action—an "observer-relative" property or feature of a state of affairs in Searle's sense—is a normative (often only *instrumentally* normative) notion concerned with what one should (or should not, etc.) do.[10] Or at least the reason is a factor that favors or demands the action in question (or is normative in some other way).[11]

A *motivational* reason for action is one that the agent accepts as his reason for action and that he is at least prepared to act on. This kind of "intentional" (acceptance-based) reflexivity of the notion of reason is no problem as long as one can meaningfully use and "live with" the notion. As will be argued later, it does not require that the agent has more than a vague notion of reason, nor does it require that the agent reflexively takes a fact as his reason (cf. "this fact is my reason for opening the window"). Rather, the agent is just required to be disposed to function (perhaps unreflectively) in an appropriate way. (The notion of intention is quite analogous.)

II. The Individual Case

My main concern below is to give an account of an agent's motivational reason in a context where he is intentionally responding to the demands of a certain statelike part of his environment (thus physical or social, including institutional, environment) that is taken by him to obtain or that he purports to make obtain. I will, however, aim at a general account so that in principle, for instance, moral reasons are covered.

Consider an example. When the weather is hot (q), a person, x, actively comes to want and possibly to form the intention to get refreshment, and as result of his motivationally processing the external information he goes swimming (A). Action A, assumed to be intentional, requires a relevant intention to achieve something p (e.g., to get refreshment or perhaps to go swimming) conducive to the action. The agent may have deliberated about what to do in the situation at hand, and after having considered potential motivating reasons pro and con swimming he is assumed to have formed the intention to swim. The fact that the weather is hot (q) was the agent's motivating reason (or at least part of his reason) for his going swimming (A). For this to be the case the agent must have accepted q as his reason and connected it to his motivational system. Here he must be taken to have believed or at least assumed that q obtains.

Schematically speaking, we have here an external reason state q, the agent's belief that q, his want to get refreshment, his belief that he can best obtain refreshment in the present situation by going swimming, and his intention to go swimming. The obtaining of state q can in the present context thus be said to be the agent's reason (in the sense that the reason-predicate is applied by him to q) for his going swimming. There is accordingly a reason-preserving chain of thought motivationally leading to his going swimming.

Consider next a different case. The agent is in a stuffy room and wants to ventilate the room. Let q now stand for the state of the room being properly ventilated.

The agent here believes that he can ventilate the room by opening the window, and he accordingly forms the intention to open the window, and this leads to his opening the window. Here the wanted (obtaining of) state q was the agent's reason for his action of opening the window.

I accordingly propose the following schematic account of motivational reasons, namely, reasons *for which* the agent acted:[12]

> (*RA*) State q was the *reason for which agent x performed action* A in situation S if and only if in S
> (1) x had some want(s), belief(s), and (prior) intention(s), relevant to his performing action A (relevance partially spelled out below), and some of these propositional attitudes were about q—typically the agent must have either wanted q or believed that q (*existence of motivational set*);
> (2) x acted and performed A on the motivational set, namely, on the wants and beliefs, and intentions in (1) (*acting on motivational set*).

In (*RA*), (1) is a rather obvious and common requirement for a motivating reason, and it does not require deeper discussion here. In (2), the word *on* expresses that the agent purposively acts on his action plan (in general including his intention to perform A), striving to satisfy it. This involves the agent's guiding and monitoring his action in the right way in accordance with his plan of action (cf. the problem of wayward causal chains), all this making it true that he at least functionally has taken q as his reason. This means that "acting on" here is understood in a special, strong sense entailing that the agent will take q as his reason for action (see below for how weakly this can be understood). This may (but need not) involve that he rationally reasoned from q to his action. (*RA*) does not require that the reason state q be the total reason for which the agent acted but allows that he has picked out a specific element (viz., q), which he regards as his motivationally salient reason. Such a reason can be regarded as an insufficient but necessary constituent of an unnecessary but a sufficient condition for the agent.[13]

In (*RA*), wants and intentions are supposed to have the world-to-mind direction of fit of semantic satisfaction while beliefs in general have the opposite, namely, mind-to-world direction of fit. The agent's motivational set may include not only the elements in (1) but further mental states, for example, emotions. However, it can be argued that only the cognitive components of emotions can function as reasons and that thus the other elements in the motivational set take care of the "reason effects" of emotions.

As pointed out, my account allows that the reason state q is about the agent's own mental state. The agent may thus respond to his belief state or to his want state in some cases. To find examples, one has to deal with such special cases as "Agent x believes that Martians are after him to kill him" or "X wants to drink a can of paint." In the belief case x may respond to his haunting want—that he regards as a crazy one—and seek help from a psychoanalyst, or he may respond to the content of the belief and seek help from the police.[14] In the case of the special want, the agent may either respond to the mental state of wanting and seek help to get rid of this mental state or he may respond to the content of the want and go to

a hardware store to buy a can of paint. It can thus be seen that different actions are performed depending on whether the mental state or its content is responded to. There is a clear functional difference at stake here.[15]

In general, a motivating reason serves to explain an action and justifies the action for the agent—independently of whether the reason state objectively exists. Hot weather can be cited as an explanation of x's action of having a swim. The explanation is of course only a partial one. A fuller explanation states that the agent took hot weather to be his reason for seeking refreshment; so he formed the intention to get refreshment; and as swimming was believed by him to be the most appropriate way to get refreshment, he went swimming.

As both beliefs and wants (viz., proattitudes) can provide reasons and serve as their psychological basis, we need to consider both possibilities, which are not fully symmetric (as at least in general wants have a primary role in that they "tell one what to do"). The idea is to think that the agent can on different occasions take either a belief-based reason or a want-based reason as his (main) reason. This is his motivationally salient reason, which in general may fail to be his total reason for action. For example, an agent may have a *standing* want that something p were the case (e.g., he wants to eat cashew nuts). When he acquires the *occurrent* belief that action A will secure p (the belief can have the content that there are cashew nuts for sale in a certain shop, action A being the relevant buying action), he acts on this belief and takes the belief content as his reason for buying nuts. From an objective third-person point of view, he is actually but perhaps only unreflectively and possibly even unconsciously acting on the whole conglomeration of states in his motivational set, especially on his want and the belief. The reason (viz., that a certain shop is selling cashew nuts) here is generated by his belief. But in other cases the agent may only need to have the "registration" belief that the state obtains (e.g., that the red light is on in the car stopping case, which presupposes that a normal agent also believes that he is required to stop his car at red lights in order to fulfill the normative requirement).

Conversely, an agent may have a *standing* belief (e.g., one with the content that there are cashew nuts always available in a certain shop) but that he only on a certain occasion acquires the *occurrent* want to achieve p (e.g., to eat cashew nuts). Here the mental state in focus is the want state and its content (a want-based reason). Of course, wants and beliefs being interdependent and wants generally primary, he still at least tacitly acts on the whole mental state conglomeration in his motivational set. In some other cases, all the states might be equally salient and he might or would then act on this total reason, and then no separate treatment of belief-based and want-based reason is warranted.[16]

Consider still the example of a want-based reason where the room is stuffy and the agent wants to ventilate it. In this case the reason content (q) would be expressed by "x is in a ventilated room" (or "the room is ventilated"). The action A would be opening the window. While in the hot weather case the agent's *belief* that the weather is hot would be a salient motivational element in the agent's motivational, reason-expressing set, in the ventilation case the agent's *want*, in contrast, is the central motivational element in the total psychological reason. In the first case the

agent will accept the external state q as status quo while in the second case q is to be brought about. In the want-based case, in contrast to the belief case, the direction of fit is world to mind and the reason state, say q, so far does not exist. This also means that it can only be a *type* of state in contrast to the belief case where q is a *token* state.[17]

Understanding beliefs here broadly to be attitudes with the mind-to-world direction of fit and wants as having the opposition direction of fit, the want-based and the belief-based cases in view of our above discussion exhaust the reasons there can be for single-agent actions.

As to the notion of taking a state as one's reason for action, it involves giving this state a special status that the agent "respects" by his action in terms of which he responds to the reason. In a standard case, the agent is disposed to refer to the reason as his reason for performing the action and is disposed to perform relevant practical inferences on its basis vis-à-vis his action as well as to take the reason to explain his action. According to (*RA*), acting for a reason involves that the agent connects the (typically external) reason state to his psychology in the functionally right way. Thus, the agent must have a proattitude to act in that right way relative to the reason state in question. When the state is the content of a mental state, for example, when he, by his wants or beliefs, generates the state, he will have to act in the right way related to the direction of fit of the wants (or beliefs) in question. He monitors the situation and guides his action in view of the reason state in relation to his motivational set. Sometimes this requires him to perform certain practical inferences related to his plan of action that has as its output his intentional performance of a relevant action (here A). A minimal amount of rationality is required here concerning consistency in thinking and acting. However, the falsity of the connecting belief in question is allowed by my account, thus nothing more than subjective rationality needs to be involved.

However, an agent acting for a reason need not reflect on his reason and need not even have the concept of reason in an articulated cognitive sense (think of a small child acting for the reason of getting candy). His plan of action (however rudimentary it may be) contains as its central element the reason state that we have above dealt with. Taking something as one's reason need not cognitively involve more than what intentionally responding to the state in question requires—the right functional process basically suffices (cf. my unreflectively moving my body in the right way in a narrow corridor for the reason of not bumping into somebody else—all this can be highly routine). However, in general, the agent must be disposed to articulate his reason and to reason adequately on the basis of it. The disposition here might not be directly realizable but require learning. This is pertinent in the case of, for example, three- or four-year-old children, as they seem capable of acting for a reason.[18]

As seen, my account of single-agent reasons gives a unified way of speaking about all kinds of reasons, be they the agent's motivational reasons or not, and it allows for reasons that are created by the agent and would not exist without his relevant mental states (primarily wants, intentions, and beliefs). On the other hand, as emphasized, there are also "desire-independent" reasons and "agent-externally" created reasons.

III. Audi on Reasons for Action

Robert Audi has discussed reasons for action in several works, and especially in moral contexts. In 1986 he published a paper on acting for reasons, which I regard as a very fine and fundamental paper on the topic. There he makes several important distinctions, and he proposes and defends a number of central theses about reasons. His account regards reasons as contents of wants and intentions in the first place (2001a), while contents of beliefs are reasons in a more secondary sense. Wants are primary in the sense that they provide motivation for action (and I agree with this). Reasons are basically expressible by infinitival clauses to the effect that my reason for taking my umbrella with me is *to stay dry*. He also says that psychological states "provide" reasons. That would seem to amount to saying that the mental states either constitute or at least generate the reasons, as the latter occur in the contents of the former.[19]

While Audi's newer work on reasons is in some ways richer than the early work, I still find the basic 1986 paper more precise and will below concentrate on it. It seems that Audi has not in his recent work (such as the 2001 book and his 2004 paper) substantially changed his view on what reasons basically are.

To begin, in his 1986 paper he discusses the following example about an agent, Sue. In the course of mailing impersonal invitations to a conference that she is organizing, she puts John's aside. Her reason here according to Audi is (*) *to delay it until after she sends him a condolence letter (his mother has died)*.

The immediate reason here is the delaying while the further reason is that John's mother has died. Concentrating on the immediate reason, this example can be formulated in terms of a state reason and a that-clause as (**) *that she will delay sending the conference invitation until after sending a condolence letter.*

There is also the difference between Audi's and my views that I take reasons for which an agent acts to involve the agent's *taking* the state in question as her reason. Here Sue's taking (**) as her (want-based) reason amounts to her intentionally putting the letter aside because of (**), which expresses a state that Sue intends to make and keep satisfied. At least she must be acting functionally in the right way as if she had taken the state to be her reason (recall my earlier discussion). This is the case even if she may not have used the term "reason" or any of its cognates and might not have thought about her action as a reason-based action. Every such reason is to some extent normative, roughly in the sense of favoring or perhaps demanding the action in question or "reasoning" the action.[20] As is commonly accepted, if an agent performs an action for a reason, his action, so described, therefore is necessarily intentional.

Audi requires of typical cases of acting for a reason that (1) the action (such as Sue's action above) must be explainable by the reason in question (viz., [*]), that there is the right kind of (2) connecting belief (Sue believes that her putting the letter aside will delay it), and (3) a disposition to appeal to the reason in question when asking what she did and why, (4) the action is a response to the agent's reason, and she has some (5) degree of awareness of a reason, (6) she knows that and why she is putting the letter aside, (7) her action is rational relative to her reason and her connecting belief, and (8) she has control over whether to perform the action in ques-

tion or not.[21] (These requirements are later in the paper built into generally applicable conditions that are named as follows: the explanatory condition, the connecting belief condition, the attribution condition, the nonaccidentality condition, and the normal intermediaries condition.)

Audi ends up with the following account of reasons that is meant to apply to basic actions (as opposed to nonbasic ones, such as actions generated in terms of the by-relation):[22]

(I) x's A-ing is an *action for a reason*, r, at t, if and only if, at t, x performs A, and there is a connecting relation, C, such that

(1) x wants to r and believes C to hold between her A-ing and r, or believes something to the effect that C holds between her A-ing and r;

(2) x's A-ing is at least in part explained by this motivating want and at least one connecting belief, and is guided by the belief(s);

(3) x is noninferentially disposed, independently of seeking reasons she has had, or might have had, at or before t, for A-ing, to attribute her A-ing to the want and (explaining) belief(s);

(4) x's A-ing is nonaccidentally produced by the want and (explaining) belief(s); and

(5) the want and explaining belief(s) do not bring about (or sustain) x's A-ing via an alien intermediary.

In Audi's above formulation r is the content of a want, and thus we are dealing with the *want-based* case of motivating reason here.

I take both Audi's above conditions of adequacy and his above account to be basically compatible with my account of reasons. There are, however, some differences that I will list here and discuss afterward. Thus, I have added the (i) *taking a state as one's reason* condition. Audi regards this condition as too strong but does require internalistic "accessibility" to the reasons of action.[23] This latter requirement is weaker. Audi does not say what his argument against (i) is. Recall that my wide sense of taking a state as one's reason does not really require more than acting in the right way on the motivational set in question. But this seems to amount more or less to Audi's above alien intermediaries condition that prohibits, or is meant to prohibit, the existence of wayward causal chains. Thus, concerning (i) my approach after all may be compatible with Audi's.

I have also added (ii) *belief-related* reason-cases, where a belief plays a salient motivational role. For Audi, belief contents can also be reasons, although not in a primary sense. In his 2001 book (see, e.g., 121) he does accept that also beliefs may play the role of reason (although subordinated to wants), so perhaps my approach also concurs with his recent views if not his 1986 paper on this score.

In my account all motivational reasons are taken to have some (iii) *normative force*. This seems to contrast with Audi's view, as he in his classifications of reasons distinguishes between normative and other kinds (e.g., motivational) of reasons, unless he is willing to accept that all those other kinds of reasons are normative as well.[24] When I require reasons to be normative, I also include mere "technical" or instrumental normativity, while Audi might have in mind only substantive (e.g., rational or moral) normativity in this context. As to technical normativity, I may want to open the window and have formed the intention to do that. The (so-far

nonexistent) state that the window is open conceptually and "hermeneutically" requires in this merely technically normative sense that I act appropriately in relation to the reason state to be tokened. Audi would hardly object to the requirement, basically involved here, that if you intend that a state obtain you have to act appropriately.[25] The result then seems that Audi would have to accept also the present kind of technical normativity to count as a species of normativity, and there seems to be a dilemma for Audi here—or at least the reader would expect a clearer statement about the normativity of reasons.

A related point is that Audi says on page 121 of his 2001 book that (iv) *an irrational desire does not provide the agent with any kind of normative reason, though it can explain action*. I do not agree with this. Suppose an agent desires to ski to the North Pole and acts on this desire content. To me this is an example of an irrational desire (although not to all of us). Another case would be provided by the desire to drink a can of paint. As soon as the agent has formed the intention and committed himself to realize the content of the desire, there is technical normativity included. So here my view seems to differ from Audi's if he is committed to regarding technical normativity as a species of normativity.

IV. I-Mode Social Reasons for Action

In this section I will consider "I-mode" social reasons for single-agent and collective action. Briefly, an I-mode reason is a person's private reason for action, and in this section the reason will be taken to have social content. The central case of this is where the others' thinking and acting thus and so is a person's reason for thinking and acting so. I-mode reasons contrast with we-mode reasons. A we-mode reason is a social group-reason for action. Roughly, a person acts for a we-mode reason if he takes it as his main reason for action that his group wants, intends, or believes something that also requires his participation as a group member.[26] I claim that I-mode reasons are basically covered by the earlier analysis (*RA*) but that we-mode reasons, which obviously are not reducible to I-mode reasons, require a somewhat more complex analysis.

As to the centrality of social reasons, the basic point here is that human beings are fundamentally social beings disposed to cooperate and share intentional states. This probably genetically based sociality aspect is a many-faceted thing.[27] Thus, people on the whole need, and also enjoy, the company of other human beings, and they in their thinking and acting tend to take into account what others think and do, especially what is generally expected of the group members. Accordingly, others' approval and disapproval of one's ways of thinking and acting are also important motivational elements. All this induces an element of conformity and cooperativity into human life. Obviously there are other, especially private reasons and motives for action, and they sometimes supersede the cooperative ones.[28]

My explication of social reasons in the above sense of responsiveness to other people's thoughts and actions (be they singular actions or parts of collective activities) involves taking the social reasons in question to be shared "we-attitudes" (to be defined shortly).[29] Here the "we" need not be a "togetherness-we" but may be only a "thin" one. I have argued elsewhere that a central class of intentional collective

social actions and social practices, including institutional ones, is formed out of actions performed because of a shared we-attitude (such as a socially shared goal or belief), either in the I-mode or in the we-mode.[30] Given this, my present account (in this and the next section) will cover perhaps the most central social cases.[31]

We can speak of we-attitudes in two different ground-level senses. There are plain we-attitudes and there are reason-based ones. An agent has a plain we-attitude ATT (e.g., want) toward a content s, expressing, for example, action, precisely when he has the attitude ATT(s), believes that the others have the attitude ATT(s) and, furthermore, believes that it is mutually believed in the group that the group members have the attitude ATT(s). A reason-based we-attitude in contrast is characterized as follows: an agent has a reason-based we-attitude toward s precisely when he has the attitude ATT(s) and has it in part because, namely, for the reason that (he believes that) the others have the attitude ATT(s) and (believes that) it is mutually believed in the group that the group members have and, perhaps, ought to have, the attitude ATT(s). The activity that we here have a reason for in the first place is the agent's forming his we-attitude, which in turn "reasons" the overt action (s).

My interpretation of "because" (and "on the basis of") as reason involving in the case of a reason-based attitude seems warranted because the agent is assumed to be intentionally responding to the "becausing" factor. Thus more than mere naturalistic causation must be involved. The intentional formation of a we-attitude is an action, and thus the agent's reason for forming his conformative attitude is of the kind already discussed and accounted for in section II.[32]

Consider the example where some agents plan to have a swim, where they are not required to swim together but can do it alone. Each person decides on her own whether to join the action, but each person is also here assumed to be conformistically disposed to act on a shared we-attitude (plan). Here the attitude can be simply the goal of having a swim.[33] Considering an agent's full reason content, her reason for going swimming is not only that it is her intended goal but rather that the others will go swimming and that there is a mutual belief in the group about this. When all the agents in the group go swimming on the basis of the we-goal to go swimming, we get a kind of collective social action guided by a shared we-goal.

Suppose agent x then instantiates s and intentionally has a swim. What was x's reason here in precise detail? It consisted of the fact that the others went swimming and that there was mutual belief in the group about this. The agent is conformatively responding to his present social environment, and this gives him the mentioned reason.[34] It must be acknowledged that agents often act because of a multiplicity of reasons and the content of those reasons may be complex. Mixtures of we-mode and I-mode reasons seem to be common in actual life.

To make our example more interesting, let us take our agents' goal to be getting refreshment tonight. Then assume that the agents share the belief that having a swim tonight (s) will be required for their getting the wanted refreshment (r). We need not discuss the various kinds of practical inferences that might be involved but stipulate that swimming in this case is necessary for getting the refreshment. The reason r here is the collective state of the agents' getting refreshment such that this state-to-be-realized was the agents' shared we-goal. Here r is differently tokened by each x (in contrast to the proper joint action case, where the token must be a joint

token). They realize r by going swimming, and each participant's action can be explained by the we-attitude in question. In this explanation it is assumed that each agent x believed that the others have the same goal with content r and that this is mutually believed. We may assume of reasons here that the participants in general would not have had their conformative reason unless the others also act on the basis of the same-shared we-attitude. (However, this statement only holds "in general," but it does not hold for *all* group members as there obviously must be some "initiator," some members' attitudes or actions available for imitation.)

As the collective action in question consists of each group member's going swimming, the explanation of this collective action consists simply of the aggregate of the explanations of the above kind. In this case the explanation has the same form for all agents. It refers to the shared goal and to the relevant shared beliefs of the agents. There are of course more complex and more interesting I-mode cases where the participants can have different goals but still be dependent in a social sense. For instance, an agent may have a certain attitude in part because the others or some others have a certain different attitude.[35] In institutional cases such as organized groups, this is typically the case.[36]

To end this section, it can be noted that Audi's account (I) of section III can be extended to the present kind of I-mode social case simply by requiring the reason r to be a conformative social reason as follows. Instead of the previous analysandum "x's A-ing is an *action for a reason*, r, at t" we now use "x's A-ing is an *action for a conformative social reason*, r, at t" and require in addition to the earlier clauses (1)–(5) that the following also holds true: (6) x's want that r is a we-want that r, namely, x wants that r in part because the others in the group want r, under conditions of mutual belief about this.

V. We-Mode Social Reasons for Joint Action

There can be we-mode reasons, namely, full-blown group reasons, both for individual and joint action. I will below discuss joint action in its fullest sense, which is joint action *as a group* (or as one agent). A we-mode reason is a group reason an agent has when acting as a group member in the full sense.[37] The group may have formed an intention expressible by "we will perform action A together" or a group belief expressible by "we believe that p," and so on. Thus, when acting as a group member an agent will have to perform his part of A in part for the reason that he as a group member accepts the group's intention and analogously he will have to respect the group belief and take it as his group-based reason when thinking and acting as a group member in the full sense. The group here can be an "ephemeral" one concerned only with the joint performance of some task. (Note that whenever there is a joint intention there is a social group, a group capable of action.) As group reasons clearly are not analyzable in terms of private reasons, the analysis of we-mode reasons for action goes beyond the account (*RA*), which was above argued to cover I-mode reasons and only them (if nothing else is added).

To consider joint action, we first notice that, on conceptual grounds, there is necessarily a social reason involved in any intentional joint action. This social reason,

indeed a joint reason, is, or at least contains, the joint intention required for making the action intentional. Consider the example of some agents forming the joint intention to paint a house together. Then the joint intention will serve as their reason for their painting the house together and it also serves as each participant's reason for his performing his part of the joint action. Notice that the joint case is different from the singular case, where an intention cannot (at least in many cases) be a reason for the action that the intention is about. This is because one can rather arbitrarily form an intention and act on it. Such bootstrapping is not, or need not be, involved in the case of a joint intention, because of the "interaction effect" in joint intention forming—joint intentions often are not aggregates of the "proposed" we-mode participant intentions (viz., "we-intentions) and are not aggregates of their private, I-mode intentions.[38] If the participants have an underlying joint want, then that want (or its content, rather) can serve as a deeper joint reason than does the joint intention formed on its basis. But think of cases involving negotiations and bargaining between the participants that finally lead to a joint intention that is a compromise and is against some participants' underlying wants (at least first choices). Here the jointness of the intention indeed is informative, and the we-mode proattitude that it involves can serve as a motivational reason.

While forming a joint intention expressible, for example, by "We will do A together" thus can generate a reason for the participant' joint performance of A, this perhaps is not what we typically are interested in when discussing reasons of for joint action. As compromise cases show, there need not be a joint proattitude underlying the joint intention. But if there is such a joint attitude, especially a stable underlying attitude (either an I-mode or a we-mode one), it may give more unity to joint action than mere joint intention does, especially in the case of repeated joint action.

In contrast to the cases (RA) covers, here we are explicitly dealing with several singular agents sharing a joint reason (minimally a joint intention whose content will constitute the joint reason), and more needs to be said. The only thing we actually have to do in order to apply the single-agent account to the joint case is to write down the conditions for a group acting for a reason in terms of the group members' relevant attitudes and actions. We get this account in the case of a standard kind of joint action, A, to parallel (RA):

> (JRA) State q was the *joint reason for which the agents* $x_1, \ldots x_m$ *performed action A jointly* in situation S if and only if in S
>
> (1) $x_1, \ldots x_m$ had some (possibly joint) want(s) and belief(s), as well as, especially, a joint intention, relevant to their performing action A jointly (relevance partially spelled out below), and some of these joint propositional attitudes were about q—typically the agents must have either (jointly) wanted q, (jointly) intended, or (jointly) believed that q (*joint motivational set condition*);
>
> (2) $x_1, \ldots x_m$ acted and jointly performed A on the joint motivational set, namely, on the joint intention and the possibly shared attitudes (wants, and beliefs) in (1) (*acting on joint motivational set condition*).

Corresponding to (JRA) we may now have a case involving an external reason, q (e.g., the weather is hot, or the house needs to be painted), and assume, in analogy with the single-agent case, that it generates a motivational process in the agents

that will make the reason a motivational one. As a result the agents arrive at a suitable joint intention (e.g., to get refreshment or simply to go swimming together rather than separately or to paint the house together). I will assume in my formulation that the intention here is a full-blown (or "we-mode") joint intention. Similarly, action A is assumed to be a full-blown joint action. Given these two assumptions, we are able to deal with the most interesting case of acting jointly as a group.

As in the single-agent case, we must also here distinguish between the belief-based and the want-based (or proattitude-based) case—and more generally, between a reason primarily motivationally related to an attitude with mind-to-world direction of fit and one related to an attitude with the world-to-mind direction of fit. (The world-to-mind case will include joint intention.) In the former case the mutual belief that q will be the central element in the (partially) joint motivational set of $x_1, \ldots x_m$, while in the latter case q is the jointly wanted state or action.[39]

Because acting on a joint motivational set requires joint intention, on conceptual grounds, in the case of full-blown joint action, each agent must then have—as his "slice" of the joint intention—a "we-intention" toward the participant's joint performance of the joint action in question, and he must also have the intention to perform his part of the joint action A. Each agent then performs his part of the joint action A partly for the reason that the agents will perform A jointly in accordance with their joint intention. Their reason for participation is accountable in terms of (*RA*) if it is assumed that (a) they are *acting and functioning as group members* (thus having the attitudes in the motivational set and their acting on them as group members) rather than as private persons and if (b) (*RA*) is allowed to deal with joint attitudes.

Given our broad understanding of the want-based and belief-based shared reasons, respectively, as reasons either with the world-to-mind direction of fit of satisfaction or the mind-to-world direction of fit, they can be taken to exhaust all (proximate) reasons for joint action.[40]

I-mode reasons and we-mode reasons may lead to dramatically different results in the case of strategic social action, for example, in collective dilemma situations. To illustrate, consider a simple two-person two-choice case of a prisoner's dilemma with the familiar choice alternatives C and D, where the row player's preference ranking is DC, CC, DD, CD, and the column player's symmetric ranking is CD, CC, DD, DC. Here the agent may choose either C or D when he acts for an I-mode reason. Thus, considering the single-shot case, if the agent—the row player, here "I"—thinks strategically and intends to maximize his value or utility he can reason thus: I prefer the joint outcome DC to all the other joint outcomes; however, I realize that if you are reasoning similarly and planning to go for CD, we will end up in DD. For wanting to avoid the worst outcome CD, I cannot rationally aim at the Pareto optimal CC. So I choose D and rationally expect DD to result in the single-shot case. In this case the agent's reason for choosing D is to secure at least the third-best alternative given his beliefs of the nature of the game and the other player's rationality. Thus, as long as change of the game structure is not allowed, in a standard single-shot PD mutual defection is the rational outcome, and this result holds independently of how egoistic or altruistic the participants' preferences and utilities are.[41]

In contrast, acting for a we-mode reason can lead to the group members' rationally cooperating in collective action dilemmas. Considering what individualistically viewed is a prisoner's dilemma situation from the point of view of the group members' private preferences, we now adopt the group agent's point of view and assume that the group ("our group," "we") accepts the dominance principle ("higher payoff dominates lower payoff") and thus intends to choose C (over D). Thus, as the group can be taken to intend and act only if the members correspondingly jointly intend, as a group, to cooperate and accordingly jointly cooperate, it follows that the participants form the joint intention to realize the joint cooperation outcome and accordingly perform their parts of the group's achieving it, namely, do C. As the group members are assumed to act as one agent, the joint outcomes CD and DC simply drop out. The prisoner's dilemma in question simply does not arise for the group agent. The agents act collectively rationally when they act for a we-reason as here, and indeed there is no room for private individual action and private rationality at all. The upshot then is that while strategically acting agents acting on the basis of the relevant I-mode reason involving maximization of private value (utility), they can rationally choose only D in the single-shot case, but in the case of agents acting for the relevant we-mode reasons (e.g., "our group will maximize the value for the group by acting appropriately") the agents will rationally choose C. Of course, it must be added to all this that in actual practice groups may not act fully as units. There may be freeriders functioning on the basis of their I-mode utilities and this will normally require group-based sanctioning. Thus, from the backdoor, as it were, we may get the prisoner's dilemma alternatives back.[42]

Let me return to Audi's theory of single-agent reasons and extend it to the social case. Given my above discussion, his 1986 account of reasons can easily be extended to the social jointness case. Thus we assume or think that the agent x in Audi's account I (see section III) is a collective agent (viz., a social group). This is the right move to make precisely because joint reasons for action concern joint action as a group (viz., we-mode joint action). Using holistic language, the group has a reason for its action, and this is what we are trying to understand in less holistic terms. As a group can only act if (some of) its members act, we will speak of its members', or its relevant authorized members', actions.[43] A group carries a table upstairs when its members (or its authorized members) jointly carry it. A group builds a house when its members perform certain part actions purported to be functional to the house getting built, and so on. Along similar lines we can also speak of a group wanting, intending, having a reason, believing, and so on.[44] Audi's account can now be extended in terms of the following analysis of acting jointly for a want-based reason (the belief-based case can be rendered analogously):

(I_s) A joint A-ing by $x_1, \ldots, x_m$ is a *joint action for a joint reason*, r, at t, if and only if, at t, $x_1, \ldots, x_m$ jointly perform A, and there is a connecting relation, C, such that

(1) $x_1, \ldots, x_m$ jointly want to r and mutually believe C to hold between their jointly A-ing and r, or mutually believe something to the effect that C holds between their jointly A-ing and r;

(2) the joint A-ing by $x_1, \ldots, x_m$ is at least in part explained by this motivating joint want and at least one connecting mutual belief, and is guided by it;

(3) $x_1, \ldots, x_m$ are noninferentially disposed, independently of seeking reasons they have had, or might have had, at or before t, for joint A-ing, to attribute their A-ing to the joint want and (explaining) mutual belief(s);

(4) the joint A-ing by $x_1, \ldots, x_m$ is nonaccidentally produced by the joint want and (explaining) mutual belief; and

(5) the joint want and explaining mutual belief(s) do not bring about (or sustain) these agents joint A-ing via an alien intermediary.

Here, r is the content of a we-mode joint want, thus we are dealing with the want-based we-mode social case of motivating reason. Assuming that the we-mode joint action indeed is intentional and entails a relevant joint intention (in the we-mode), (I_s) is compatible with my account (*JRA*); and, given how I have clarified (*JRA*) it entails (I_s). Whether Audi is willing to think that the joint action here indeed is a we-mode joint action based on a group reason I will leave open to interpretation here. (Furthermore, the reader should recall the discussion in section III—it applies mutatis mutandis to the present case.)

VI. Conclusion

In this paper reasons for intentional action have been discussed and informative truth conditions for statements attributing reasons have been given both in the individual and the social, many-person case. All motivational reasons in the case of intentional action must, nevertheless, be intentionally "processed" by the agent via his want-belief system and, indeed, accepted by him to be his reasons for action. The agent must form beliefs about facts that will be his reasons and will have to have a proattitude toward the potential facts that he purports to make actual. The engagement of the agent is needed when acting for a reason already before acting: it is up to him what his effective reasons are at least in the sense that in principle he could have chosen to act on other reasons than those he did.

    The approach of this paper has been to explore Robert Audi's theory of reasons. Although the two approaches are largely compatible, Audi's approach only concerns the single-agent case. This paper shows how it can be extended to cover at least the central case of joint reasons for joint action. Some remaining problems that still require analysis have been noted, among them the question of the normativity of reasons and the content of reason-clauses, but they must await another occasion.*

Notes

    1. My main sources for Audi on reasons have been his 1986 paper and his 2001a book.
    2. See e.g., Searle 2001 for these.
    3. When the agent takes his want content (what he wants) to be his reason for action, he is responding to the possible future state of affairs in which, according to his belief, the want is satisfied. One can construe such possible facts in different ways, e.g., realistically by assuming that such possibilia really exist or by reference to either so far merely conceptually

existing facts or to "intentionally inexisting" facts in a Brentanoan sense. I will leave such ontological questions open in this paper, but let me indicate a technical approach that can be used here: the possible worlds analysis of propositional attitudes. Consider the case of belief: x believes that p = in all the possible worlds compatible with what x believes, it is the case that p. More precisely, "x believes that p" is true in a possible world w if and only if p is true in all the worlds accessible (in the belief-accessibility sense) from w. This set of possible worlds (states of affairs complete with respect to the framework in question) in which p holds true then can be taken as the semantical *analysans* of the agent's reason for action, and the reason is real if the actual world belongs to the set. In this paper I speak of *states* and *attitude contents* almost interchangeably as reasons, and the present kind of possible worlds account is one way of making this possible.

4. In this paper, I take wants and beliefs, respectively, to be representatives of states with the world-to-mind direction of fit and those with the mind-to-world direction of fit, but for ease of exposition I just speak of wants and beliefs.

5. Stoutland (2001, 2005) also seems to adopt this view, although I do not know what his account says about want contents as reasons, and although he seems to require that in the false belief case the agent must still be responding to something real in his environment.

6. Audi 1986 speaks of a connecting belief in this kind of situation.

7. See Dancy 2000 for the notion of favoring.

8. I take this claim also to encompass moral reasons, although this may not be a very popular view.

9. See Tuomela 2002 for a collective acceptance account of collective social constructions and a discussion of things that are in the meant sense up to us to collectively create and maintain.

10. See Searle 1995 for the notion of observer-relativity.

11. This paper concentrates on "*pro tanto*" reasons as contrasted with "normative requirements" in Broome's (2000) sense. *Pro tanto* reasons can be weighed against each other in contrast to normative requirements; see section IV.

12. Compare Audi 1986, Searle 2001, and Miller 2005 for resembling accounts.

13. At least in some cases of motivating reasons, the reason, required to be a state the agent takes to be her reason, might be regarded from the agent's point of view as an *inus* condition of the action, say X, it "reasons" (perhaps relative to suitable background conditions). An inus condition, as explicated by Mackie 1974, 62, is an *insufficient* but *nonredundant* part of an *unnecessary* but sufficient condition of X. To illustrate this, assume that there is in the circumstances in question a necessary and sufficient condition for action X and assume that this condition be expressible by a disjunction of conjunctions of some factors of the kind "(ABC or DEF or GHI)" where A, B, C, . . . are conditions for action. Perhaps they are only subjective ones, namely, factors the agent regards as relevant for his undertaking the action in question. Here each conjunct, such as ABC, is sufficient for the action, and it is also assumed to be minimally sufficient, namely, none of its conjuncts A, B, C is redundant and no part of such as AB is sufficient for X. Each single factor, such as A, is neither a necessary nor a sufficient condition for X. However each such single factor, a reason in the present account, is an inus condition of X. (Compare the discussion and comments on inus conditions in Tuomela 1976.)

In the case of a *total* reason q it may be advisable to require that the agent would not have done A unless q had obtained (as Miller 2005, does). In the total reason case the whole

motivational set of the agent is referred to. In the case of want-based reasons it is wants rather than beliefs that will be regarded as motivationally central (salient), and analogously for the belief-based case. Recall, however, that wants and beliefs are interdependent (as is generally accepted). Acting for a reason in general requires both wants and beliefs.

14. Dancy 2000, chapter 6, discusses similar belief cases.

15. In my approach, the reason for which the agent acts need not actually exist at all—consider the rain expected to happen today or an agent's want to get a ticket for the next day's concert. Whether or not the reason state will actually come into existence is immaterial from the point of view of the agent's taking a content to be his reason and acting on it. The same goes for false beliefs (I turned the knob believing it will open the door when it actually closed the window) and even for mental states such as believings as reasons (one might be in error about them to some extent, as well). The relevant knowledge on the part of the agent would of course solve the "real reason" problem, but such knowledge may not be available and in the case of future-directed attitudes cannot even be available at the time of action.

16. Here is a sufficient condition for a belief-based reason (reason$_b$):

(RA_b*) q was a *reason$_b$* for which x performed A in situation S if x believed that q and there was a state or event r such that q generated x's want that r, and x believed that he can achieve r by and only by his performing action A, and x performed A on this want and this belief.

In the present analysis, content q is a proposition assumed to express a potential or actual fact.

17. Here is a simple sufficient condition for a want-based reason (reason$_w$) in analogy with (RA_b):

(RA_w*) q was a *reason$_w$* for which x performed A in S if x wanted q and x believed that he can achieve q by and only by his performing action A, and x performed A on this want and this belief.
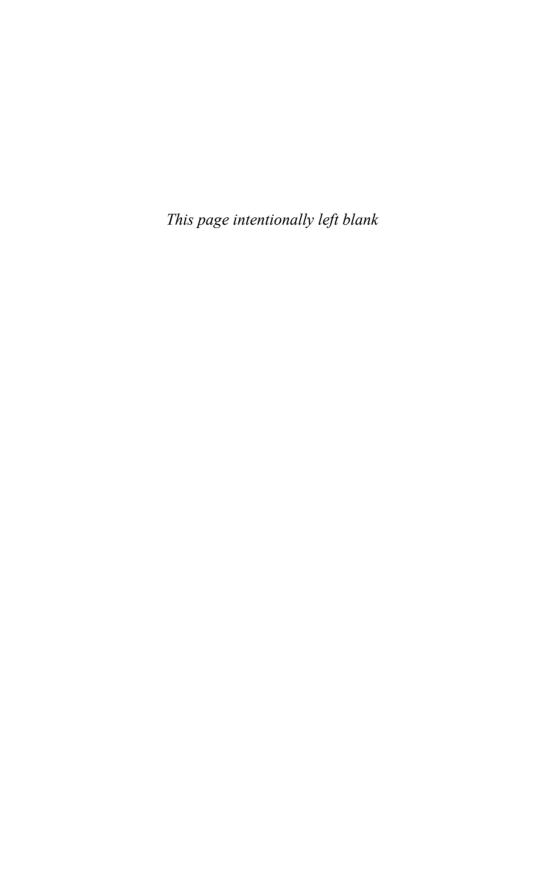
18. Routinization of relevant thinking and action also in the context of acting for a reason is a central psychological phenomenon. Whatever the degree of routinization involved, the conceptual situation still requires the elements incorporated in (RA). Thus, even such institutional actions as stopping at a red light or obeying one's promises must involve a proattitude toward these reason states (viz., that the red light is on or that one has meaningfully made a promise). These can be regarded as desire-independent states. Accepting this, my "routinization view" still requires that being connected with promising there must be not only a disposition to keep the promise—in normal cases—but also actually existing, although perhaps only a tacitly possessed, a proattitude toward keeping it (and ditto for other similar institutional cases). That a reason is a motivationally effective reason clearly requires that there is something relevant going on in the agent's psychology, which brings about the right "acting on" process.

19. I am presently discussing especially what is to be found on p. 511 of Audi's 1986 paper and 2001a book, especially p. 115.

As to Audi's claim that mental states provide reasons, I am not sure how tight a connection "providing" here is meant to establish between mental states and reasons, but I will not here press the point.

20. See sections I and IV and, especially, Broome 2000.

21. Audi 1986, 511–12.

22. Ibid., 537–38.

23. See ibid., 523; and Audi 2001b, 33.

24. Compare, e.g., the clear statement in Audi 2004: "Reasons based on desires may be called *internal* to contrast them with normative reasons viewed as *independent* of what the agent wants and in that sense external and objective." This passage also clearly shows that Audi does accept so-called desire-independent reasons—central in my account (also cf. Searle 2001).

25. Indeed, in Audi 2001b, 21 he says: "To be normative (for a notion of judgment or proposition) is to be governed by a standard that is proper, in a sense implying that violations of the standard merit a kind of disapproval and that conformity merits a kind of approval and, commonly, praise (as where conformity is difficult)." This I take to apply also to technical normativity. If so, it is difficult see what kinds of reasons are not normative in Audi's account, understanding normativity in the quoted sense.

26. I have elsewhere (in Tuomela 2007) presented the following accounts of I-mode and we-mode reasons:

> (*IMR*) Reason R is a group member's motivating *I-mode reason* for performing an action X if and only if R is the agent's main *private* reason for his performing X. Typically, R is a state that the agent wants or intends to be the case or a state that, according to his belief, obtains; and X is an action that is a means to R or an action that R requires for its obtaining such that the agent is privately committed to performing X on the basis of R.
>
> (*WMR*) Reason R is a group member's motivating *we-mode reason* for performing an action X if and only if R is the agent's main motivating group reason for his performing X. Typically, R is a state that the group in question wants, intends or requires to be the case or is a state that, according to the group's belief, obtains; generally speaking, R is a state that is "for the group." X is an action that is the individual's part of a collective action that is a means to R or a collective action that R requires for its obtaining, where the group members are collectively committed to performing the collective action for reason R and mutually believing so.

My criteria speak of an agent's "main reason" for action, which notion is taken to involve that for a reason to be a main reason it, considered alone, typically would (or at least ought to) suffice for (her commitment to) the action in question. In (*WMR*) X can be, e.g., a collective (or group) action with multiple tokens (e.g., going to church on Sundays) or a joint action like cleaning up a park as a many-person action. As the group members are collectively committed to performing the collective action in question for reason R (a state expressible by a that-clause), they are also socially committed to the group members to performing their parts of the collective action for reason R. A full-blown, viz., a we-mode, group reason will have to satisfy the *Collectivity Condition* (a kind of "common fate" condition) that says, roughly, that necessarily, if a group member has that reason, every other group member has it, on the basis of the members' collective acceptance of the reason as the group's reason.

Having a private commitment means in (*IMR*) that the person privately (rather than as a group member) has psychologically bound himself to a "content," e.g., to performing an action for a reason. In general, private commitment is dependent on an intention, here the intention to reach a goal. Notice that functioning in the we-mode is necessarily connected to a "thick" group

reason, to what one's group has committed itself in the situation at hand, where the group's commitment serves as an authoritative reason for the participants. In contrast, functioning in the I-mode is at most only contingently connected to a group reason (when there happens to be one involved). ([WMR] and [IMR] can be generalized to cover any voluntary attitude and action.)

27. See Richerson and Boyd 2005 and Tomasello et al. 2005 for recent discussions defending the importance of cooperation and (in the latter text) also shared intentionality as evolutionary adaptations peculiar to humans as species.

28. With several qualifications, people tend to be cooperative in in-group contexts but competitive and even hostile in out-group context.

29. See Tuomela 1995, and especially Tuomela 2000 for a detailed account.

30. See Tuomela and Bonnevier-Tuomela 1997.

31. My account concentrates on conformative we-attitudes, but there are also nonconformative (and "neutral") we-attitudes but they will not be discussed here. (See Tuomela 2007 for various other cases where the social reason is based on others' attitudes different from ATT.)

32. The so-called toxin puzzle may be taken to be problematic here, but actually it is not, for my account does not depend on the assumption of the sameness of the social reason for the attitude and the reason for the ensuing action.

33. ATT could even be the intention-in-action to have a swim, when swimming already is taking place.

34. Mutual beliefs must, under an iterative construal of them, involve loop beliefs of the kind "I believe that you believe that I believe that s" and thus it cannot be strictly said the agent x here is responding to his environment. The agent's own belief state is built into the notion of an iterative mutual belief. However, if this is considered to be a problem, mutual beliefs must be understood differently—e.g., in terms of a fixed point analysis (which is only at the infinite limit equivalent to the iteration analysis (cf. Balzer and Tuomela 1997). Alternatively, the mutual belief could be a weaker one, e.g., "general belief" in the group without assuming the iterative account of this notion.

35. I have elsewhere, in Tuomela and Bonnevier-Tuomela 1997 and Tuomela 2002, 2007, discussed collective social actions as collective actions performed *because of* a shared we-attitude.

36. Also, social norms and obligations can function as social reasons in the present account. What we have been discussing above are *protanto* reasons, viz., reasons that do not need to entail that the agent strictly ought to do the action that he has the reason for, even if in a sense the totality of such reasons—if the agent had them all—can be taken to entail that the agent *ought to do* the "reasoned" action. In contrast, in the case of social norms we are dealing with *normative requirements*, in the terminology of Broome (2000). Thus, consider the norm "everyone ought to do A when in circumstances S." If this has the logical form "if in S, then one ought to do A" we can still be dealing with a *pro tanto* reason. But, if in contrast to this "external" form, it is taken to have the "internal" normative logical form "one ought (to do A when in S)" we are dealing with a normative requirement. That is, here the social norm entails a normative requirement but not a *pro tanto* reason for the target persons (perhaps unless in cases where the participants have not yet decided to obey the norm or the norm in question conflicts with another norm). The action can be derivatively detached (from the norm premise) in the case of a *pro tanto* reason but not in the case where the norm represents a normative requirement. For instance, promising to do A is putting oneself under

an obligation to do A. Here the obligation represents a normative requirement. But when one fulfills the promise to do A, then, *for the reason of having promised*, one does A. That one has promised here is a desire-independent reason (viz., normative requirement) for performing A. Analogously, institutional actions are in part normatively required and in that sense "reasoned" by institutional norms, e.g., "redbacks count as money in our group" or "the university regulation is that every professor must attend faculty meetings." Thus the use of money (redbacks) and a professor's taking part in the faculty meeting are in part "reasoned" by the institutional norms (and other relevant facts). For instance, stopping at a red signal is such an institutional case, the state of the red light being on here being a socially created normative reason that the agent may or may not accept as his motivational reason for stopping his car. In this kind of case the reason state may in general be a state to which we attribute the normative content that one ought to do A when in situation C.

37. As said in a previous note, a full-blown group reason (we-mode reason) must satisfy the collectivity condition expressing the participants' "common fate" and their "standing or falling together" due to their collective acceptance of the reason as their group's reason.

38. As to bootstrapping, see Bratman 1987, chapter 2, for the single-agent case and Tuomela 2007, chapter 6, for the joint case.

39. We get the following partial characterization for the belief-based case (cf. [RA_b*]):

(JRA_b*) q was the *joint reason$_b$ for which $x_1, \ldots x_m$ performed* A in S if, in S, $x_1, \ldots . x_m$ mutually believed that q and there was a state or event r such that q in a reason-preserving way generated the joint want or intention that r of $x_1, \ldots x_m$, and $x_1, \ldots x_m$ mutually believed that they can achieve r by and only by their jointly performing action A, and they jointly performed A on this want or intention and this belief.

As to the want-based (or proattitude-based case), let me suggest the following simple and rather obvious sufficient condition for a want-based joint or shared reason (reason$_w$):

(JRA_w*) q was the *joint reason$_w$ for which $x_1, \ldots x_m$ jointly performed* A in S if, in S, $x_1, \ldots x_m$ jointly wanted q and $x_1, \ldots x_m$ mutually believed that they can achieve q by and only by their jointly performing action A, and they jointly performed A on this want and this belief.

40. In cases of joint action falling short of acting jointly as a group (I speak of I-mode joint action here), we can still use the same notation, but the joint intention will then only be understood in the weak we-attitude sense, which does not entail, e.g., that collective commitment to the content is present, nor does it entail that the collectivity condition (CC) discussed in Tuomela 2000 and 2007 is satisfied).

41. In contrast to this strategic case, we may consider an agent who is maximally conformative and thinks that the other player similarly will cooperate conformistically and non-strategically ("parametrically"). Thus the row player has as his goal to cooperate if the other cooperates and as his belief that the other one will indeed cooperate, all this being mutually believed. Aiming at CC (the result of both agents choosing C) is the agents' we-goal here. Here the agents must act on dominance, dominance meaning that CC is better than DD for each of them. (Notice conformist agents could alternatively end up choosing DD if for some reason they would believe that the other one will choose D.)

42. See Tuomela 2007 for detailed comparison of several kinds of cases from the point of view of I-mode and we-mode reasons.

43. For such authorized members that I have called "operative" ones, see Tuomela 1995, chapter 4; and 2000, chapter 6.

44. For detailed and up-to-date analyses of these cases, see especially Tuomela (1995, 2007). My qualifications concerning the possibility that the authorized members can act for the group of course applies also to my extension of Audi's account below.

References

Audi, R. 1986. Acting for Reasons. *Philosophical Review* 95: 511–46.
——. 2001a. *The Architecture of Reason*. New York: Oxford University Press.
——. 2001b. An Internalist Theory of Normative Grounds. *Philosophical Topics* 29: 19–46.
——. 2004. Reasons, Practical Reason, and Practical Reasoning. *Ratio 17*: 119–49.
Balzer, W., and R. Tuomela. 1997. A Fixed Point Approach to Collective Attitudes. In *Contemporary Action Theory 2*, ed. G. Holmström-Hintikka and R. Tuomela, 115–42. Dordrecht and Boston: Kluwer.
Bratman, M. 1987. *Intention, Plans, and Practical Reason*. Cambridge, Mass.: Harvard University Press.
Broome, J. 2000. Normative Requirements. In *Normativity*, ed. J. Dancy. Oxford: Blackwell.
Dancy, J. 2000. *Practical Reality*. Oxford: Oxford University Press.
Mackie, J. 1974. *The Cement of the Universe*. Oxford: Oxford University Press.
Miller, K. 2006. Social Obligation as Reason for Intention and Action, *Cognitive Research Systems* 7: 273–85.
Richerson, P., and R. Boyd. 2005. *Not by Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago Press.
Searle, J. 1995. *The Construction of Social Reality*. London: Allen Lane.
——. 2001. *Rationality in Action*. Cambridge, Mass.: MIT Press.
Stoutland, F. 2001. The Belief-Desire Model of Reasons for Action. Uppsala University: Department of Philosophy.
——. 2005. Reasons for Action and Psychological States. Paper presented at the conference "Action in Context," University of Zurich.
Tomasello, M., M. Carpenter, J. Call, T. Behne, and H. Moll. 2005. Understanding and Sharing Intentions: The Origins of Cultural Cognition. *Behavioral and Brain Sciences* 28: 675–735.
Tuomela, R. 1976. Causes and deductive explanation. In *PSA 1974, Selected Proceedings of the 1974 Biennial Meeting of the Philosophy of Science Association*, ed. R. S. Cohen, C. A. Hooker, A. C. Michalos, and J. W. Evra, 325–60. Dordrecht and Boston: D. Reidel.
——. 1995. *The Importance of Us: A Philosophical Study of Basic Social Notions*. Stanford Series in Philosophy. Stanford, Calif.: Stanford University Press.
——. 2000. *Cooperation: A Philosophical Study*. Philosophical Studies Series. Dordrecht and Boston: Kluwer.
——. 2002. *The Philosophy of Social Practices: A Collective Acceptance View*. Cambridge: Cambridge University Press.
——. 2005. We-Intentions Revisited. *Philosophical Studies* 125: 327–69.

———. 2007. *The Philosophy of Sociality: The Shared Point of View*. New York: Oxford University Press.

Tuomela, R., and W. Balzer. 1999. Collective Acceptance and Collective Social Notions. *Synthese* 117: 175–205.

Tuomela, R., and M. Bonnevier-Tuomela. 1997. From Social Imitation to Teamwork. In *Contemporary Action Theory, vol. 2: Social Action*, ed. G. Holmström-Hintikka and R. Tuomela, 1–47. Dordrecht and Boston: Kluwer.

This page intentionally left blank

REASON AND INTUITION IN THOUGHT AND ACTION

Replies and Extensions

This page intentionally left blank

Intuition, Reflection, and Justification

ROBERT AUDI

This chapter will both respond to the critical papers in part 1 and extend my views. I cannot reply fully to such challenging papers. I must be highly selective, emphasizing differences with the authors rather than our areas of agreement; but I can pave the way for continuing reflection and debate on the central problems. I begin with ethics. Chapter 16 concerns epistemology, with some supporting new work in a related area of philosophy of language. Chapter 17 addresses the philosophy of action—including moral psychology. It draws on the previous parts, but each part and indeed, each response to a given chapter, is largely self-contained.

In ethics as elsewhere in philosophy, intuitions are commonly accorded prima facie credibility. This is not because of their content but because of how they are grounded. Paradigms come from the domains of perception, consciousness, memory, and reason (where logical intuition is central). Ethical intuitions may have either singular or general propositions as objects, and these propositions may be either empirical or a priori. My concern will be only empirical intuitions and mainly quasi-perceptual intuitive moral judgments. These are not unqualifiedly perceptual, but the properties on which their truth is grounded may often be directly apprehended and are sometimes perceptible.[1]

Reflection versus Inference in the Grounding of Moral Judgments

Ethical intuitionism is in part based on the view that moral knowledge and justified moral judgments need not be inferential. A plausible intuitionism, however, grants that premises may be available for such direct knowledge or justification. It seems to some philosophers that premises *must* be available. Walter Sinnott-Armstrong's case for that view is subtle and challenging.

Sinnott-Armstrong's View of Intuition

For Sinnott-Armstrong, my intuitionism implies that "reflection is what makes moral intuitions justified" (20). He maintains, moreover, that judgments justified as what I call conclusions of reflection fail to meet my condition of noninferentiality:

> In Audi's poetic example, "one judges from a global, intuitive sense of the integration of vocabulary, movement, and content." This "sense" seems to be articulate, since the reflecter is reading and reflecting on language. . . . At the very least, the reflecter needs beliefs about which words are in the poem, or else he could not know which poem he is reflecting on. Thus, whatever they are, some beliefs are formed during reflection and are needed for that reflection. Let's call all of them together "the beliefs of reflection." (22)

Using my notion of one belief's being inferentially grounded in another when the first is "held on the basis of the other" (my words), he maintains that "this definition seems to imply that conclusions of reflection are inferentially grounded in beliefs of reflection" (22). I have two points here.

The first concerns what beliefs, if any, are required for adequate grounding of a conclusion of reflection. Sinnott-Armstrong speaks as if one had to have beliefs about what words are in the poem. On my view (supported by the distinction between dispositional beliefs and dispositions to believe developed in 1994), one need only *understand* those words and appropriately *respond* to them. Suppose that a basis on which we hold the conclusion that the language is artificial is the presence of the French phrase *de trop*. Granted, a belief that this phrase is "highfalutin" *could* be one ground of a judgment that the language is artificial. But must such a classificatory belief be antecedently held or formed on the occasion?[2] Could we not simply *sense* the artificiality of *de trop* in the context? Perhaps we have heard it in pretentious speech and associate it with artificiality. That association might have some evidential value. Again, we might be *disposed* to form beliefs expressing these points, but we need not.

My second point here should now be apparent: the perception of a property can ground a judgment without doing so by yielding beliefs that supply premises for that judgment. Consider, for instance, facial recognition regarding someone you have not seen for many years. If the judgment that the person is, say, an old friend from high school, arises from thoughtfully contemplating facial properties, but is not based on *beliefs* of supporting propositions, we may call it a conclusion of reflection even if the person could formulate "corresponding" premises. The judgment may, then, be both noninferential and intuitive.

Second-Order Beliefs and Inferential Resources

Background beliefs of various kinds may be important as necessary conditions for justification. Sinnott-Armstrong rightly wonders why I speak of the intuitive judgment emerging "from a global, intuitive sense of the integration of vocabulary, movement, and content" unless the person believes these are relevant. One might believe this, or be disposed to believe this; but it is a contingent matter just how much, if anything at all, we have to believe *about* the relevance of elements in order

to learn how to judge *by* them. Indeed, in the order of learning, one would expect that judging on the basis of a ground is prior to beliefs about its relevance to judgments of that kind.

Supposing I am right about this, Sinnott-Armstrong still maintains that "the reflecter needs at least some implicit beliefs about when reflection is adequate in order to know when to stop reflecting" (25–6). Is this so? Much as, when we look into a room to see if the table is set, we stop when we see enough, we can stop reflecting on whether a poem's language is artificial when we read enough. This is usually when we reach the judgment we sought to make (or some judgment that brings a sense of something like closure). Again, we *may* believe we have reflected long enough; but we need neither believe this nor have a criterion for its truth in order to reflect enough to make a judgment. Compare, too, recognizing a painting, which may require simply looking at it until its identity 'hits us'. A justified conclusion of reflection, then, does not need a "second-order belief about the adequacy of the reflection" (26). Some destinations are identified for us only by road signs that provide information we must read on signs and believe; others are recognized directly.

If Sinnott-Armstrong accepted all this, he would still hold that the person should consider "whether his reflection is adequate. This reflecter is [otherwise] not very reflective . . . a reflecter needs to think about whether he should trust the source of his belief. Not to do so is irresponsible" (26–7). I doubt that all moral beliefs are subject to the indicated condition. But, for the sake of argument, suppose they are. It might be clear enough, on reflection about whether to trust one's grounds, what they are and that they are strong enough to sustain our judgment. *These* points could become clear and be justifiedly accepted without inference from premises.

In any case, I doubt the supposition: even apart from reflection on the status of our grounds, we might acquire justification simply by appropriately responding to adequate grounds, where, in contrast with taking these grounds as premises, our response is noninferential. If the grounds *are* justificatory, this response will serve — as where one simply has someone read the poem again or reads it with the right expression — without our having to infer our judgment from them or show *that* the grounds are justificatory (a requirement that generates a regress).[3]

The last problem to be considered here emerges from my granting that "in principle, where one arrives at a conclusion of reflection, one could determine why and then formulate, in premises, one's basis for so concluding" (2004b, 47 [hereinafter referred to as *GIR*]). I added "that a ground of intuitive judgment *can* be formulated through articulation of one's basis for judgment does not entail that the ground must do its justificatory work in an inferential way" (2004b, 48). Sinnott-Armstrong's response is that "a skeptical regress arises from the need for an ability to infer" (28).

A regress would arise if *every* evidentially necessary ground of a conclusion of reflection were itself such a conclusion. But I am not committed to this implausible view. Take a case in which a ground of a judgment that I ought to help a friend is my sensing his need for my help. I may have to listen to him a long time to get the signs of illness or pain on which the judgment is based. I might be able to formulate indications of these, such as a hoarse voice; but do I need to be able to formulate my ground for believing him to be hoarse? Believing this can be simple

recognition. A skeptic may demand premises; I am sympathetic with appropriately *selective* skeptical demands, and I would grant Sinnott-Armstrong that reflective people tend to seek premises more often than others. But demands for premises need not be considered reasonable across the board.

I have not implied that *every* intuitive moral judgment need be a conclusion of reflection. Some moral judgments and moral beliefs are nonreflectively as well as noninferentially justified. A more sensitive observer of my friend might realize virtually immediately that she should help. And if a babysitter annoyed with a toddler burns it with a cigarette to stop its humming, this can be noninferentially seen to be wrong. The wrong is so obvious that 'intuition', used for the belief that it is wrong, is misleading. But my point does not depend on appeal to that contested term.[4] Allaying the doubts of the (relatively) few who doubt there can be noninferential justification here, however, requires nothing less than the full-scale epistemology I have set out in 1993d, 2001a, and 2003.

Much can be learned from Sinnott-Armstrong's subtle challenge to the case for possible noninferential intuitive justification. I have tried to suggest how moral beliefs and judgments can possess it. I have distinguished between believing something—whether reflectively or not—*on* a ground and inferring it from the proposition *that* this ground is present. I have suggested that second-order beliefs are not needed for the justification of either kind of belief or judgment, even if they are appropriate to *showing* that one has such justification; and I have contended that ability to infer a proposition is not a general requirement for justifiedly believing it.

Rational Disagreement as a Challenge to Ethical Intuitionism

Conscientious rational persons who find others disagreeing with them often wonder who is right. As we saw in considering whether beliefs are ever justified by a ground that the believer cannot articulate in a premise for inference, there may be beliefs that the person cannot argue for by appeal to premises. But suppose we think that, for such a belief on some important moral matter, a person who disagrees with us is as rational and generally informed as we. Should we still adhere to our view?

The Epistemic Role of Consensus

Roger Crisp is inclined toward a negative answer. His essay particularly emphasizes what he calls Sidgwick's consensus condition. In Sidgwick's words:

> [D]enial by another of a proposition that I have affirmed has a tendency to impair my confidence . . . the absence of . . . disagreement must remain an indispensable negative condition of the certainty of our beliefs . . . if I have no more reason to suspect error in the other mind than in my own, reflective comparison between the two judgments necessarily reduces me temporarily to a state of neutrality. (32)

With this in view, Crisp poses a serious problem for my position:

> R [who believes that p] comes across another thinker, S, who asserts *not-p*. Immediately, according to the consensus condition, this should impair her confidence

in *p*. . . . But unless S's epistemic state is significantly faulty in one or other of these ways, *R* may well feel that her trust in *p* has been dented. (33)

In Crisp's example, R is supposed to be justified in believing that S is in roughly as good an epistemic state as R. Should R's confidence diminish? One difficulty with this case is that confidence is a psychological notion whose bearing on justification needs clarification. The strength of our confidence that *p* need not correlate highly with our degree of justification for it. Furthermore, Crisp is thinking of self-evident propositions, which are only one kind I hold to be (often) intuitively knowable. Even when they are known, however, they need not be believed with "certainty"— an elusive concept I have treated in detail elsewhere (2003, chaps. 8 and 10). There are also degrees of justification; hence, ample justification for believing a proposition might not be of the high degree appropriate to certainty.

Crisp raises a related problem, pertinent to *any* kind of proposition and not undermined by these clarifications. "R . . . has merely to ask herself whether it is possible that another person, in as epistemically good a state as she is, might reasonably hold *not-p*. If it is, then she should suspend judgment on *p*." Whatever he thinks Sidgwick may have been committed to saying about this problem for intuitionism, Crisp says that it "rests on a dim view of human epistemic capacities" (33). From his example concerning disagreement about what kind of bird is before us, it is clear that he takes the notion of an epistemically good "state" to include things like visual acuity and favorableness of the light. This notion encompasses a huge number of variables. Background beliefs are included, as are inferential capacities, reliability of memory, and conceptual sophistication. He might grant, then, that we are commonly not justified in believing that someone else is in as good an epistemic position as we.

More important, the fact that it seems very clear to me that, for example, there is a robin before me is some reason to take a person who disagrees *not* to be in as good a position. It might seem that if we could test every epistemically relevant variable, we might instantiate the unfortunate position Crisp describes. But suppose we conscientiously test every variable we consider relevant. Doing so (for the purpose in question) presupposes that we trust our own beliefs enough to make a reasonable comparison of someone else's grounds with our own, now checked out. You must, for instance, ascertain that I discern someone's hurt feelings as clearly as you, if my disagreement is to force you to doubt your judgment that an apology is owed for having hurt them. Now, however, if you finally see, as you list your freshly scrutinized grounds, that I still cannot be brought to agree that she has been hurt, you may have *better* reason to think you missed some relevant difference between us— a difference favoring you—than to think you should suspend judgment.

In the very act of conscientiously comparing my epistemic situation with yours, you must, in a certain way, trust *your* judgment along the way if you are thereby to arrive at a conclusion you may justifiedly hold. In particular, you must anchor your judgment that I am in a good epistemic position by presupposing some judgments of yours as to, say, whether I am aware of certain elementary, perceptible facts. You need not *always* favor your judgment over mine when we disagree, but the confidence you have that I am in a good epistemic position must come from relying on *your* epistemic position, and the strength of *that* position will be confirmed by the

entire exercise of reconsidering both positions. Dogmatism is undermined; justified conviction need not be.

Dissensus as a Threat to an Intuitionist View of the Self-Evident

As I have developed intuitionism, it is essential that the claim (common to at least the most prominent versions) that some moral principles are self-evident *not* be taken to entail their being *obvious*. I do not think they need even be psychologically compelling: considering them with adequate understanding does not entail believing them. Indeed, if one holds background beliefs, or accepts theories, that call for rejecting a self-evident proposition, disbelief may be possible in this same case. It is not always clear, however, that denying a proposition entails disbelieving it, as I think the case of self-deception (among others) confirms.[5] Our inferential behavior and our other beliefs are also important for determining what we believe. With all this in mind, I have distinguished agreement *on* reasons from agreement *in* reasons. Let me explain its bearing.

Suppose a Rossian promissory principle is self-evident and that it is correspondingly self-evident that promising to A entails having some reason to A. There is disagreement on both points,[6] but someone who denies that promising entails a reason to A may regularly regard the making of *particular* promises as implying such a reason. Might such a person still accept as true someone's denial of having a reason to A, after having promised to A? Suppose so. We would wonder why. Holding a theory requiring this denial is one possible explanation; another—compatible with this one—is an inadequate understanding of some relevant concept, such as that of *having some reason*. This concept as understood here embodies defeasibility, which is no easy notion to grasp.

These points are among those that make clear how the kind of self-evidence required by (rationalist) ethical intuitionism is compatible with theoretical and other kinds of opposition. But theoretical opposition applies mainly to disagreement *on* reasons and need not be manifested in disagreement in them. This is where Crisp raises further doubts. He grants that there is sufficient consensus on such self-regarding normative principles as "pain" ("Non-deserved suffering of any sentient being that would be caused in or by some action of that agent counts [though not always decisively], for A, against the performance of that action by A" [34])—to provide a sound basis for intuitionism. But, he asks, "how should R proceed in philosophy? And how should she live?" (36). In answering, he says, "Once she realizes that these principles fail to meet the consensus condition, it will become clear to her that they have no claim to self-evidence, to being objects of knowledge, or to any directly justified role in her deliberation" (37). Several points are in order.

(1) For reasons already offered, the consensus condition is too strong. But suppose it is sufficiently plausible to require suspending judgment on whether Rossian moral principles are self-evident; (2) it does not follow that the principles have no claim to being *known*. Both empirical and certain kinds of a priori principles can be known without being self-evident; (3) it also does not follow that the principles have no directly justified role in deliberation, say, that they must be supportable by

utilitarian or Kantian considerations. Noninferential justification for believing them would suffice for the kind of defeasible role my intuitionism requires.

The Dualism of Practical Reason

Suppose I have been right so far in defending my ethical intuitionism. There remains the problem of how to deal with conflicts of obligations and, especially, of self-regarding reasons with other-regarding ones. Sidgwick despaired of solving this problem, and Crisp seems to share much of his pessimism:

> That ethical intuitionism should lead to a result in which each of us is left decid-
> ing which ethical theory to follow on the basis of how much each will advance our
> own self-interest is somewhat odd. But if normative philosophical ethics can
> progress to the point of widespread convergence on a comprehensive view, it would
> provide a practical role in our lives for ethical theory to which such theory has not
> yet earned any entitlement. (38)

Crisp's second, conditional point may well be correct. If a "comprehensive" view is a master principle view, such as Kant's, widespread convergence on it (on suffi-ciently good grounds) would entitle it to a practical role. But my efforts in ethical theory have been aimed in part at suggesting that Rossian intuitionism as developed in *GIR is* comprehensive in the basic sense of capturing at least the vast majority of our obligations. Its principles may thus serve—and very often do serve—as common ground to be accommodated by competing theories such as Kantian and conse-quentialist views. They also provide an important *area* of "convergence."

As to Crisp's first claim, both of us, as well as Sidgwick, Ross, and many others, hold that there are basic self-regarding and basic other-regarding reasons and that these often conflict. Sidgwick despaired of a good solution; Ross appealed to prac-tical wisdom to deal with such conflicts; and I have provided two approaches to resolving the problem. The first, developed in my *Architecture* (2001a, mainly chap. 6) defends the reasonableness of altruism (for people with certain not uncommon characteristics) and places the conflict in question in the context of a comprehen-sive theory of rationality. The second, developed in *GIR*, supplements both that approach and Ross's by appealing to the intrinsic end formulation of the categori-cal imperative, which I interpret in an intuitive way that does not depend on any particular interpretation of Kant, and to a theory of value.

None of this implies that there is a perfectly adequate way to determine when self-regarding reasons override other-regarding ones. In my view, however, this is a problem for any plausible ethical theory. It does not, however, require countenanc-ing what Sidgwick called "irrational impulses." Perhaps it would commit us to that if practical reason were "divided against itself" in a sense implying incoherence; but conflict of prima facie reasons does not entail incoherence.

Even if our only resort were Aristotelian practical wisdom, there would still be intu-itively plausible solutions to individual conflicts that command, if not assent by all who are adequately rational and sufficiently informed, then their respect. Indeed, suppose nor-mative properties are consequential on natural ones (in ways suggested in *GIR*). Then the facts underlying a conflict between self- and other-regarding reasons should ground

either a preference for one choice or support a disjunction of equally permissible choices. The latter case may be more common than generally supposed. Here certain kinds of emotional or other preferential inclinations might properly decide. Emotions can indeed have evidential value in the first place (as stressed in *GIR*, chap. 2).

A final point here is implicit in the idea that the institution of morality serves human flourishing. I take this to entail a role in sustaining and promoting cooperation. It is entirely appropriate to that role that morality should often force us to ask whether we do enough for others. To place the general good so far above yours or mine that we are obligated to approach treating ourselves merely as means to enhancing it is one mistake. To think that morality should make it easy for us to tell whether are doing enough for others is another.

Moral Motivation and Volitional Intuition

So far, I have focused mainly on how moral principles and singular moral judgments can be known or at least justified. But ethical theory goes well beyond moral epistemology. A major dimension of it concerns the connection between moral judgment—especially when both self-addressed and justified—and motivation. Although I have attacked motivational internalism conceived as the view that motivation is *intrinsic* to (holding) moral judgment, I have also argued for an integration between moral judgment and motivation *in rational persons*. In them, holding a self-addressed moral judgment normally implies motivation.[7] Suppose, however, that one posits a sense of duty as a distinctive kind of propositional attitude with both cognitive and conative elements. Might one then integrate reason and motivation *in the will*?

McCann's Conative Theory of Moral Judgment

Hugh McCann's theory portrays moral judgment—at least where it has an adequate basis—as embodying motivation. It is not just an apprehension of moral properties; we should

> [B]egin with the will: specifically, with . . . *felt obligation*—that is, the sense of duty . . . felt obligation lies at the heart of moral behavior—that is, of acting from duty. . . . It is a kind of movement of will or emotion, the thrust of which is that although, subjectively, there is a choice to be made, there is a certain objective settledness about the options . . . although felt obligation is at least largely a conative state, it may legitimately be viewed as a form of *objective* awareness, and hence . . . ground for judgments. (46)

I accept much of this. My main reservation concerns packaging so much into what seems to be a single psychological state: felt obligation is conative and has a behavioral and typically future-directed object; but awareness of the relevant kind is cognitive and has a set of properties or of propositions as object. Should we posit a single state with two such different kinds of objects, one kind truth-valued (or at least factive, since one cannot be aware of what does not in fact exist) and the other not?[8] Why not grant that the conative state—which may be part of a moral emotion (as McCann credits me

with implying in *GIR*)—may *evidence* a moral judgment in the cognitive sense, and may also motivate the conduct that the judgment favors, rather than build motivation into (holding) the judgment? McCann posits a unity of different elements in a single psychological state where I would distinguish two kinds of state and provide an account of when they achieve the integration that McCann melds into a unity.

Prospects for a Conatively Based Intuitionism

If such conative elements as a sense of duty can be evidence for the truth of a moral judgment, might we go further and argue that some kind of volitional response to morally relevant facts is the basic kind of intuitive evidence for moral judgment? For McCann:

> Our cognitive recognition of them [the moral features of actions] is *indirect*, grounded in the conative experience of felt obligation, in which we apprehend our duty by feeling impelled toward or away from a given course of action. . . . It is through this kind of experience that duty is first grasped. . . . The principles that emerge, since they are propositions, must still be treated as cognitive intuitions, but the intuition occurs only by means of conative experience. . . . (50)

There are connections here to the is-ought issue, to moral development, and to the interconnected roles of intellect and will in grounding moral knowledge and justification. I take these in turn.

When McCann says that descriptive features—say, the property of being a fatal stabbing—do not entail evaluative ones, the Humean is-ought (and fact-value) gap comes to mind. I agree that there is no formal entailment; but on even a moderate rationalist view, there are nonformal entailments of normative propositions. To be sure, an act's being a fatal stabbing (of someone else) does not entail that it is wrong *on balance*. But that it is prima facie wrong does follow on any plausible rationalist intuitionism. Is that proposition not normative ('evaluative')? I consider it so: it expresses a moral reason for avoiding such an act. McCann need not deny this, but his ostensible endorsement of a Humean is-ought view makes it harder for him to appreciate the resources of a moderate intuitionism that does not presuppose the essential role he assigns to the conative.

Concerning moral development, it is possible that if children could not be given a sense of moral demandedness, they also would not develop the cognitive capacity to know moral principles. McCann may be presupposing this important point. I am inclined to suppose it is true. It does not entail, however, that moral knowledge is "grounded in the conative experience of felt obligation, in which we apprehend our duty by feeling impelled" (50). A genetic condition for the development of knowledge need not be a ground of it. I agree, however, that our cognitive grasp of moral features of action (e.g., obligatoriness or wrongness) is indirect. Moral knowledge, even if noninferential, epistemically depends on knowledge of, or justification for cognition regarding, some set of natural (roughly "descriptive") properties on which the relevant moral property is consequential.

I do not, then, unqualifiedly accept McCann's conative grounding of moral intuition. Granted, a child who never felt obligated—and (to take a kind of example

McCann does not stress) never felt a moral revulsion to being abused or to seeing its mother squelched by a domineering father—might acquire no moral knowledge. But a child who never experiences shapes might be unable to see the truth of the proposition that nothing is round and square.

To be sure, the geometrical analogy applies to general moral knowledge. But would McCann deny that once that knowledge is acquired, *some* moral judgments, say that it is wrong to pay Juan more than Juanita simply on grounds of different gender, can be known by subsumption under a general principle rather than by a conative inclination? Even apart from this, whatever the genetic role of conative experience, it seems possible to know that an act is wrong (or is obligatory) on the basis of cognitive awareness of its "descriptive" properties, even if, as with weakness of will, one lacks the appropriate conative inclinations. It is perhaps easiest to see this if we consider third-person moral judgments that do not even directly call for any action by the agent.

I find in McCann's paper, then, resources that may advance the kind of ethical theory we share, but I do not accept either the moral psychology that gives a fundamental role to a kind of mixed theoretical-practical state or the moral epistemology that posits a universal dependence of moral knowledge on the will. That there may be such states, however, and I see no need to deny that moral knowledge may arise from volitional elements. Moreover, that we must educate the will as well as the intellect, and that emotion and motivational inclinations can have evidential value, is a major point of agreement between us.

The Content of Moral Principles and the Scope of Morality

The relation between the good and the right—more broadly, between the axiological and the deontic—is a perennial concern. Consequentialist ethics treats the good as more basic. So does Aristotelian ethics. Ross is more difficult to describe: he sees some duties as grounded in the value of the consequences of the obligatory acts in question, but refuses to take all obligations as so based (1930, chap. 2). Kant is similarly deontological in taking some acts to be wrong or obligatory on nonconsequentialist grounds, yet he also says that the *value* of persons underlies the categorical imperative (*Groundwork*, sec 428).

My view is Rossian in affirming a deontology that takes promotion of the good to be one obligation, Kantian in providing a framework in which moral principles can be seen as standards whose internalization expresses respect for persons, and Aristotelian in taking morality to be an institution that serves human flourishing understood partly in terms of the realization of virtue. One can, however, conceive morality on Hobbesian natural law lines. On this view, powerfully and comprehensively represented by Bernard Gert, moral principles are those that all rational, adequately informed persons would put forward as a public system of rules.

Moral Principles and Moral Ideals

It is instructive to compare Gert's views with mine concerning the content and ground of (basic) moral principles. For him,

> [E]very rational person regards these rules as moral rules because he wants these rules obeyed to protect him and his friends, and realizes that if these rules are to be accepted by all as moral rules . . . all must put forward these rules as protecting all moral agents. Audi seems to accept the standard philosophical view that morality is primarily a guide to conduct that people adopt for themselves. (56–7)

I doubt that the view Gert describes as standard is so if taken as a motivational thesis about why people adhere to moral rules. In any case, the question dividing us is what *justifies* the set of sound basic moral principles. Here I doubt we differ greatly, since I conceive the principles (at least the Rossian ones) as such that our internalizing them tends to protect us all. But given their self-evidence, I deny that knowledge of their doing this is necessary for justifiedly believing them. Gert accepts some of this:

> Audi and I agree that my first five rules, or his first middle axiom [that we should not injure or harm people], are self-evident. However, Audi, as an intuitionist, regards coming to accept the middle axioms [Rossian principles of prima facie duty, explicated in chap. 5 of *GIR*] on the basis of a proper understanding or grasping of their content as adequate for being justified in accepting them, and I do not. . . . It is necessary to show that appropriately described rational persons would put forward the rules as public rules governing the behavior of all rational persons. (56)

Perhaps Gert is thinking of *accepting* them *as justified* and is making a plausible claim about what it would take to *show* that they are justified or true, as opposed to simply justifiedly believing them. I doubt that he must be understood as articulating a condition necessary for justifiedly accepting them. He later says, however, that what explains "why rational people would regard this middle axiom [that one should not lie] . . . or the equivalent rules, as self-evident, can explain why they would sometimes want them not to be followed" (58). The explanation is that rational persons would not want these principles violated toward them *unless* there is a rationally acceptable exception. Here Gert's conception of self-evidence differs from mine and is closer to something like the notion of intuitive plausibility.

The importance of this difference should not be exaggerated. I agree that moral principles can be supported from the general point of view of rationality, as argued in *Architecture* (chap. 6). But since, unlike Gert, I grant a kind of epistemic autonomy to Rossian moral principles, neither their justification nor their interpretation depends on "external" grounds in the way he supposes. With this in mind, let me comment on some further differences between our views.

Our main difference regarding Rossian principles concerns whether moral obligations extend to "imperfect duties," such as the duty to do good deeds. He has no difficulty agreeing that we are obligated to avoid doing injustice, but denies "that a monk or a scholar has even a prima facie obligation to leave his monastery or library in order to rectify or prevent injustice" (59). I grant that this prima facie obligation can be so clearly overridden that it is misleading to *assert* it. But suppose a monk sees that he could simply step outside and, by speaking briefly, dissuade an angry family from lynching a teenage offender who, being hungry, stole an apple from their tree. Might it not be morally wrong for him to abstain for the sake of an additional stretch of meditation?

One thing Gert says that is highly pertinent to this is that 'moral rules' refers to those "it would be immoral to violate without an adequate justification. These rules—for example, Do not kill, do not deceive, and do not cheat—can be obeyed impartially with regard to all moral agents, and unjustified violations of them make one liable to punishment" (56). By contrast, "moral ideals, which encourage people to prevent others from suffering harm or evil, for example, help the needy . . . cannot be followed impartially with regard to all moral agents, and failure to follow them does not make one liable to punishment" (56).

This is not the place to frame an account of punishment. Perhaps public moral censure is one kind. The imagined monk *is* liable to that. Still, unless the notion of punishment is taken considerably more broadly than this suggests, it appears that liability to punishment is not a condition for (unjustifiably) violating even a moral principle Gert considers such. Think of certain lies or broken promises to a spouse. In any case, there is a kind of moral criticism that, whether or not we consider it punishment, might be properly felt to be *punitive* and may be appropriate, even in public, for violating an obligation of beneficence. Perhaps Gert could agree that this, rather than something like deprivational or physical punishment, is a better candidate to help us delimit ethical obligation.

As to impartiality, is it true that the kinds of apparent obligations in question cannot be impartially followed? If this means that one cannot do the same amount and kind of good for all in the world, it may be true; but surely one could have a *policy* that governs relations to all equally even in a quantitative sense. Does morality require adherence to such a policy, however? Gert and I both deny this. Obligations to family members can outweigh obligations of beneficence. This does not entail, however, that there are only ideals, rather than obligations, of beneficence.

The Scope of Morality

In some ways, it is of minor importance how we delimit the realm of specifically moral principles, as opposed to that of normative standards that properly guide action. Once it is agreed that punishment is not always appropriate for violations of moral principles and that some kind of criticism *is* appropriate for (even if not advisable for) violations of any normative standards, we can consider what should and should not be done and what steps are appropriate when someone violates a sound normative principle. Still, for purposes of ethical theory, we should explore the distinction between moral and other normative principles.

Consider Gert's interesting claim, "On my view, and that of Mill and Hobbes, morality is concerned only with the way you treat others. A rational person need not care how others treat themselves" (61). I would divide these points; one might hold the first and deny the second. I begin with three comments on the latter point. First, Mill's valuational hedonism apparently commits him to taking the rationality of action to be based on its contribution to the net balance of pleasure as against pain, and in places he indicated that *everyone's* pleasure and pain is relevant. Second, beneficence surely requires us to care about the good of others even when that is a matter of how they treat themselves. (Granted, the acts appropriate for us here are, as Gert might stress, toward others.) Third, can we appropriately specify fully what

kind of things rational persons care about without considering their life circumstances and their psychology? A rational person *could* be so situated as not to care how others treat themselves; but it is imaginable that a rational person could be so situated as not to care about whether others lie (such a person could be reclusive, self-sufficient, and able to tell when others lie).

My fourth point concerns whether morality can be self-directed as well as other-directed. Consider the difference between a Hobbesian natural law view as Gert understands it, and Mill's position. The former view is most plausibly understood as supporting principles to be backed, as a legal contract is, by the force of coercion, including legal force. People may properly be forced, on pain of punishment, not to kill, defraud, commit injustices, and so forth for the most stringent principles. This view is like many contractarian approaches in beginning with the idea—or idealized assumption—that rational persons are self-interested. For Mill, however, we are to promote the good as best we can. Moral obligation derives from the requirements of doing this and extends to how we treat ourselves. Perhaps Gert takes Mill to be closer to Hobbes than he is because of the centrality, in Mill's political liberalism, of the harm principle, which prohibits our coercing a competent adult except on grounds of (sufficiently) probable harm to *others* (Mill 2003). But this is a prohibition of *coercion*; endorsing it does not commit Mill to holding that only violations of this principle are moral wrongs.

It is interesting to note in this connection that many terms for people who violate the beneficence and self-improvement principles seem, if not intrinsically moral, to be normative in a way that at least suggests moral import. A person who does nothing for others when it is not highly demanding is considered selfish, thoughtless, insensitive, and the like. One who does nothing toward self-improvement is considered, for example, slothful, lazy, uninspired, or self-indulgent. Gert could point out that the latter terms, at least, do not imply immorality. Still, the first person could not be said to be morally sound, and, as Gert could grant, neither may such a person be said to have good moral character. Perhaps another applicable term is 'poor ethics'. In any case, it seems plain that such people do not fall short only of *ideals*. They do not even approach any moral ideal in the relevant domains, roughly those of social conduct and self-regarding motivation and behavior.

A final point here is that both self-improvement and beneficence concern not just reducing suffering and enhancing enjoyment but also improving people in regard to certain virtues. Ross stressed this, but I would add that, assuming the institution of morality should serve human flourishing, it is altogether appropriate that its principles should in part direct us toward strengthening our adherence to some of their kind. Whether, as Gert might agree, it is best in the end to treat the prima facie obligations of beneficence and self-improvement as moral and as expressible in moral principles is less important than whether we consider them sound normative standards. It continues to seem to me appropriate to consider them moral.

Kantian Intuitionism and Ethical Pluralism

The Good in the Right defends a revised Rossian intuitionism but goes on (in chap. 3) to develop a Kantian intuitionism that appears superior to either Rossian intuitionism

or Kantian ethics taken alone. This theory is intended to retain what is sound in each. The Rossian framework gains in unity; the Kantian theory gains in clarity and usefulness in guiding moral judgments. An integration like this is liable to criticism from both sides. Thomas Hurka vigorously challenges it from the Rossian side.

Conflicting Duties and Kantian Imperatives

In resolving conflicts of duties, Ross held, no general theory helps any more than applying practical wisdom. In arguing that the categorical imperative framework can help, I tried to retain some of what is best in Kant, but not to provide a full interpretation. On my view, we can interpret the two central notions of Kant's intrinsic end formulation of the Imperative—the notions of treating persons merely as means and of treating them as ends—in a way that is both intelligible apart from Kant's overall ethics and illuminated by it. The former—*merely instrumental treatment*—is understandable in terms of the notions of using an instrument and of being motivated in certain ways that go with this. This notion is factual and—in a certain sense—'descriptive'. Treating as an end is largely a matter of acting for the good of the other for its own sake, say by relieving suffering. The notion is prima facie normative, but not specifically moral. Accordingly, both notions can provide non*moral* grounds for moral judgments, just as the paradigmatic Rossian grounds do. Moreover, I take it to be intuitively clear that treating people merely as means is prima facie wrong and treating them as ends is prima facie morally good and in some cases prima facie obligatory.

   Now consider Hurka's claim regarding the case where, to tend a sick child, one breaks a promise to meet a friend. Hurka says that Kant's universalizability formulation is no help because the friend would accept my decision "only because he would see that in this case the moral duty to prevent harm is stronger than the duty to keep one's promise, which is just Ross's claim" (66). Must we say "only because"?

   One answer in *GIR* is based on the intrinsic end formulation of the categorical imperative (I should add that my *main* use of the universalizability formulation in *GIR* is in testing decisions and principles, not in arriving at them). First, then, we may see the preference for tending the sick and later explaining the broken promise to the promisee as ways of treating both parties as ends, and this may clarify our reflection. Second, since there is both a duty (certainly a moral reason) to treat the child as an end and a duty to prevent harm to the child, the case can be seen as one in which the promissory duty conflicts with *two* others one has toward the sick child. The comprehensiveness of the notion of treating as an end does not prevent the duty to do this from being intuitive and from adding to the overall reason to tend the child.

   A different example will also show how a comprehensive duty can help in resolving a conflict of more specific duties. A duty of reparation to the child would be comprehensive in scope (though less so than the duty to treat the child as an end), and, in a similar way, unspecific: one might owe the child amends for previous neglect, but find the field of appropriate compensations very wide—wide enough to include rendering aid now. Even if we might still intuitively see the prevalence of the duty of beneficence apart from the duty's receiving support from the duty of reparation, the point is that a wide, unspecific duty can help in dealing with a conflict between two

more specific ones. The duty to treat persons as ends is extremely comprehensive, and, granting that we might intuitively see the prevalence of the duty of beneficence over that of fidelity without conceiving the case in terms of treatment of persons as ends, so conceiving it can indicate an independent moral reason for the priority and *also* provide an explanation of the intuitive sense of prevalence.

This is in no way an anti-intuitionist point. I take the intrinsic end formulation to be intuitively plausible if understood as I propose. In Rossian terms, I bring more comprehensive intuitions into the picture to supplement those underlying the Rossian duties. For one thing, if, to avoid breaking a minor promise to A, I let B get sicker, I would typically be in effect weighting A's good over B's. That the notions of treating as an end and avoiding treating merely as a means are comprehensive does not prevent their being intuitively applicable in a way that gives them a measure of epistemic authority regarding singular moral judgments. The intuitions that accord with this approach have an authority much like that of those grounded in, say, a sense of inequality or of injury. That an act would restore equal treatment and that it would reduce injury are reasons that support it; so would the facts that it would avoid using someone merely as a means or would treat a person in the kind of positive way that bespeaks viewing the person as an end.

At one point Hurka both makes a concession and raises a doubt. He says that the intrinsic end formulation yields

> two basic duties: to treat persons as ends and not to treat them just as means, where the second . . . is stronger. If successfully established, this priority claim can do some explanatory work, showing why specific negative duties such as not to harm . . . can all outweigh positive duties. But how much epistemic work can it do? Will it be easier for us to decide whether the duty to keep a specific promise outweighs a specific duty to aid if we ask how much, in general, the duty not to treat as means outweighs the duty to treat as ends? . . . The more abstract question seems harder to answer. (66)

This seems to acknowledge the plausibility of taking the intrinsic end formulation to express duties. But it also suggests that my Kantian intuitionism takes the avoidance of merely instrumental treatment automatically to outweigh the duty to treat as an end. I have argued, however, that intuitionism is not *hierarchical*, with one or more kinds of duty *always* prevailing over some other kind(s), and I have not implied that we can, in general, determine by "how much" one duty overrides another when it does. I grant that comparison between these two kinds of duties is "more abstract" than between most pairs on Ross's list; but that does not imply either unresolvability of conflicts between them or, more important, that such conflict can be addressed only by deciding conflicts of duties using just Ross's list. In my view, treating someone like a dispensable tool is objectionable even if it does not violate one of the Rossian duties, say, that of noninjury. We do not always damage tools we use, even if we are ready to discard them after the job is done.

The Beneficence Problem

Hurka and I share a concern with how to limit the duty of beneficence, but we differ on whether Kantian intuitionism helps in doing so. Demandingness is a problem

for *any* plausible moral theory, and we should not expect a precise answer. One point I made is that if we move in the direction that act-utilitarianism takes us—and in which Ross needs a good rationale for not moving too far—the intrinsic end formulation (as I interpret it) is illuminating. I suggested that if we regard ourselves as obligated, overall, to maximize the good of persons, we are liable to treat ourselves merely as a means to that goal—where, in a troubled world like this, we ourselves may matter, at most, minutely (*GIR*, 97). I specifically allowed that one could *voluntarily* undertake selfless philanthropy. But Hurka says,

> If . . . I ask Bill Gates to donate his fortune to humanitarian relief and he does, I do not treat him merely as a means, since I act only with his consent. And acting only as another consents to is normally thought sufficient for treating him as an end. But if I voluntarily contribute my own fortune to humanitarian relief . . . does that not answer any charge of treating myself just as a means? Does consent not cancel disrespect? Audi thinks it would if contributing the fortune were merely a voluntary ideal . . . but not if the beneficence is a moral duty. But how can the specific content of my motive, and in particular whether I am acting as I think I ought, affect whether my act is voluntary in the way that amounts to consent? (68)

First, I do not think that our consent to what is done toward us entails its constituting treating us as an end. Consent can be manipulated, misinformed, even irrational. One *could* consent to conduct that treats one merely as a means. The *basis* of consent is crucial for whether it legitimizes the conduct consented to. Part of what should be considered in distinguishing consent that warrants a kind of treatment from consent that does not is whether the conduct would be treatment of the person as an end (in terms of what is for the good of the person) or would be, or would approach, treating the person merely as a means.

Second, as to the content of motivation, this *can* be relevant to whether, in agreeing to something, I am being treated as an end. If I give money to a cause only because I want to avoid severe punishment for failing to (so that my motive has a kind of self-preservational content), the fact that this is my actuating motive typically implies that I am not acting freely.

My third point is more closely connected with the *basis* of motivation. Moral principles should be internalized and are properly backed by at least social pressure, such as liability to public criticism for failure. If there is an overall obligation to act a certain way and one acts that way on the basis of internalizing the relevant principle and so for a morally approvable motive (say, to avoid harming someone), there is a sense in which one did as one *must*. This case has indeed been called moral compulsion. My overall point in *GIR* was in part that a principle that makes what should be voluntary in the sense of *discretionary* morally obligatory is mistaken, and I think that the Kantian idea helps us to distinguish between the obligatory and the discretionary. It thus helps to explain why the act-utilitarian standard is too demanding. It also helps us see why inculcating in children a certain kind of utilitarian commitment to beneficence would be moving toward treating them as means to maximal promotion of the good in general. To subject adults to the pressures of such a demanding standard goes in a similar direction. Granted, there will be hard cases in which it is unclear whether enforcing a demand for beneficence, even by

the social pressures of threatened moral censure, would tend to treat someone merely as a means to promoting the good of others; but having this way to explain why such a demand may be excessive, as opposed to using the resources of Ross alone, provides an advantage over Ross's account.

I should emphasize that I have not tried to give what Hurka calls "a Kantian account of beneficence" (69). Rather, I employ the intrinsic end formulation to add to our intuitive reasons not to let the duty of beneficence swamp all or most of the others. He apparently doubts, however, whether my understanding of beneficence is even compatible with Kant's ethics: "how can a value that is unchanging, and based on a property that is likewise unchanging, make it right to favor some changes in a person rather than others, say, ones that make him happier? . . . How can a value that is independent of any states of affairs lead us to promote some states and not others?" (69). My main response to this is that although Kantian dignity does not admit of degrees, people may be treated more or less fully as ends, and, related to this, their good may be served well or poorly. Moreover, I do not rely on the notion of Kantian dignity as a major element in understanding treating persons as ends; the latter, as I explicate it, seems prior in the order of explanation. Even apart from that priority question, we can take the values of states of affairs to be important for treating persons as ends without implying that persons differ in the degree of their dignity.

Prospects for Unification of Intuitive Moral Principles

Hurka and I agree that even self-evident principles can be supported and unified in relation to wider principles. Here I find his positive use of a Thomistic view plausible. We

> see the duties to promote and not destroy the good as irreducibly separate. . . . The view then applies its two duties to each of the different goods. If pleasure is good and pain evil, there is a positive duty to promote pleasure and prevent pain but an even stronger duty not to cause pain directly, so that even if torturing one person would stop two others from being tortured, the torturing is wrong. . . . This Thomist view adds several elements of unity to a Rossian deontology. First, it relates all moral duties to intrinsic goods and evils, understood as being located in states of affairs. This should be agreeable to Audi. . . . (70–1)

I find much that is plausible here. I prefer to take it into account on the basis of the axiological grounding of moral principles set out in chapter 4 of GIR, but even Kantian intuitionism can accommodate the perspective. Suppose that the basic goods that concern us are (as I proposed) experiential, hence occur in the lives of persons, and that the evils to be avoided are either cases of treating persons merely as means or at least of acting in opposition to treating them as ends—this last case may be a matter of such disrespectful treatment as talking loudly during someone's lecture, which need not be instrumental treatment of the lecturer at all but is diametrically opposed to treating the person as an end. We can take it as intuitive that the latter concerns have priority over the former. This priority is apparently also entailed by Kantian ethics. We are thus close to accommodating Hurka's framework.

Kantian intuitionism, then, is more accommodating, and less dependent on Kant's own claims, than Hurka implies. If I have proposed any marriage at all, it is between a revised Rossian view and one that is Kantian in lineaments. It seems compatible with the Thomistic unification Hurka suggests, and that, in turn, is surely compatible with the full theory I develop by providing an axiological grounding for Kantian intuitionism itself.

Ethical Method, Moral Principles, and Intuitive Judgments

Ross did not see Kant as a deontological ally, and his appeal to the singular moral judgments of the "plain man" would not have had the high epistemic authority in Kant's eyes that they had for Ross. The suggested methodological differences should not be exaggerated, however, and they do not undermine my integration between Rossian intuitionism and a Kantian ethical view centered on my interpretation of the categorical imperative. A major meeting ground of the two theories is intuitive moral principles—the kind that Ross took to express prima facie duties and that are, in Kantian language, categorical imperatives with a small c. Here I revisit both perspectives in the light of Candace Vogler's novel interpretation of Kant and Ross.

Moral Principles: Abstract Content and Internalized Guides

In *GIR* I noted that "whereas Ross stressed intuitive induction . . . Kant resoundingly asserted that one could not do morality a worse disservice than to derive it from examples" (198–99). Vogler responds (in part) that

> [W]hatever their differences, it is false to characterize Kant as embracing "a top-down conception of the determination of moral obligation," whereas Ross gives us a "bottom-up theory." Rather, powerful Kant is best read as a "bottom-up" man, the chief weakness in Ross traces to the "top-down" cast of Anglophone moral philosophy. (73)

We may gain clarity by dropping the metaphors. Kant also said, "Even the Holy One of the gospel must first be compared with our ideal of moral perfection before we can recognize him to be such" (*Groundwork*, sec. 408). This and many other Kantian statements suggest that knowledge of generalizations and ideals is epistemically prior to knowledge of singular moral judgments. For Ross, by contrast, knowledge of singular moral judgments is epistemically prior to knowledge of moral principles. This is not to say that Kant is committed to the impossibility of knowing a singular moral truth other than by *inference* from a principle, such as a categorical imperative. The point is that—in these and many other passages, at least—the knowledge of general moral truths seems to be a ground of the *capacity* for knowledge of singular ones. Similarly, Ross is not committed to our knowledge of moral principles' being inferential; indeed, as self-evident, they are (on his view) never known on the basis of premises. He did hold, however, that knowing them requires a certain conceptual capacity ("mental maturity") and that the very concepts essential in them are grasped only by "intuitive induction" (a procedure clarified in *GIR*, chap. 2).

Granted, Kant may in fact have learned much from reflection on particular cases, and certainly Ross's generalizations allow for subsumptive (hence inferential) knowledge of singular judgments, for instance for inferring the obligatoriness of saving a child from the fact that only I can safely do it and the generalization that we are prima facie obligated to save others' lives when we alone can safely do it. *Both* philosophers allow that one can achieve moral knowledge by subsuming a case under a general principle; but this similarity is consistent with my contrasts between them. There is, however, a contrast that Vogler thinks I miss:

> [T]he categorical imperative formulae are unlike the other principles [such as *Promises are to be kept*]. . . . First, it is not clear *what* they enjoin. Second . . . they will, ideally, underwrite everything that a finite, dependent rational being does. . . . This is clearest for the universal law formula. Imagine yourself trying to operate as though you are legislating and enacting universal law every time you lick a stamp. . . . Turning to the second formula . . . ends-in-themselves include friendship, health, pleasure, and justice. . . . Humanity is not an end in that sense. . . . The idiom of ends, and of ends-in-themselves, is instead productive. So is ordinary practical reasoning. . . . Worse . . . ideally, humanity will be both end and means in everything we do. What could possibly count as means and at the same time end of moving a piano, making a paper hat . . . or your favorite intentional act-type. (75)

Much is said in this passage. I agree that it is not clear just what Kant's formulations enjoin; but if we impose a high standard of clarity, this holds for some Rossian principles, for instance the principles calling for beneficence and self-improvement. As to the idea that Kant's formulae "ideally, underwrite everything" we do, I grant that Kant sometimes talks as if all rational action is based on a maxim in a way that instantiates an almost self-conscious guidance by the categorical imperative; but I doubt that he must be read as having so intellectualistic a theory of action. In any case, I see no reason why Vogler need deny that the Imperative implies that there *are* reasons for action and that, once we interpret it so that it applies to some of our options, rational commitment to it can *give* us reason to act whether or not we must call it or any rule to mind in order to determine what to do.

Regarding her views about ends, I am puzzled by the claim—in criticism of Kant and perhaps me—that nothing can "count as means and at the same time end of [action]" (75). Kant's view requires that *people* can be treated as means and *also* as ends, and it allows that the same action, such as asking someone for advice, can be a case of treating partly (though not merely) as a means and partly as an end; it does not imply that the same *act* can be both means and end of performing another. Given my explication of treating persons as ends versus merely as means, the first point is perfectly clear. I can ask your advice, thus using you as a means, while also treating you as a valued friend. Moreover, contrary to Vogler's suggestion in the stamp case, Kant is not committed to the view that all action, even all action affecting a person, is "treatment" of some person. Much is surely not, as where your buying the last ticket affects the person behind you in a normal queue, and there is thus still less reason to think that Kant must take the Imperative to govern every action in the suggested way. (How it does govern action is detailed in chap. 3 of my 2006b.)

The Internality of the Moral

Although my use of Kant's work is not intended to reflect his view in detail, my hope was to produce a theory that, in its use of the categorical imperative, is at least in the spirit of Kantian ethics. If Vogler is right, even this hope may be unfulfilled:

> [T]reating the categorical imperative as of a piece with other principles involves trying to use the formulas in a way that they ought not to be used. . . .
>
> For Kant, there can be *no* question why quick-witted, healthy, adult humans take an interest in ethical matters . . . it is *in* us, actually, to be tuned to ethics . . . it is cosmically impossible that there should be a finite, dependent, rational being that was not accountable to the moral law by its own lights. . . . In Audi's system, by contrast . . . it is hard to see how the ethical is anything other than one among the many fields of engagement that might attract our interest. (77)

An interesting question here is what it is for a principle to be internal to a creature. The context suggests that a kind of *internalization*—perhaps occurring only when maturity is reached—might be what she has in mind. But why should Rossian principles (or mine) differ from Kant's on this score?

Regarding the claim that it is "cosmically impossible that there should be a finite, dependent, rational being" not "accountable to the moral law by its own lights," if this is plausible, it is in part because the kind of rational being in question is conceived as *reflecting* on the categorical imperative in relation to other such beings. There, however, the contrast with Rossian principles is less sharp—and certainly not sharp in a way that undermines my points about the relation between Rossian and Kantian principles.

Vogler also suggests that in my system the ethical is nothing more than "one among the many fields of engagement that might attract our interest." Our "interest" contrasts with our being "accountable to the moral law" by our own lights. Accountability is a normative notion; interest is a psychological one (in this context). I can grant that our actual motivation to be concerned with ethics is psychological, though I think it is not merely one among other kinds of motivation: I leave open that given our social nature, it may be a psychological law that we *tend* to be rule-following (even in a way that implies being ethical to some degree). But on the normative matter of accountability, I have argued (elsewhere) that *rational* persons under certain common social conditions will tend to hold moral principles under which they are accountable for their actions (*Architecture*, chap. 6). Here we may be close to agreement.

In her conclusion, Vogler locates me as further from Kant and even Ross than I acknowledge: "neither sees the task of moral theory as providing comprehensive normative guidance" (80). If moral *theory* is taken to be moral *philosophy*, this is arguable; but surely a moral theory in that sense is in part an attempt to account for principles that *do* provide normative guidance. Whether a set of moral principles is "comprehensive" depends on how strong that vague term is. As to the internality of the grounds of moral judgments, given my emphasis on reflection and on the experiential aspects of value, I doubt that there is a plausible kind of internality of the ethical that cannot be accounted for on my view. Why might Vogler think otherwise?

Contrasting me with Ross, she says, "*Nowhere* in Ross's account do general principles take center stage as the basic stuff of ethical life. Upbringing, habituation, character, practical wisdom, interpersonal relations, relative power and vulnerability, merit, differential welfare—these inform conduct essentially" (79). If the basic stuff of ethical life is constituted in good part by the experiential elements central in conducting oneself as a moral agent, I not only make room for the elements she lists; I have also indicated how some of them work in ethical matters, for instance in describing the role of emotion in supporting moral judgment (e.g., in *GIR*, 56–57). But she is quite correct in suggesting that I have been more concerned with ethical theory than with moral phenomenology (though I have outlined a position on that in my 1998 which supports the views defended here and in response to Sinnott-Armstrong and others). Nonetheless, the ethical theory I have developed is at least hospitable to the rich phenomenology she has portrayed.

Notes

1. The relation of moral to natural properties is discussed in detail in my 1993c.

2. I do not think it plausible to posit such a belief if we clearly distinguish between dispositional beliefs and dispositions to believe. Suppose 'it' is in the poem; must we, when we see the word while reading, come to believe that 'it' is in the poem? We *would* come to believe this if asked (when we see it) whether 'it' is in it. But suppose we did come to believe this. That belief would not be an inferential *ground* of our conclusion that the language is artificial. This crucial distinction (which is often overlooked or underemphasized) is defended and developed in my 1994.

3. The issue here is whether we can be justified by our grounds without being able to show them to be reliable. I have argued that this is possible, in 2003, chap. 10.

4. For a case against excessive reliance on intuition, see Williamson 2004. It is not clear whether any of his points raise doubts about the weak prima facie justification I accord to intuitive beliefs. It is *grounds* of beliefs that I take to confer justification on them (in ways described in, e.g., 2003), and believe his view requires a similar position.

5. This point is defended in some of my papers on self-deception, especially 1988.

6. The main issue here is the particularism-generalism contrast; my generalist view is defended, most recently, in 2006a.

7. This matter is discussed in detail in my "Moral Judgments and Reasons for Action," in 1997a.

8. An extensive case for cognitive and conative attitudes having different kinds of objects is made in my "The Grounds and Structure of Reasons for Action" (forthcoming b).

Justifying Grounds, Justified Beliefs, and Rational Acceptance

ROBERT AUDI

In discussing the justification of moral beliefs and judgments, I relied on examples to show the plausibility of taking sensory experience to provide noninferential justification for the kinds of perceptual beliefs that—being crucial in identifying certain grounds of obligation—are essential in understanding justification in ethics. One question raised by the epistemology I presupposed is what sorts of beliefs may be noninferentially justified. Another is the sense in which grounds of justification may be internal. Still another is the range of attitudes that admit of justification and (more broadly) rationality. I will pursue these questions in that order, with the aim of making a cumulative case for an important part of my epistemology.

Perceptual Justification and Experiential Content

In everyday life we navigate the world in the light of myriad perceptions. We commonly do this with no sense of making inferences about what we see, hear, and otherwise perceive. But even if perceptual beliefs are not inferential, they may be inferentially dependent on other kinds of beliefs: structurally inferential, even if not episodically so.[1] Their status in this respect is a major issue dividing epistemologists.

Visual Experience as a Model

A number of my writings defend the thesis that visual experience grounds prima facie justification—a kind that is defeasible even when "conclusive." I have also defended the a priori status of certain principles ascribing this kind of justification to experiences. For instance, a clear and steadfast arboreal vision—the kind we consider characteristic of seeing a tree—grounds justification for believing there is one before us. Laurence BonJour is inclined to accept such principles but doubts they can be appealed to as I have in defending the noninferential justification of perceptual

beliefs, such as that there is a tree before me. His critical examination begins with an interpretation of part of my view:

> [E]ven though we cannot know that accepting beliefs according to this principle [one ascribing prima facie justification to visually grounded beliefs] will lead us to the truth, we can still know that beliefs accepted in this way are justified. While the visual experience principle is not . . . "partly constitutive of truth," it is nonetheless "partly constitutive of epistemically permissible attempts to show truth." (91)

My claim was that we cannot know *a priori* that there is an *objective* likelihood of truth, one entailing that in at least the majority of relevantly similar possible worlds the proposition in question is true. That paper takes seriously the possibility that a Cartesian demon could produce arboreal experiences where there is *not* a tree before us. I am supposing that we cannot know a priori that the demon world is not "close."[2]

What is wrong with this approach? For BonJour, although, on my epistemic realism, it is "*part* of our very conception of such objects that they lead to experiences of that kind and also that beliefs about them accepted on that basis are justified. . . .

> [S]uch objects are of course *also* conceived as having . . . physical and causal properties that are entirely independent of such epistemic claims. And Audi's view seems to me to amount to saying that while both of these aspects are involved in our conception of material objects, we have in the end no reason at all for thinking that a case in which the justification conditions are satisfied is thereby one in which the other features required for such an object to exist are likely to be present. (92)

The critique continues: "all we can say is that *these* experiences are justificatorily relevant to the belief that a large deciduous tree exists in a certain location simply because *these* are the experiences whose relevance is partly constitutive of the concept of such a tree. But then any experiences . . . could . . . have . . . played that role" (93).

I do not think my view has this consequence. I grant (at least for the sake of argument) that a Cartesian demon could cause us to have rosebush-indicating experiences when we are in fact visually affected by a tree (I leave open whether this could be seeing a tree *as* a rosebush). But I am not committed to holding that the kinds of arboreal experiences we normally have are only contingently related to the *concept* of a tree. I think BonJour and I agree here; it is essential to the concept of a tree that trees have, for example, branches, and it is essential to an arboreal visual experience—one *as of* a tree (a noncommittal expression I use to cover both veridical and nonveridical experiences)—that it exhibit, for example, branches. A demon could prevent trees from causing such experiences but could not prevent them from grounding prima facie justification for believing there is a tree before us, provided we have the concept of a tree. Their justificatory relevance, then, is a priori. The difference between us is that I do not take it to be a priori that they are truth-conducive in the objective, reliability sense I sketched. What is a priori here is their noninferential role in grounding justification as a normative status, one such that the relevant beliefs are rational on the basis of such appearances.

A related question is whether we could have a concept of a tree without actually experiencing one. For the sake of argument, I will assume that we can. But even

if we could not, it would not follow that we can determine a priori that our arboreal experiences reliably indicate trees. Perhaps, however, that is so. I would welcome this news.

The Connection between Justification and Truth

BonJour is not a philosopher to observe a malady of the serious kind he thinks may afflict me without prescribing a cure (I attribute this to some combination of beneficence and an irresistible attraction for philosophical theory). He finds it through a conception of explanation:

> [O]ur sensory experiences give us good reasons for our correlative beliefs about material objects because the existence of the objects in question provides the best *explanation* for the existence of such experiences. Defending this . . . is notoriously fraught with difficulties. But . . . it is very hard to see how it can fail to be true if accepting beliefs about the material world on the basis of that experience is to be a reasonable way to seek the truth. (94)

The idea (which I have myself expressed in a number of places) is that what best explains the kind of experiences that ground a perceptual belief that there is a tree before one is that a tree (of the relevant kind) is producing it. I can further agree with BonJour, at least for the sake of argument, that "the full experiential basis for supposing that a three-dimensional object of that general shape is to be found in the world would, at least in principle, include the shapes reflected in all of the various experiences I might have in (apparently) approaching the tree and moving around it" (94). The crucial question is not whether this is true, but whether it sustains an a priori claim to the effect that such experiences are objectively reliable in the sense I have sketched.

In pursuing this, it may help to work with a more modest epistemic principle than any so far formulated: if our only good explanation for a proposition we are amply justified in believing entails the likely truth of a further proposition, we are prima facie justified in believing the latter proposition.[3] This is more readily defended than the best explanation principle for two reasons: first, our "best" explanation, comparatively understood, may not be good; second, even a good explanation might be rivaled by one almost as good. The latter case would cast doubt on our having the kind of justification adequate (where the other conditions for knowledge are met) to ground knowledge.

Granted, it is widely supposed that the existence of trees is essential for our *only* good explanation of arboreal experiences. On certain plausible assumptions about explanation, this supposition is arguably a priori. But the case for it seems strongest where the concepts of explanation and of justification in question are internalist, in a sense entailing that beliefs so justified need not be true in most relevantly similar possible worlds. If, however, only *true* propositions can explain, then even if we know the principle a priori, I do not see how we can know, *a priori*, that in most cases of arboreal experiences, an instance of its antecedent—that there being a tree before us *is* our only good explanation (much less *the* only good explanation) of our arboreal experience—is true.

BonJour argues (in this paper) only briefly for the apriority of any such explanation principle. Granting that "there are as many possibilities as human imagination can devise . . . [that] fit experience perfectly: brain-in-vat hypotheses, Cartesian demon hypotheses, and so on" (96), he suggests that such explanations are inferior to the commonsense hypothesis because "by introducing an extra layer of complexity, they are less likely to be true. . . . For these reasons, I believe that a case can be made . . . that the commonsense explanation is indeed superior to the apparent alternatives on a purely a priori basis" (96). His main thrust, however, is toward showing that we must presuppose such a principle, objectively understood, in order to justify taking external object beliefs to be noninferentially justified. He says, for instance, that given the best explanation assumption in question "our perceptual experiences can constitute *in themselves* good reasons for thinking that the perceptual beliefs about such objects that we accept on the basis of them are true" (30).

Let me first take up the apriority question. May we assume, with BonJour, that some version of Occam's razor—understood as roughly the principle that in framing and adopting explanations and theories we should prefer (on grounds of likelihood of truth) the simpler, other things equal—is a priori? To assess this principle, we should distinguish internal from external epistemic considerations: roughly, between justificatory elements—evidences—that are accessible to reflection or introspection and, on the other hand, objectively truth-conducive factors (BonJour's concern) that need not be thus accessible or, if they are, cannot be known a priori to be objectively truth-conducive. Is the simpler view better justified, on internal grounds (other things equal)? This may seem obvious. I doubt that it is. What may be obvious here is something easily conflated with the principle conceived epistemically. This is the principle that it is preferable to *work* with a simpler view. That, however, is a practical principle and could hold even if there were no difference in terms of evidential value between more and less simple views. Call this *the principle of least effort*.

Belief, Acceptance, and Intellectual Economy

The kind of strong connection between justification and truth that BonJour seeks to make is not established by the principle of least effort or any other practical principle. Let us ask, then, whether we have reason to think that from an external perspective simpler views are more likely to be true. Perhaps so, given common experience and the track record of scientific hypotheses. The simpler have tended to be better confirmed (though this might be at best hard to show owing to the difficulty of separating considerations of simplicity from those of degree of confirmation). I do not see how to demonstrate this. We might entertain an evolutionary explanation say that since we do in fact prefer simpler hypotheses, we would not have survived if the ways of nature were not better evidenced by simpler hypotheses. This would provide at best limited support, however, and would *presuppose* that we have perceptual justification in the first place. It also provides only empirical justification for the epistemic preferability of the simpler.

These points do not undermine the intuitively plausible idea that the simpler of two otherwise equally plausible (or equally explanatory) hypotheses is *preferable*. This is plausible, and a version of it may be a priori. To see this, let me start with an

important distinction often not observed between belief and acceptance. For most of the everyday cases of belief-formation, 'accept' does not even seem natural. Where 'accept' is natural (for propositions), it contrasts with 'reject' and 'withhold'; and in that probably central use, acceptance does not entail belief. We can accept *p* while withholding—or simply not forming—belief that p. Now there *is* an a priori principle to the effect that if one needs to act and cannot do so without certain information (e.g., directions someone gives to a place one must find on pain of death), then, in the absence of reasons to doubt an apparent source of such information, one should *accept* its deliverances and is justified in so doing.

Skepticism at the theoretical level, then, is one thing; but in practice we would not be fully rational if, when there is no positive reason to doubt the testimony of the senses or indeed of someone else (or that it provides information we need), we did not take it as worthy of acceptance, however tentative, as a basis for action. The implicit acceptance principle, however, is pragmatic, not epistemic. Its own a priori status (if it is a priori) does not imply the same status for its epistemic counterpart, in which belief replaces acceptance. If no epistemic counterpart is a priori—whether for testimony or for perception—the case for the apriority of some version of the best-explanation principle (Occam's razor understood epistemically) is weakened. But must that principle be a priori in order to play a role in grounding the noninferential justification of perceptual beliefs?

Consider a reliabilist view on which there are no substantive a priori principles (a notion of the a priori clarified and defended in my 1999b and forthcoming c). On such a view, perceptual beliefs can be both noninferential and, being reliably grounded, justified. The *principle* that perceptually grounded beliefs are mostly true, however, is not taken to be a priori; and it is granted that it cannot be justified without circularity (as argued in detail in Alston [1993]). We would have to rely on perception, for instance, to determine how often it apparently delivers true beliefs. But they could be reliable—and could even be based on sensory experience in the way BonJour outlines—even if one could not show this without circularity.

Neither BonJour nor I is satisfied with reliabilism concerning justification. But I have been content to support the apriority of certain epistemic principles and to show that, given these together with internally accessible grounds for belief—such as visual experiences—we may be noninferentially justified in beliefs about the external world and may indeed provide a noncircular argument to the effect that we have beliefs justified in the internalist, as opposed to reliabilist, sense. This answers global skepticism about internal justification. BonJour agrees on the apriority of the principles I appeal to, but takes their (objectively) truth-conducive counterparts to be a priori. He can thus argue against skepticism about external (objectively truth-conducive) justification. My doubts about this view turn mainly on the scope of the a priori. The case I have made for its having narrower scope than on BonJour's view is one I would gladly lose.

Testimonial Justification, Trust, and Credulity

Historically, epistemology has focused mainly on the single-agent case. But even if this is the epistemologically basic case, we all rely heavily on one or another kind of

testimony for at least a huge proportion of the things we know. My epistemology of testimony seeks to do justice to both its essential role in human knowledge and what I have called its *operational dependence* on perception. I have also given different accounts of two of its major roles: grounding knowledge and grounding justification. Elizabeth Fricker has developed a different view, and her paper instructively contrasts her position on testimony with mine.

Testimony and the Social Transmission of Knowledge

We agree that one can come to know a proposition on the basis of testimony only if the attester knows it in the first place. Our difference lies in how this thesis is to be explained. In her view, the thesis—"K-Nec"—cannot stand on its own nor be adequately defended on reliabilist grounds. As "the key suppositions of Audi's attempted grounding of K-Nec," she cites:

> i. If an attester A does not know what she asserts, then her assertion does not express a belief of hers that is reliably true (this is why she lacks such knowledge).
> ii. If A's assertion does not express a reliably true belief, then that assertion itself is not reliably true.
> Therefore: iii. Forming belief on the basis of such an assertion is not doing so reliably, or via a reliable method, in the fashion required for knowledge. (102)

She regards this way of grounding K-Nec as a failure because Lackey (1999, and forthcoming a) "succeeds pretty well in undermining the reliabilist argument for K-Nec" (4). Lackey's counterexamples are indeed challenging, and I want to examine a representative case of hers before considering Fricker's own approach. (I will vary the case somewhat, partly to reflect a similar scenario by Graham [2000] with additional challenging features.)

Imagine a teacher (Luke, let us say) who disbelieves the theory of evolution but teaches it conscientiously. He tells his students, on the basis of his correct reading of the theory and his knowledge of fossil discoveries, that there were *homo sapiens* in a certain place. He thus gives his students correct information for which there is adequate evidence. May we conclude that their testimony-based belief can constitute knowledge without Luke's knowing the proposition in question (which he disbelieves)?[4] Is their belief adequately grounded, if the teacher would have taught a false theory in the same disbelieving way, had this been required by his job? Even if the theory *itself* is (an item of) "knowledge" (being known by someone), *he* isn't a reliable link in the chain from the fossil record through the theory, since he neither knows nor even believes the theory, hence does not believe it on the kind of ground that would protect him from error in the way the (truth-conducive) grounding of knowledge does. By his lights, he is *deceiving the children*—a point important in itself for the epistemology of testimony. Moreover, it appears that he would have been as likely to state a false proposition if the school required his teaching a mistaken theory that seemed to him no more pernicious than this one. Such a person might well be teaching a false theory or one that is not well evidenced and just happens to be correct.

The case has more plausibility on the assumption that the school would not require anyone to teach a theory that is not true or at least well evidenced. Suppose so. But do the students perhaps believe or presuppose something to this effect? In any case, might an essential part of their *basis* for believing Luke, or for knowing from his testimony, be a similar background belief or presupposition? If so, either their belief would be bolstered by background elements in a way that implies that it is at most in *part* testimony-based and hence not what normally counts as a testimony-based belief or, on the other hand, their belief would not count as knowledge, owing to the falsity of their presupposition as applied to Luke. In order to believe what (in this kind of case) Luke says—or at least in order to know its truth—they would presumably have to believe or presuppose something to the effect that this is what the school is teaching. Suppose, however, they simply take his word. Then they are taking the word of someone who will deceive when job retention requires it and (let us charitably assume) the person takes it that there is a plausible rationale for the proposition in question. It is highly doubtful that this kind of testimonial origin would be an adequate basis of knowledge.[5]

Fricker is quite right in holding that I am appealing to reliabilistic considerations in defending the view that the attester's knowledge is necessary for testimony-based knowledge on the recipient's part. But one important suggestion she makes must be qualified. She takes it that I conceive knowledge based on testimony to be "exactly parallel in its internal grounding, . . . to perceptual knowledge . . . and this yields entitled belief, in just the manner that, for instance, 'I may just believe that a bat flew by if I see one zigzag across the evening sky' [where this belief is justified by virtue of seeing]" (103). This claim overlooks my point that although testimony is like perception in being a source of noninferential belief, it is unlike the latter in having an essential epistemic dependence on the operation of another source: one must *perceive* testimony in order to know on the basis of it, whereas no nonperceptual source of knowledge plays a comparable role for perceptual knowledge. A further difference is that the phenomenology of perception is distinctive and in a sense basic, whereas the phenomenology of testimony depends on that of the perceptual mode of its reception, say, hearing versus reading. To know on the basis of your testimony, I must receive it; but it can be loud or soft, written or in sign language, and in English or Italian. I know on the basis of *semantic* representations, not *phenomenal* representations.

Two Kinds of Trust

Even supposing that my defense of the knowledge condition needs no supplementation, it is important to see how trust figures in the genesis of testimony-based knowledge. One of Fricker's important points on this is that "the speech act of assertion is . . . governed by the norm: one should assert that P only if one knows that P" (104). The accompanying epistemology explains both what kind of trust is needed for the acquisition of testimony-based knowledge and, given that, implies that I am mistaken in a sufficient condition for such knowledge. She maintains that since

> [H]uman nature is susceptible to the many motives driving deception, and to honest error. . . . I do not think that we are entitled, as Audi suggests . . . to accept testimony at face value in the absence of empirical warrant for believing the speaker

to be both sincere and competent about her topic . . . we must have empirical war-
rant to take her to have roughly this property: 'Not easily would she assert that P,
unless she knew that P.' (104)

This view may or may not be combined with a requirement that the recipient
believe the attester to have this property—roughly, that of being trustworthy relative
to what is attested. Fricker later endorses a belief requirement (though I do not see
that her overall view requires doing so). Since "part of the recipient of testimony's
basis for her belief is justified belief that the attester speaks from knowledge, then if
that is false her own belief is based on a false premise, and so, even if true, is not
itself knowledge" (104). The reference to premises suggests a further, inferentiality
requirement, on which the recipient's belief that p is grounded inferentially, and
not only justificationally, in the belief that the attester speaks from knowledge in say-
ing that p. Several comments are in order.

First, she has in effect indicated two kinds of grounds for trust: empirical war-
rant for taking the attester to know—warrant for a *fiduciary presumption*, we might
say—and actual belief that the attester does know, a *fiduciary belief*. Let us employ
this distinction to identify two kinds of trust. Both kinds imply a tendency to believe
the attester and are combined with a justification for this—hence normative. But
one is noninferential and the other inferential (exhibiting an inferential depend-
ence on the justified belief even absent actual inference). By contrast, the kind of
minimal trust I take to be required for testimony-based knowledge is simply a natu-
ral credulity of the kind Thomas Reid described. This is neither a presumption nor
an attitude; it is a kind of structural characteristic of human nature. No justification
of the kind Fricker has in mind is necessary for this natural trust or for the acquisi-
tion of knowledge through it.

I surmise that our difference on what is required for testimony-based knowledge
goes with her viewing justification as essential for knowledge. For justified testimony-
based belief, I have proposed a weaker version of her weakest requirement: merely
having justification (1997b) for taking the attester to be credible. It is unclear when
in human development it is appropriate to speak of a child as justified in believing
testimony; but it seems characteristic of this stage that when it is reached we may also
speak of the child as *unjustified* and may criticize—or at least critically correct—the
child. *Critically* correcting a child is not entailed by simple correction; we correct
children when they are only just beginning to speak and are being simply *trained*.
Training does not entail that the erring trainee is justified *or* unjustified. In the case
of knowledge, it seems that children normally get it from adults as soon as they can
form beliefs, which may be at the earliest stages of linguistic comprehension or
before. It surely seems earlier than the age of justification.

If, as I hold, knowledge is possible without justification even for those capable
of justification, there is further reason not to require the strong normative trust
Fricker specifies. I cannot argue for this possibility here (as I have in 2003, chap. 8),
but I agree that typically in receiving testimony from adults we do have justification
regarding their credibility. I doubt that in every such case in which we acquire
testimony-based knowledge, we are justified in believing anything as strong as "not
easily would the person assert that p without knowing it" (where 'not easily' pre-
sumably means 'not likely'). We may also wonder whether this condition is strong

enough. Could I know that *p* from someone's attesting to it if I *needed* a credibility belief but could have one weak enough to allow for a significant chance of error? Why is the requirement not more like justification for believing the attester *would* not say that *p* without knowing it? This requirement would block Lackey's purported counterexamples of the kind described, but it may be too strong.

There is much more to say on both the principles Fricker considers. At stake is the basic question whether testimony is transmissional or generative, and, if so, how. Another is the conditions the recipient must meet for acquisition of testimony-based knowledge. Still another is what kinds of trust are required for such acquisition. The similarities and differences between our views on these questions deserve continued examination.

Internalism and Externalism in Epistemology and Semantics

There are many kinds of internalism and externalism. I have defended an epistemological internalism regarding justification and an epistemological externalism regarding knowledge. I have written little on internalism versus externalism in the theory of content and meaning, but have suggested that a certain kind of semantic externalism is compatible with my internalism about justification. This suggestion partly motivates a challenging critique by Timothy Williamson. My response will substantially extend the position in question (which largely accounts for its length), in part by proposing a preliminary theory of belief content.

Justification, Reference, and Content

On my internalism about justification, one's justification for holding a belief entails having a kind of introspective or reflectional access to justifying grounds. Why should this seem inconsistent with a plausible content externalism? For Williamson,

> According to content externalism, two internal duplicates may differ in what they believe . . . Oscar on Earth in 1750 believes truly that there are pools of water . . . not . . . that there are pools of twater . . . a liquid on counterfactual Twin Earth with the same superficial characteristics as water but an utterly different underlying nature, for in Oscar's world there are no pools of twater . . . on counterfactual Twin Earth, Oscar's duplicate Twin Oscar believes truly that there are pools of twater . . . [but not] that there are pools of water, even though he is in exactly the same internal states. . . . (107)

The paper Williamson is considering might have done more to indicate that I do not unqualifiedly accept this description. What I said (in part) was that

> there is some proposition (one with "narrow content," in some sense) that we are each justified in believing, though only one of us knows the external proposition that there is water in the glass. In a way, this is a demon world solution: what we know can be systematically cut off from what we have justification for believing. But it leaves intact the point that an epistemological internalist might be a kind of content externalist. (2001b: 32)

Since I did not specify the content externalism in question (or cite an earlier paper that did do so in outline ["Mental Causation," 1993d]), I can see why Williamson considers it "legitimate to assume the content-externalist description of the case in discussing Audi's claim." As to my reference to a kind of narrow content common to both Oscars, he says, "Such contents remain the pious hope of some internalists . . . even if Oscar and Twin Oscar do have some justified beliefs in common, the externalist argument was that since they do not have *all* their justified beliefs in common, justified belief does not supervene on internal states" (108).

Much must be said in response. First, it is crucial to distinguish conditions for *justification* for holding a belief that *p*—should one hold it—from conditions for *believing p* at all. (Justification for holding a belief, as where one has amply sufficient evidence for its truth but has not considered this evidence or had any other occasion to come to believe the proposition, does not entail actually believing it). Second, Williamson does not specify the sense of 'internal' crucial for the externalist. Even if there is no narrow content in some sense he has in mind, what *justifies* either belief can still be internal in *my* sense, that is, accessible in a certain way to introspection or reflection. Suppose our perceptions (e.g., seeing paper) are accessible to us despite their external content (something Williamson would grant). They could then serve as internally accessible justifiers of beliefs having external content. As this case indicates, the consistency of justificatory internalism with a kind of content externalism does not preclude that even what justifies a belief has a kind of external content. I do not consider Williamson committed to denying either of these points, but some of what he says might suggest otherwise. In any case, there remain theoretically interesting differences between us.

Suppose that, from qualitatively identical experiences of drinking and being refreshed, each Oscar believes something apparently expressible by 'Water is refreshing'. Can the two believe the same proposition here? Their beliefs are, after all, *about* different things (though they cannot identify any chemical differences) and are properly expressed by sentences with different meanings. I can agree with externalists who hold that meanings are social and not just "in the head." But this vague claim leaves much in need of clarification, including the relation of meaning to content. Let me begin the task of clarifying the notion of content by noting that in the epistemology of perception it is standard to distinguish between *de re* and *de dicto* belief attributions. To clarify the difference (which Williamson does not address[6]), I have distinguished between *objectual* and *propositional* beliefs. Compare:

(1) believing an object, say, a bush, to have a property, such as being a bear;
and
(2) believing a *proposition*, say *that* the bush before me is a bear.[7]

The paper under discussion presupposes that the subject is *propositional* ("de dicto") belief: the kind we commonly ascribe using that-clauses whose content, expressed by wording with the form of 'that *p*', indicates *what* the person believes. This is perhaps the clearest and most common sense of 'content', though what follows will refine my earlier terminology. In any case, it is uncontroversial that we cannot have an objectual belief, such as (1), unless there *is* an object that we take

to have a property. Plainly such beliefs do not supervene on internal states, and for these I agree that "justified belief does not supervene on internal states."[8]

What of propositional belief attributions? We do not usually assertively use locutions of the form 'S believes that the x before S is F" unless we presuppose that there *is* an x that S believes to be F. Given this presupposition, belief attribution (2) would *also* imply that there is a bush; but whereas the position of 'a bush' in (1) is referential and *also* transparent with respect to substitution, in (2) it is referential but *not* transparent. Substitutivity fails: if the bush is the shrub hiding a bomb, I need not, in holding the belief specified by (2), believe that there is a shrub hiding a bomb before me. Still, (1) has one natural reading on which it is not *wholly* transparent. Its content (on this natural reading) is roughly this: there is a bush such that S believes, of it, that *it* is a bear. The second occurrence of 'it' is opaque and substitutivity will again fail. Without information about S that may be quite difficult to obtain, we cannot formulate a proposition about the bush that reflects a description under which S believes that it is a bear; and, if we find one, substitution of a co-referential expression need not preserve the truth of the resulting belief-ascription.[9]

In the light of these distinctions, it is clear that Oscar believes water *to be* refreshing (believes, *of* water, that it is refreshing) and Twin Oscar—"Toscar"— believes this of the counterpart of water, "twater." These beliefs are about different things, hence are naturally considered different beliefs. They are also about something external and hence have external (and wide) content in what we might call the *referential sense* of 'content'. It is also natural, however, to say that both believe the *same thing* about different objects. Is there, then, some propositional belief that both also hold, or at least some content—call it *predicative content*—that both ascribe to the object in question? One might think there *must* be some proposition they both believe. How else could they believe the same thing about entities that differ only in ways they are not aware of? This same thing believed cannot, however, be the singular proposition that *water* is refreshing—more perspicuously, they cannot both believe water to be refreshing—since only one of them has a belief about water; and we shall soon see that there need be no propositional belief they share.

This view of property-ascriptive beliefs preserves their external, referential element, but leaves open what propositional beliefs the Oscars have, and, if they do have such beliefs, in what sense these are *about* the things to which the properties are ascribed.[10] What must be emphatically added is that if we call propositional beliefs such as that there is a bush before me, understood to presuppose there *being* one before me, *beliefs* as opposed to (nondoxastic) property ascriptions—we cannot take beliefs to be psychological properties in the narrow sense: the sense in which their possession at a time does not entail the existence, at that time, of any contingent object outside the believer's mind. Only in this narrow sense of 'psychological' does my internalism imply that psychological duplicates are alike in their justificatory resources. Clearly, belief-locutions have importantly different forms. Content-locutions also vary significantly, as will soon be evident.[11]

Meaning, Content, and "Aboutness"

Given these background points, let us consider belief content further. If this is taken to be a matter of the *meaning* of a sentence that the believer may naturally use to

express the belief, one would consider the content of Oscar's water-referring belief wide. One might even think that there *is* no narrow content, since meanings are not in the head and such semantic content is socially determined. Thus, the content we must attribute to beliefs expressed by the sentences so conceived is at least in part socially determined. We should distinguish, however, between the meaning of what one says and what one means to say—more accurately, between the meaning of the sentence one utters in saying something and what one means (intends) to say in uttering it. The contrast is, roughly, between semantic and intentional meaning. The former is social in an important sense; the latter seems individual and, in a certain way, determined by what is "internal" to the subject.

Supposing the distinction between semantic and intentional content is sound, how, in ascertaining belief content, can we tell which (if either) is in question? We *cannot*, as in some interpretations of Twin Earth, simply consider the subject's environment and the semantic meaning of the 'that' clause the person would most readily use to express a belief such as 'that water is refreshing'. We often use propositional locutions in implicitly attributing an *objectual* belief. Most commonly, when we speak or write, it is about things we are referring to *of* which we have beliefs and *about* which we want to express or convey information. I tell you there is at last a sunny day; you say that you are happy about it, since you must fly today, and so on. In making these references, we presuppose that the speakers are in contact (presumably causal contact) with what they are referring to and are talking *about* it in a sense of 'about' entailing its existence. Call this the *referential presupposition* of propositional belief attributions.

We also tend to presuppose that typically native speakers say what they mean and believe what they say. Call this the *coincidence presupposition*, since it expresses our very common presupposition that the intentional content of what is believed coincides with the content of what is said taken at face value. It may be that, almost as often, we make these presuppositions about our own thinking insofar as our thoughts are expressible in the ways just illustrated. There are, however, interesting exceptions to both presupposition principles. We can talk about imaginary entities, or even impossible ones, such as round squares. Here we do not presuppose reference in the ordinary sense entailing the existence of the apparent referent. The sense of 'about' might be called *topical* to distinguish it from the more common, referential sense in which what is talked about is not merely a kind of subject matter but something that exists. More important for our purposes, the coincidence presupposition plainly does not hold when, as is common, we take one person or thing for another. Someone might be described as, for instance, 'believing that I was my brother' or 'thinking, in the moonlight, that the bush was a bear'. Here propositional locutions are loosely used in place of accurate ones, such as 'believed me to be my brother' or 'took the bush to be a bear'.

Is the suggestion, then, that the coincidence presupposition fails for Oscar (transported to Twin Earth) when he looks at twater in a glass and sincerely says that drinking it would be refreshing? Certainly he believes the twater in the glass *to be* refreshing. He does not believe *that* the twater in the glass is refreshing, where this is a propositional belief and not assimilated to the former property-ascription. But, since he is talking *about* the twater and is saying something about *it*, why doesn't he have a *propositional* belief with wide content?

Recall the fraternal confusion example. What proposition (if any) the person who takes me for my brother believes will depend on how he thinks of me on the occasion. Perhaps, for example, my clothing and gait are salient but he does not see my face; he might then believe that the man walking past in a tweed jacket is AA. This indicates that the proposition believed may well not be one formulable on the basis of the coincidence presupposition. The observer may have some working criterion, *c*, for identifying my brother, may take me to satisfy it, and may believe the proposition that the guy with *c* is AA. The person may, however, just strongly associate some properties of mine that stand out (my gait, e.g.) with my brother and believe that the guy with these is AA. Let me stress, however, that as natural as it is to use propositional locutions in such cases, the most perspicuous description of them may be in terms of property ascriptions. Then, what is ascribed to the thing in question may or may not be true *of* it, but I suspect there need be no proposition, in the sense of an abstract, truth-valued content of thought of the kind that serves as an object of propositional belief.[12]

What has been said so far does not indicate how to construe cases in which a belief is expressed by a sentence employing a referring expression to which nothing answers, such as 'the attacking bear'. There is still something believed and it is natural to say that it is about something. Let us explore such "aboutness," in relation to reference and content. We may ask not only what a person's belief (or knowledge) is about but also what sentences, propositions, discussions, and many other things are about. My main concern is beliefs. Objectual beliefs are by their very nature *of* something in a sense of 'of' entailing their being about it. We may call this a *referential sense* of 'about' if we conceive reference in an ordinary way. For that sense encompasses the topical sense, in which we can be talking about the ghost of Hamlet's father, though he is fictional, and even about what he might have thought about his brother Claudius, which is in a way doubly fictional. We can even be talking about round squares, as in saying they are philosophers' examples of the impossible.

For a certain *basic* kind of objectual belief, however, namely perceptual ones, there must be the object(s) the belief is about and (in my view) the believer must even have some causal connection with what the belief is about such that it is in part in virtue of this connection that the belief *is* about that. In one respect, this kind of objectual belief is the clearest case both of reference and of aboutness. But as we have seen, it is also important to account for what is *believed* about something—whether it exists or not. Call this *contentual aboutness*. The (or a) contrast between wide and narrow content can be raised for this kind of aboutness as for propositional beliefs. Just as, when an objectual belief attribution is made, we should determine, in the light of the particular situation of belief, how the person conceives what the belief is about, so when a propositional belief attribution is made we should determine what is believed about something in the light of particular variables. Let me illustrate.

Suppose that (with no inkling of it) Oscar is transported to Twin Earth. If, having drunk some twater, he (sincerely) says, 'Ah, water is so refreshing', how should we tell what he means to say? We cannot say that it *must* be something false since he must mean what is standardly expressed in *that* world by the declarative sentence he uses as applied to the watery substance he drank. If we ask what he means by

'water', he may say any of a number of things, some ostensive, some not. This is important; people often err in saying what they mean, and in any case cannot in general be expected to give *definitions* of their referring expressions, as opposed to indications of how they identify what they refer to. Thus, 'the sort of thing we all drink when thirsty' might or might not tell us what he means. Suppose it does. Whom does 'we' include? His thinking of his earthly peers would anchor his reference to water. But he might have said 'this kind of liquid, which everyone around *here* drinks', in which case he would be correct only if twater has the relevant refreshing quality. In these cases his reference is, in a certain way, socially determined.[13] By contrast, his inferential and linguistic behavior is best explained on the hypothesis that he is conceiving water in terms of its perceptible properties acquired through our common uses of it (perhaps of 'water' too). Then what he believes *may* be both narrow in content and true: that a certain perceptible kind of thing commonly drunk is refreshing. Here he has generalized and is connecting sets of properties with each other in a way that does not entail the existence of any objects.

Beliefs of, Beliefs about, and Narrow Content

A general point here is that there is no necessity about whether any conventional or social anchoring notion figures in what one believes *about* what one is drinking, in the contentual sense of 'about'. I grant that just as there is a presumption that people are talking about what their words, understood conventionally in the context of utterance, refer to, there is a presumption that people believe, about what they refer to, what one would think they believe about it if one took their utterance at face value—that is, in the conventional sense (if there is a single one). What one would think they believe about it *tends* to be what they would (sincerely) say in expressing this predicative content of what they believe. These presumptions go with what I have called the referential and coincidence presuppositions. But all three patterns fail in important cases. We cannot always reliably infer people's beliefs from even their sincere utterances.

Supposing, then, that there is such a thing as narrow content, how is it to be conceived and determined? These are large questions I cannot fully answer, but several brief comments can provide background for the epistemological position I will shortly outline. I assume that in the natural order of human development, experience gives us a familiarity with many properties of things, and that these properties figure centrally in perception, thought (including belief), and language. We experience colors and shapes, learn to ascribe some of the color and shape properties experienced, develop conceptions of them, and master predicates that express them. We identify objects by their properties; and much of what we say—and believe—is essentially a matter of a certain kind of property-ascription.

I have spoken of the natural order of human development on the assumption that thought, belief, and language causally depend on experience, presumably experience of an external social world. But (to focus just on cognitive development in relation to belief) causal genetic dependence on the external does not entail contentual dependence on it. We can indicate what a belief is about in the referential sense by citing properties (perhaps including relational ones) by which one can identify the thing in

question.[14] We can indicate what we believe about something by citing the properties that we ascribe to it in virtue of having the belief. Call the theory I am sketching here for the latter case a *property-ascription theory of narrow content*. The theory does not imply that *every* property-ascription embodied in a belief indicates narrow content, but the kind that does is indicated by some of my examples.

Recall the Oscars. They have different objectual beliefs if Oscar believes water to be refreshing and Toscar believes twater to be refreshing; but why not take their beliefs to have the same ascriptive content *provided* each believes, of the stuff in question, something to the effect that it is refreshing? This is highly plausible if they can discern no difference between water and twater. To be sure, it may be difficult to determine just what properties figure in their conceptions of the stuff to which they make the ascriptions; but I see no reason why their having different objectual beliefs—in virtue of ascribing properties to different *things*—must imply that they differ in their *property*-ascriptions and thereby in a kind of narrow content of their propositional beliefs. One might object that 'refreshing' itself is wide, but it need not be. They may have come to understand this property in different settings, but both can still ascribe that same property to objects.[15]

Indeed, on a strict understanding of 'psychological duplicates', I find it difficult to see how they could differ in *those* "beliefs," that is, in what they believe about the referent in question, a matter of ascriptive content. (Since Williamson does not specify what he means by internal duplicates, I do not know that he would accept this suggestion; if [like Burge] he means roughly *identical in physical and functional properties*, I have not claimed that duplicates in this sense have identical justificatory resources.) They will presumably coincide both in how they *conceive* what they are talking *about* and in the ascriptive content of what they assert of it. This is perfectly compatible with holding that the usual attributions of beliefs by way of names or definite descriptions of particular things are wide in being referentially about something. But that does not preclude the (objectual) beliefs in question having narrow content, in virtue of property-ascriptions essential in them.

The Scope of Internally Grounded Justification

Given the suggested property-ascription theory of narrow content, it should be plain why I hold that a *kind* of content externalism is consistent with my epistemological internalism. Objectual beliefs must be specified partly in terms of the objects they are about, in an external sense of 'about'; plainly, then, where there are two objects of the relevant property ascriptions, there are two such beliefs. There is, however, a plausible hypothesis suggested by my position that applies even to objectual beliefs: what *justifies* our ascribing a property to something (or to some apparent thing) is internal—though the conditions for actually believing, *of* an external thing, that *it* has that property are partly external. Consider hallucination. If, awaking in the dark acutely thirsty, I have an experience qualitatively just like seeing a glass of water on the night table, this experience justifies me in believing that there is one—and in believing the table to have a glass on it. This *sensory* experience is qualitatively just like its *perceptual* counterpart and is a justification for the same beliefs as well as for an objectual belief missing in the first case: believing *the glass* to be on the table.[16]

I can now explain how I am replying to Williamson's description of a fallback position, that is, "that although internal duplicates may differ in what *gets justified*, they do not differ in what *does the justifying* . . . Oscar's internal state would justify him in believing that there are pools of water but not in believing that there are pools of twater" (109). I see no retreat. My position is that the Oscars need not differ in what they believe *about* the liquid in question; they do differ in their objectual beliefs; and they need not differ in the *kinds* of accessible elements that serve as justifying grounds, which will be internal in any case.[17] The crucial point for my epistemology is that what justifies be internally accessible, not that people with equivalent justificatory resources need have the same *beliefs*—nor, as we have seen, is there only one kind of belief in question.

Williamson would likely not be satisfied with this response. He continues:

> Could Audi restrict his claims to propositions that both subjects can grasp? . . . We can add . . . that a traveler once showed Oscar a tiny phial of twater and told him (truly) that . . . it was called "twater," had similar superficial characteristics to water but . . . was utterly different (in unspecified ways), and occurred only in minute quantities, not pools (in this world). . . . Similarly, a traveler once showed Twin Oscar a tiny phial of water and told him (truly) that the liquid in it was called (confusingly for us) "twater." . . . Although Oscar *associates twater with different descriptions from those that he associates with water, and likewise for Twin Oscar,* externalists about content will typically hold that in such circumstances Oscar can grasp the proposition (false in this world) that there are pools of twater while Twin Oscar can grasp the proposition (false in that world) that there are pools of water. After all, *we share the belief that there are pools of water with people whose beliefs about water are quite different from ours.* Nevertheless, Oscar is justified in believing that there are pools of water but not in believing that there are pools of twater, while Twin Oscar is justified in believing that there are pools of twater but not in believing that there are pools of water. (109–10, italics mine)

Granted, both can understand the propositions introduced by the travelers. But the first problem the scenario poses is whether they remain psychological duplicates in my sense. The *properties* in virtue of which they understand these propositions include some that individuate either the informant, or the substance described, or both. Williamson and I agree that the *reference* of 'water' and 'twater' is normally to different substances for the two; but this alone does not suffice for their differing in propositional beliefs in the way he suggests. Can some of their water- and twater-beliefs not also differ in narrow content? This might occur in terms of properties that figure in their belief-formation and cognitive dispositions at the time (e.g., properties by which they judge what constitutes being water, being refreshing, or being what a certain informant pointed out); these properties will differ in a way that implies a difference *both* in the totality of their purely psychological properties and in what they are justified in believing. The difference might be between recalling the two informants in terms of different phenomenal properties instantiated in the respective perceptual circumstances.

I do not deny that there is—for any view of belief content—the problem of articulating criteria for determining *what* proposition each believes, or what properties each ascribes to the object in question, on a given occasion. Are their "associations" dominant

in a given case, or is the traveler's description dominant? Ascriptive content must not be too tightly connected to the external object the belief is about; for there are indefinitely many ways to "connect" (even causally) with such an object, and ascribing properties to it does not require any particular way. Believing, of some x, that *it* is F allows wide latitude concerning the range of the believer's possible property-ascriptions—other than *being F*—to x. In any case, none of the plausible ways of determining belief content undermines the view that what justifies beliefs is internally accessible to the believer.

Does Knowledge Entail Justification?

If justification is internally grounded, and—even apart from that—if knowledge does not entail justification, then the traditional view that justification is necessary for knowledge must be qualified. I have argued that memory is a domain that may exhibit knowledge without justification.[18] Williamson challenges this. Noting that, from learning it years ago, we can know that Trincomalee is in Sri Lanka, yet might have no internal grounds for the belief, he says:

> Since my belief that Trincomalee is in Sri Lanka is not justified. . . . I do 'deserve criticism (from the point of view of the effort to reach truth and avoid falsehood)' for believing that Trincomalee is in Sri Lanka, even though I know that Trincomalee is in Sri Lanka. Such claims seem quite implausible. One ought to give up a belief once one realizes that one is not in the right in holding it. But it would be silly for me, having read Audi, to *give up* my belief that Trincomalee is in Sri Lanka. That would be to give up some of my knowledge. (111)

The case is underdescribed. Williamson is supposing that one cannot remember learning this or provide premises for it ("We cannot remember how we acquired the information, and it may be relatively isolated"). But, first, one would normally have good inductive grounds to support this as the *kind* of thing one would not believe if one had not learned it; and, second, on my view a memorial sense of having believed a proposition is enough to justify retaining belief of it (barring defeat). Third, he is imagining one's not *finding* justification of the kind I require and thereby being justified in giving up the belief. But I would deny that in this case not finding implies not having. For one thing, the gaze of introspection may alter the landscape it is supposed to reveal.

   I would also argue that here "giving up a belief" suggests taking some kind of action that may not be justified by one's total evidence; and, even apart from whether giving up a belief is a kind of action, it is not justified every time one justifiedly thinks one lacks justification for the belief. As to the implied argument that we should not give up the cognition because we would be giving up knowledge, the situation sketched is one of *second-order* reflection; hence we would be giving up a cognition we would *describe* as knowledge. If we do not assume that a cognition we are evaluating is knowledge, then we *can* be justified in "giving it up," even if it *is* knowledge. Knowledge is often defeasible in this way.

   With these points in mind, consider Williamson's ingenious example:

> S knows *p* without being justified in believing *p* . . . at the same time S has a justified belief . . . $p \rightarrow q$ without knowing $p \rightarrow q$. . . . Now S applies modus ponens. . . . What is the status of her belief in *q*? It seems to be good in some way, since it is held on the

basis of competent deduction from premises, each of which was either known or believed with justification. However, S does not know *q*, for S does not know one of the premises . . . Equally, her belief in *q* is not justified, for S lacks justification for one of the premises. . . . Audi seems to be missing an epistemological category. (111–12)

I again find the case underdescribed. We agree that she does not know that *q*, but is her knowledge that *p*, which is not a case of justified belief, also one of patently *un*justified belief? If so, her belief that *q* might be *excusable* but not justified. Suppose, on the other hand, that the belief that *q* has support from some elements but not *enough* support to be unqualifiedly justified; it may, then, have some degree of justification. Another possibility is that she believes *q* is justified, though without grounds good enough to imply that. She can, however, "speak for it." Here we might consider her believing *q* understandable, perhaps expectable and minimally rational. We might also consider some such beliefs *creditable*, though not quite justified. Moreover, my epistemology accommodates objective probabilities and objective evidence. These might be estimated on the basis of details concerning the grounds of her knowledge and of her justification. Thus, I do not see why, as the case is described in detail, my position does not make it quite possible to assign some appropriately positive status, say, credibility, having *some* degree of justification, or having high probability.

Accessibility, Justificatory Practices, and Normative Identity

The notion of accessibility admits more than one interpretation. In holding that the grounds of justification are internally accessible, I have not denied that there is a sense in which much that is external is accessible in an importantly similar way. Here Williamson and I are closer than may be apparent. Let us first explore this passage:

Have I internal access, in Audi's sense, to the presence of the computer screen before me? . . . [I] have it 'in consciousness' in some sense. However, [Audi] . . . takes for granted that the contents of consciousness are confined to more traditional items, such as mental images. On his view, an experience as of a computer screen before me may be 'in consciousness,' but not the computer screen itself . . . [and] on Audi's view, I lack internal access to Trincomalee's being in Sri Lanka. (113)

The screen is, I take it, in consciousness in virtue of perceptual consciousness of it. My view *would* be incompatible with this if I held that what is in consciousness is only something like a sense-datum replica of the screen, but I do not (see, e.g., 2003, chap. 1).

On my view, the screen is in consciousness in virtue of being directly seen, but what is in consciousness as a phenomenal element is a sensory state that is in fact perceptually grounded in seeing the screen. Seeing the screen entails the perceiver's having a physical relation to it, and this *relation* is not in consciousness. But I am directly aware of a sensory state that *is* in fact an experience of a screen. This awareness is not only not mediated by any object of (ontologically) prior acquaintance (as on a sense-datum theory); it is also not mediated by any inferential ground. I would normally neither believe (nor disbelieve) I have it (for reasons provided in my 1994), nor need I even think of it or try to focus on it as a state separable from

seeing the screen. We can thus do justice both to the internal grounds of the justi-fication of the belief and to the sense in which the physical object is directly per-ceived and is in consciousness when it is so perceived.

The memorial case is different. Williamson considers me committed to the view that "although I consciously remember that Trincomalee is in Sri Lanka, what is in my consciousness . . . is not that Trincomalee is in Sri Lanka, but only that it seems to me that Trincomalee is in Sri Lanka (or that some related narrow con-tent seems to me to be the case)" (113). But on my view, where we have a normal memory belief that *p*, we would not normally even *believe* the proposition that it *seems* that *p*; it is the proposition that *p* that is apparently remembered and in that sense in the mind. That Trincomalee is in Sri Lanka may itself be internally acces-sible (being "in mind") in much the way a computer screen is. My knowing this proposition need not be based on a memorial seeming; but even if such a seeming should be needed for justified memory belief, the seeming need not be the *object* of any belief. There is an analogy to the perceptual case: my (propositional) belief is in consciousness (when occurrent); if it constitutes knowledge, I thus have direct access to a belief that in fact *is* knowledge. In this way, I can have an external fact (or obtaining state of affairs) *in mind*.

Concerning accessibility, then, I also do not see that we need differ on another point: that "internalism . . . risks losing touch with usability in the everyday con-texts. . . . For justification is typically a *social* practice: we try to justify our beliefs to others in response to their challenges. But if what I use in justifying my belief to you is internal to me, how much use will it be to you?" (115–16). I agree that if, for exam-ple, you ask how I know the screen is lighted, I may say that I see its light on; it would be misleading to say that I have a visual experience of its luminosity. But it does not follow that some such experience is not basic in my justification or that I may not be pushed to cite it if the usual answer does not suffice. My point against reliabilism about justification was that if the reliability of the grounding of a belief is what justi-fies it, one could have a justified belief for which one might be unable to say *any-thing* on the basis of an adequate self-understanding—neither citing something like believing a publicly confirmable evidential proposition nor even saying one has an intuitive sense of the truth of *p* (for nothing like that is a condition for merely reli-ably grounded belief, as argued in some of my writings referred to above). One might then not be able to engage in the practice of justification, which my paper systemat-ically connects (as Williamson seems happy to do) with the property of justifiedness.

Note, too, that, in my view, you *can* be aware of my beliefs and even my sen-sory and other impressions through my testimony or otherwise. The awareness may even be noninferential, as I have argued it is when you have testimony-based knowl-edge of what I believe. Moreover, what you come to know through my testimony can be in your consciousness and directly accessible to you in the same sense in which what has external content can be.

My final concern here is to forestall a misunderstanding of my internalism. Suppose we

> [C]all situations 'indiscriminable' if and only if any difference between them is not
> accessible in the relevant way to the subject, and 'normatively identical' if and only
> if they are the same in the relevant normative respect. Then Audi's approach commits

him to [the principle that] . . . [i]ndiscriminable situations are normatively identi-
cal. . . . Imagine a sorites series . . . each indiscriminable from its immediate neigh-
bors, where the first and last members are very different. . . . normative identity, like
exact sameness . . . and unlike indiscriminability, is transitive. Therefore the first
and last members of the series are normatively identical. (119–20)

I reject this principle, and my view does not imply it. First, although Cartesian sce-
narios are internally like ordinary perceptual cases, this is not because they are indis-
criminable from those but because (above all) the relevant phenomenal properties
are the *same*: in both cases, for example, one has the visual impressions character-
istic of seeing a screen. Second, I have long defended the view that a belief is justi-
fied by a ground for it only if it is, in a partly causal sense, based on that ground.[19]
There is no reason to expect indiscriminability to imply sameness of causal power.
Indeed, there is some reason to think that qualitatively different psychological prop-
erties *differ* in causal power.[20]

   If the distinctions and the theoretical extensions of my epistemology set out
above are sound, then my position on justification and knowledge is not vulnerable
to Williamson's objections. My replies to some of those objections, moreover, show
that, skepticism aside, we agree on many points: that there is noninferential per-
ceptual knowledge, that there is an associated kind of direct awareness of external
objects, that there is a sense of 'content' in which certain beliefs have external con-
tent in a sense entailing the existence of objects they are about, that there is a social
practice of justifying beliefs, and that (as I would put it) knowledge is a kind of suc-
cess. The differences that remain are well worth further examination.[21]

Belief, Acceptance, and Faith

A central concern of religious epistemology is the status of theistic beliefs and their
relation to the rationality of religious faith. Faith is commonly conceived as a kind
of belief—say, belief *in* God or belief *that* God will do certain things. In a series of
papers on faith, I distinguish faith from belief and argue that there is a kind of faith
whose rationality conditions are less stringent than those appropriate to beliefs with
the same content. I refer here to *propositional faith*, say, faith *that* God is sovereign,
as opposed to *attitudinal faith*, such as faith *in* God. Although propositional faith is
compatible with believing the proposition in question, it does not entail believing
it. In recent work I have called such nondoxastic faith *fiducial* to convey the idea of
a kind of trusting, as in 'I trust that he will meet his obligations'. William Alston
challenges my account of nondoxastic faith but, agreeing with a number of my
points, proposes an original conception of acceptance that he takes to capture the
data that concern us both.

Fiducial Faith versus Belief

In the papers Alston considers, I drew many contrasts between nondoxastic faith
and beliefs with the same propositional content. Despite the amount of detail the
account of fiducial faith contains, Alston is not satisfied. He says I have

told us quite a lot about its properties—how it is similar and dissimilar to belief, what it is and is not compatible or incompatible with, some of the dispositions it does and does not involve, some of what it is associated with, and the relative strengths of various properties. But we are still left in the dark as to what all this is true of. The only substantial hint we get on this is that NDF [nondoxastic faith] is a positive cognitive attitude toward a proposition. (126)

The papers he cites do not use 'fiducial faith', though some draw attention to the locution that partly inspired it, 'trust that'. If his demand is for an ordinary near equivalent of 'nondoxastic faith', I suggest 'trusting that'. There are questions about what the relevant fiducial attitude is, but this is a general problem for philosophical explication. In any case, he thinks nothing can satisfy *all* my conditions for fiducial faith (123). Is that true?

In one place, Alston says that one of my requirements (being incompatible with disbelief) also seems to be in conflict with the person's not having "definitely accepted" *p* and not being intellectually committed to it. If (he wonders) NDF involves no intellectual commitment to *p* and no definite acceptance of *p*, then why should it be incompatible with disbelief that *p* (126)? The briefest way to answer this is to note that I have stressed that faith that *p* is stronger, in at least the convictional dimension, than hope. But even hoping that *p* is inconsistent with believing it false (disbelieving it). As to *definite* acceptance of a proposition—as contrasted with, for instance, accepting it as a working hypothesis—I took that to imply belief. One might think it is possible simultaneously to believe and disbelieve *p*, but I doubt this (for reasons given in my work on self-deception). In any case, neither Alston nor I thinks it possible for a proposition that is the object of faith.

On another aspect of the question, Alston says I do not "forthrightly assert that to be properly said to have faith that p, one must be taking p to enjoy less than ideal support, but his assertion that NDF requires less for justification or rationality than belief (3c) could be read as a backhanded acknowledgment" (128). Granting that *taking* does not entail believing, Alston is at least implying that having (propositional) faith requires having a conception of the quality or strength of the evidence for *p*. I consider this too strong a requirement. For sufficiently sophisticated subjects, there may be a *disposition* to believe some proposition(s) concerning the strength or quality of evidence; but believing, or even just "taking," the evidence to be of a certain strength is not required. If, as seems plausible, faith is not always a response to evidence at all, this point gains support, and the evidential taking requirement Alston here suggests is disconfirmed. Indeed, even if in some cases (propositional) faith is largely a response to evidence, responding to evidence does not require having a sense of how strong it is.

On the nature of belief itself, I find Alston's discussion of my 1972 account instructive. I have since written much on belief and, especially in the light of my systematic distinction between dispositional beliefs and dispositions to believe (1994), I would add many points to the earlier paper. In one case, Alston may put unnecessary distance between our accounts. He says, "If xBp [x believes that p], then if x considers whether it is the case that p, x will tend to feel it to be the case that p, with one or another degree of confidence" (131). Perhaps this normally holds, but I doubt that it should play the major role Alston gives it. If the weaker

phenomenological conditions I originally specified hold (together with all the others), it is not clear to me that this would also have to hold. Much would depend on what it is to feel a proposition to be true, for instance what this adds to a tendency to "say, assert, insist, affirm, avow, or the like, that p" taken together with other conditions specified in my earlier paper.

About one element of Alston's formulation, I have no doubt: the reference to degrees of confidence. One reason I contrast fiducial faith with "flat-out belief" more than with belief simpliciter is that there may be an understanding of belief on which a certain kind of belief that lacks a high degree of confidence *would* be a candidate to represent fiducial faith. I have not endorsed this suggestion, but have indicated that much of what I say about fiducial faith could be retained for a suitably restricted notion of weak belief.[22] One important point would be that the degree of evidence needed for the rationality of belief is (other things equal) inversely proportional to its confidence level. The strength of associated motivation, however, need not vary with belief strength in this way; it may be very great even where the confidence level is the minimum for faith that *p*, as distinct from hope that *p*.

Acceptance as a Candidate to Clarify Fiducial Faith

As to what constitutes fiducial faith, Alston and I largely agree. But I find the notion of *trusting that* a better everyday locution for intuitively anchoring the notion than (apparently) does Alston (*apparently*, because he does not consider it). He regards 'acceptance' as a better intuitive anchor:

> I find the voluntary character of the act of acceptance to be the best way of giving an initial idea of it. The act of acceptance, unlike a state of belief, is the adoption, the *taking on* of a positive attitude toward a proposition . . . a mental act. . . . But when we come to say just what positive attitude to a proposition is adopted when one accepts it, we are back to the pervasive similarity of acceptance and belief . . . accepting that p is both a complex dispositional state markedly similar to believing that p, but distinguished from it by the fact that it issues from acceptance in the other sense, a mental act. (132)

One could, then, consider a theological proposition and then accept it and thereby pass into a state of acceptance of it that is an instance of nondoxastic faith. Call the posited act *behavioral acceptance* and the resulting state *cognitive acceptance*. Alston gives a useful example:

> Consider an army general . . . facing enemy forces. . . . He needs to proceed on some assumption as to the disposition of those forces. His scouts give some information about this but not nearly enough to make any such assumption obviously true. . . . He *accepts* the hypothesis that seems to him the most likely. . . . He uses this as a basis for disposing his forces in the way that seems most likely to be effective, even though he is far from believing that this is the case. (133)

There are acts of acceptance, as the military example shows in noting the decision to use a hypothesis as a basis of action. But Alston requires a voluntary act whose result is entering a *cognitive* (truth-valued) state. Granted that we can *cause* the

formation of such states indirectly, say, by exposing ourselves to certain external stimuli (or brain manipulation). But can we do this at will? I doubt it. Even if we can, is this what behavioral acceptance is?

If you tell me something controversial and I accept what you say, have I performed an act of forming a positive cognitive attitude? Or does 'accept' here designate something like this: (1) my *not* resisting, say by asking for evidence, and (2) my cognitive system's responding in my forming the appropriate attitude—which, in this case, would normally be *belief*? 'He accepted what I said', for instance, normally implies his believing it. By contrast, our general need not pass into a state of cognitive acceptance of the proposition in question; he may simply accept it *as a working assumption*, which is mainly a matter of deciding to act in certain ways.[23]

It appears to me, then, that behavioral acceptance is not a good candidate to yield a cognitive state, and cognitive acceptance is not a good candidate to replace fiducial faith as the kind of nondoxastic faith both Alston and I consider important and insufficiently emphasized in the literature. I grant that some cases of fiducial faith may also be cases of cognitive acceptance; but the latter typically implies belief.

The term 'accept', moreover, which evokes a sense of contrast with rejection, wrongly suggests that forming the fiducial attitude in question *requires* some voluntary act. My point that a person can have fiducial faith that *p* without having definitely accepted *p* is directed toward the absence of a behavioral acceptance requirement, not toward the claim that in having such faith one cognitively accepts *p*. If Alston's point is that fiducial faith is a case of cognitive acceptance, I would mainly reply that cognitive acceptance is too strong a notion and that acceptance *as*, say, a guiding principle is too weak. The general, for instance, need not have faith that the enemy is in the relevant place.

To be sure, if I have faith that God loves us, it would be at best misleading to say that I do not accept that proposition. This may be mainly because 'do not accept' strongly suggests having *considered* and rejected, or at least having considered and not come to believe, a proposition. There is, to be sure, the locution 'accepts on faith'. We can also speak of things people accept as *part* of their faith. In these cases, however, 'accept' usually implies belief. It does not imply that the cognitive attitudes in question have been voluntarily adopted or even adopted as a result of voluntary acts. Supposing, then, that there is a kind of cognitive acceptance that is equivalent to nondoxastic propositional faith, it may also be equivalent to fiducial faith, a kind of trusting that the proposition is true. But these latter two terms are, in my judgment, more appropriate, in part because (1) neither can be used to designate an act or even an event; (2) neither of the relevant attitudes must be formed as a *result* of a voluntary act (as at least typically holds in the scheme Alston is proposing); and (3) neither is as close to implying belief as is acceptance understood cognitively.

Nothing I have said implies that acceptance—behavioral and cognitive—is unimportant for understanding religious commitment. Here Alston says much of value in this paper. My view is instead that a theory of acceptance should be placed side by side with a theory of fiducial faith. It is not an adequate replacement for it.

Notes

1. The notion of structural justification (explicated in my 1991 paper of that title) is illustrated by my justification for believing that either there is no lion in this room or this volume contains fewer than twenty essays. Before raising the possibility, I had no belief on the matter, but my cognitive structure contained background information and percepts that give me a basis for forming and justifiedly holding the belief in question.

2. The notion of the a priori I have in mind is one that treats the base case as self-evidence in the sense explicated in my 1999b. This contrasts, for reasons given in my forthcoming c, with the notion articulated by McKinsey (1994) and others, including Kitcher (1983).

3. This formulation is from my paper on the a priori authority of testimony (2004a). Related discussion of the issue BonJour is addressing is found in Graham (2007).

4. The case is from Lackey 1999. She presents others, but some of what I believe should be said about those may be implicit in this paper; further discussion must await another occasion.

5. Lackey has actually carried the case further, adding the requirement that the teacher by policy would not teach the proposition he disbelieves *unless* he thought there was good evidence for it. I doubt that this works, for reasons given in my forthcoming d. But this is a reliabilist condition and would not be welcome to Fricker, whose approach is quite different.

6. He refers, e.g., to "ordinary belief contents" (108) without noting that in different contexts there may ordinarily be (as I illustrate) different kinds of beliefs that correspondingly differ in content.

7. The bear-bush example is from Shakespeare, *A Midsummer Night's Dream* v.i: 21–2: "Or in the night, imagining some fear/How easy is a bush supposed a bear."

8. For extensive discussion of how this failure of supervenience is compatible with a version of content internalism, see McKinsey 1991. An important example McKinsey brings against Putnam's case against internalism is that of *de se* beliefs (147): however alike the Oscars are internally, if either believes that *he himself* has a property, he will have a different belief from the counterpart belief of the other—a point no plausible internalism need deny.

9. The reference of 'it' in locutions of the form of 'such that S believes that it is F' is what McKinsey (1994) calls *anaphoric*, and he takes the content of the indicated belief to be narrow (see esp. 309–15). Plainly, the position of 'it' is opaque; what is crucial is that what S believes is a matter of what properties go in for 'F', which is determined by how S *thinks* of the object at the time in question. This is an internal matter in ways made clearer below. I am reminded here of Hector-Neri Castañeda's idea of the *propositional opacity* of such contexts (which is compatible with the referential transparency of the position of 'it').

10. In my 2003 (e.g., on 51–52), I made the suggestion (not pursued in the paper Williamson is examining) that in many cases objectual belief is not best characterized as a kind of belief at all, but rather as a case of *property ascription*, in a sense implying that we may speak of an attribution true *of* the thing in question but not of believing *something true*.

11. The notion of a narrow psychological property I have sketched is much like that in McKinsey 1991, 155, employing Russell's notion of acquaintance. My distinction between internal and external content also parallels in important ways the one Chalmers (2002a) draws between epistemic and subjunctive intentions. He uses the former to develop a notion of narrow content similar to mine. Space does not permit comparing the two views, but his

paper contains much that supports my epistemological internalism and the related view of content I outline here. For a quite different assessment of narrow content see Stalnaker 1990, which examines some important internalist ideas of Brian Loar's.

12. This rules out singular propositions construed as having the individual of which one believes something as a constituent. But believing those is apparently equivalent to a kind of property-ascription. My suggestion that property-ascriptions are far more common than one would think given the dominance of *de dicto* locutions in belief-attribution is compatible with a plausible cognitive psychology, but that is something I cannot pursue here. For helpful discussion of how propositions are and are not connected with belief, see McKinsey 1994.

13. For an indication how such socially determined reference is achieved, see Chalmers 2002a, on what he calls *semantic deference* (616–18). Compare Burge's view (1982) that "to know and explicate what a person believes *de dicto*, one must typically know something about what he believes *de re*, about what his fellows believe *re re* (and *de dicto*), about the entities they ostend, about what he [his?] and his fellows' words mean, and about what entities fall in the extension of their terms" (112). Particularly in the light of how much is required of anything deserving the name 'explication', I find this not implausible, but it seems not to entail that "the conditions for individuating his [one's] attitude contents—and thus his mental states and events—make reference to the nature of entities in his environment or at least to what his fellows consider to be the nature of those entities" (114). What is sound in the former claim may perhaps be consistent with my overall view, but I do not accept the latter, individuation claim if it is applied (as seems intended here) to what I take to be narrow contents. Some clarifying discussion of this issue is provided by Goldberg 2002.

14. Space does not permit considering indexicals and qualitative identity here. Chalmers 2002a addresses aspects of the problem. Devitt (1990), like Chalmers, thinks that psychology must use narrow content. I find this plausible, but am not here committed to any particular view of how psychological science should best proceed.

15. For helpful discussion supporting the idea that 'refreshing' need not be wide and that property ascriptions may be narrow in a way that confirms my view here, see Pautz 2006. The critical response by Byrne and Tye (2006) raises doubts about some of Pautz's views but does not appear to undermine my position.

16. Burge (1986) rejects the idea that 'we could have the same perceptual representations, whether these are veridical perceptions, misperceptions, or hallucinations . . . our perceptual experience represents *or is about* objects, properties, and relations that are objective" (125, italics added). I agree on the point about perceptual experience and even on the first point *apart from hallucination*—provided we distinguish between perceptual and merely sensory experience, as Burge is apparently not doing here. Perception—e.g., seeing and touching—implies some object perceived; and if to represent is to be "about" in the referential sense apparently in question here, I agree. But then why take hallucinations to be representative at all? I find no argument in this paper that shows the impossibility of having beliefs that are representative in embodying property-ascription yet do not meet the appropriate referential condition. There *is*, to be sure, a problem about what beliefs based on hallucinations are about and just how to describe their content. The beginning of an answer is given in chapter 1 of my 2003.

17. If they are psychological duplicates then, given their equivalent justificatory resources, may we say that Oscar, never having been exposed to twater, has a justification (adequate) *for* believing twater to be refreshing? It is odd to say this since he in fact cannot

have that (referential) belief. But the oddity is pragmatic. We are talking about justificatory resources, and these apply to relevant possible worlds in which S holds the belief. In some uses, moreover, 'justification for believing' *is* used where S cannot form the belief (as with the spouse who cannot believe, despite the good evidence, that the partner is unfaithful). As I have stressed, conditions for believing *p* at all must be distinguished from conditions for justification for believing it if one should.

18. See my "Memorial Justification" (1995b), which has a better example for my case than the one Williamson cites. In replying to Williamson here, I also indicate some of what I would say to Goldman's response (1999) to my internalism about memorial justification.

19. See, e.g., my "Causal Structure of Justification" (1983) and "Belief, Reason, and Inference" (1986b), both reprinted in 1993e.

20. Although I do not hold that only causal powers individuate properties (not all of which seem to me causal at all), I do consider identity of causal powers necessary for property identity. For a plausible understanding of properties that gives causal powers a central role, see Shoemaker (1980).

21. Given the length of this section, I forgo detailed discussion of hallucination. Most of what needs saying is implicit in my response to Williamson above, but I should say that at one point in the paper he examines I was insufficiently clear. On whether the grounds of obligation may be internal while its content is external, I failed to clarify the difference between the ground of a justified *sense* of obligation to A and of an actual obligation to A (which might involve a relation to another person). A justified sense of obligation, like a justified belief, need not be veridical, entailing that one is in fact obligated, whereas an actual obligation to A does entail this and (as Williamson notes) may have an external ground, such as a relation to someone else.

22. In, e.g., my 1995c and 1996, which also contain critical discussion of other aspects of Alston's epistemology.

23. This issue is considered in detail in my "Doxastic Voluntarism" (1999a), reprinted in Steup 2001, and in my forthcoming a.

Belief, Intention, and Reasons for Action

ROBERT AUDI

The concepts of belief, intention, and reasons for action are all supremely important in ethics, and the concept of belief is central in epistemology. Chapters 15 and 16 consider many aspects of belief. Here belief must be considered in relation to action, particularly from the viewpoint of moral psychology. It also has an important relation to intention. In this chapter, then, connections with earlier chapters will be evident, but certain standards for rationality are more explicitly addressed.

Intention, Hope, and Endeavor

Intention is central in the philosophy of action and crucial for understanding persons. I have provided an account of it in terms of motivational and cognitive elements; but although I have not treated it as basic among psychological attitudes, I have represented it as basic in another way: as the fundamental practical attitude (*Architecture*, 108–10). In doing this, I distinguish intending from hoping, connect it with beliefs and plans, and indicate its role in endeavors. Frederick Adams's wide-ranging and fine-grained treatment of these topics is a basis for further inquiry.

Intention and Belief

Writers on intention have divided over whether intending to A entails believing one will A. Adams quotes me on this: "Audi maintains that . . . to distinguish intending to bring about Φ by A-ing from merely hoping . . . we need to require that S at least believe her A-ing will be a probable way to achieve Φ," where "by 'probable' he means 'more likely than not'" (145). Let me begin with a clarification of my own formulation. It might be thought to imply that intending requires the *disjunctive belief* that one will A *or* probably A, which is one way of believing "at least" (nothing weaker than) that A-ing is a probable way to achieve Φ. This disjunctive belief

would be implied by 'believing that one will at least probably A'. Instead, my claim was that intending requires an indefinitely wide *disjunction of beliefs*, the weakest being that one will probably A. The requirement (as Adams realizes) is a kind of performance expectation.

For some philosophers who note that *saying* 'I intend to A' commonly expresses confidence, it is not unnatural to attack my belief condition by maintaining that it reflects only a pragmatic point. Adams says he seeks to "make sense of both sets of quotes" (one, from Bratman, supporting my belief condition, the other, from McCann, opposing that). On his view, 'I intend to A' . . . "does not literally (semantically) imply . . . a strong positive belief that I (or S) will (or probably will) A. Instead it implies that one has A-ing as a goal, and that one has a plan to A upon which one has settled" (151–2).

Two points should be stressed here. Recall the agent endeavoring to move a heavy log but strongly doubting success. First, even if the agent has settled upon a plan, neither a plan for A-ing nor, especially, *settling* on one, is required for intending. We often form intentions spontaneously and immediately; and, especially where A-ing is basic (say a raising of one's hand in class), there need be no plan for A-ing, in any nontechnical sense of 'plan'.[1] Second, though a kind of "settledness" often accompanies intending, intending commonly reflects something weaker: simply a *predominant disposition of the will*. This need not come from deliberation, choice of a plan, or any reflective episode.

A related issue concerns the dynamic role of intentions in relation to action. Consider Adams's (widely shared) claim that "Intentions can initiate and sustain action. . . . One starts acting when one believes the time to act is now. One's continuing to . . . depends in the right way on one's continued intention to do A. And one ceases one's part in acting when the intention ends" (156). I would divide the questions of initiation and sustenance. I grant the sustenance claim; but I take initiation to be in the category of events and regard intention as a dispositional, non-event, psychological property. On Adams's view, it is more natural to call the relevant belief the initiator, but even this needs qualification. Suppose I could be caused to believe *only dispositionally* that now is the time to pay my restaurant bill, so that I do not have the (occurrent) *thought* that this is so. I might then not take up a pen. Isn't it the event of *forming* such a belief that often triggers the intention, which then plays a causal role that includes appropriately sustaining the action (if the action is not momentary)? One general point, then, is that a causal theory of intending, on which it is conceived dispositionally, requires a causal dynamic account of its event triggers (this is offered in my 2006a, chap. 6).

Intending, Trying, and Intentional Action

Intending is related to trying in many ways. Adams captures a number of them, but maintains that my account of intending and its relation to intentional action does not do them justice. After noting some of the relations, he says:

> [L]et's reexamine trying. We will find each of these functional components within trying [initiating and sustaining action, guiding and monitoring ongoing actions,

being an element in coordinative plans, and terminating practical reasoning]
lending excellent support for the view that intentions are the components within
attempts or endeavors that play these cognitive roles. (156)

My broadest point in response is that these roles are too complex to be unqualifiedly
called cognitive. A cognitive element is or embodies a truth-valued element, such
as (paradigmatically) belief, though, like a supposition, the element need not be
doxastic; I consider a psychological role cognitive (in an overall way) when some
truth-valued element is predominant in it. I have already commented on the initi-
ating of actions in relation to intentions; this seems to me to be often a partly cog-
nitive role, but I see it as performed by a *trigger* of intention rather than by intention
itself, though the success of the initiation in question depends on the intention as
well as the trigger. Let us consider the other cases.

On my view, practical reasoning that favors A-ing concludes *with* a cognitive
item, say, a practical judgment that I should A. Such reasoning is best described as
terminated by the formation of such a judgment and, often, as having as its *upshot*
trying to A. As to guiding and monitoring, doubtless if we act with an intention
to achieve G, we are cognitively oriented to do things we take to be means to G, to
adjust our ongoing action if it does not match any (predominant) plan we have for
bringing about G, and to stop when we believe (e.g., see that) G has occurred. How
all this occurs is complicated, but I do not see that anything required to accommo-
date it undermines my account of intending. In particular, sometimes trying is
needed to understand how such roles are played, sometimes not. If A-ing is raising
my hand to ask a question, I see no necessary role for trying (in any nontechnical
sense); if it is trimming a shrub to a special shape, much trying may be needed.

These points might be largely agreeable to Adams, but he still views trying as
crucial for understanding intending in ways I do not and as essential in intentional
action: "Trying is the agent's contribution to what he succeeds in doing when he
executes his intention. Any linguistic oddness to saying things like 'Benny tried to
raise his arm' when this is something Benny easily can do, is due to 'linguistic odd-
ness' (i.e., implicature)" (157). Adams here conceives trying rather as volition has
been conceived. There is much to recommend a volitional theory, but I think the
volitionalist view Adams adumbrates here is too strong. Granted, on Adams's side
we can say of someone who should have raised his hand to ask a question, 'You didn't
even try', where we are not taking it that there is any problem with his capacity. But
here the point may be that S did nothing to overcome resistance (say, timidity). To
be sure, volition is arguably required even for the mental act of saying to oneself, for
example, 'I've got to do this'. But if so, is such volition a kind of trying? I doubt that
it must be, but will not repeat my case here.[2]

Perhaps, however, Adams's view on trying is not as far from mine as it may seem.
He later says, "Trying to A is an event, in a chain of events that begins in the brain
and ends, when successful, in A-ing. . . . tryings to A begin in the brain and their ini-
tiation is the immediate effect of the formation or acquisition of a proximal intention
(a here-and-now intention) to A" (157). This sounds like a mainly empirical thesis
neutral on the different views at issue concerning how trying is *conceptually* related
to intending and intentional action. Doubtless, something happens in the brain
when we begin to act, and brain events play causal roles in our acting. Whether

the *same event* must "end" when we do the thing seems to me likely to depend on the kind of case, say immediate action versus long-term delay between intention-formation and action. But these empirical hypotheses and questions are not philosophical. My philosophy of action does not foreclose any plausible empirical option.

Suppose I am right to deny that intentionally A-ing entails trying to A (as 'trying' is pre-theoretically understood). It may still be true that hope alone cannot sustain intentional action. But Adams may underestimate its possible roles. He says, "Audi's *hope alone* is not strong enough to prompt an attempt. We all hope for world peace, but not that many of us are trying to bring it about" (158). It is important here to distinguish hoping *for* from hoping *to do*. The former does not imply the latter. Suppose there is nothing I believe *I* can do that will conduce to world peace. Then, though I can hope *for* world peace, I cannot hope—or, I grant, try—*to* bring it about. Adams's example is readily accommodated by noting that conditions are not appropriate for *any* instrumental action; the example is neutral on the question dividing us. If, for instance, I believe that contributing to Doctors Without Borders has a tiny probability of bringing about peace, I can intentionally do this in the hope that it will do this (nor need my doing it require trying to contribute, in any nontechnical sense). Granted, the strength of the desire constituent in our hoping to do something may be very great and bears on whether we do it or even try to; but neither this point nor any fact about the belief element in hoping undermines my view of the relation of hoping to intending, or of either to trying.

My last point concerns the question of why we should call intentional action intentional *unless* it must be intended. Adams says, "my view . . . explains why intentional acts are intentional by placing an intention in their causal history and production" (159). Does it explain this or just why intentional acts are so called? If an intentional act is identified using terms not in the family of 'intention' say as those performed *in order to* bring about something, then my cognitive-motivational account of intentional action (in, e.g., 1986a) does quite well in explaining why acts so identified have what seem the major properties important for intentional action, and it has the advantage of allowing that one can act intentionally, even when acting only in the hope of achieving something (since one can act in order to achieve it while only hoping to succeed). All I can say about the terminological explanation question here is that we agree (1) in placing *intentionality* in the grounding of every intentional action; (2) that the paradigmatic *fulfillment* of an intention—a notion that is significant for the concept of intending itself—is an intentional action; and (3) that the paradigms of intentional actions—and virtually all of those we have occasion to *say* are intentional—are *intended* and not just intentional. It is an interesting question why there should be the exceptions my view requires—intentional actions that are not intended. This is well worth further reflection in the light of the myriad data Adams has laid out.

Self-Deception, Delusion, and Responsiveness to Evidence

With ordinary deception one person makes another believe something the first knows is not true. If self-deception is assimilated to this case, it would embody the paradox of believing something one knows is not true. My account of self-deception avoids this paradox. We might avoid it and still try to preserve the analogy by positing an *act* of

self-deception in which one puts oneself into that state in a way that is as close as possible to the way one person deceives another by lying. I have posited no such act. Granted, one could do something that *causes* one's becoming self-deceived, but the prior question is what *constitutes* being self-deceived. In answering, I have tried both to preserve as much of the analogy to ordinary deception as is reasonable and to do justice to self-deception as pre-theoretically understood. Alfred Mele's engaging and psychologically informative essay poses objections to my use of the analogy and gives a larger role to genetic elements surrounding self-deception. He also raises a doubt concerning my distinction between self-deception and delusion. I begin with that.

The Contrast between Self-Deception and Delusion

Just as a person deceived by another into believing *p* can remain aware of evidence against *p*—and can have some doubt that *p* holds—a person in self-deception with respect to *p* can also. Moreover, deceiving others is usually self-interestedly motivated, and this, too, is a clue that both Mele and I take seriously in explicating self-deception. With these points in mind, I have argued that "self-deception with respect to *p* is a state in which S [as deceiver] unconsciously knows (or has some reason to believe, and unconsciously and truly believes) that [∼*p*], [yet, as deceived] sincerely avows, or is disposed so to avow, that *p*, and has at least one want which in part explains why the belief that [∼*p*] is unconscious."[3]

In capturing what should be retained of the two-person analogy, I have identified a kind of dissociation. By virtue of this, in one's capacity as deceiver one knows that *p*, say, that one is afflicted with cancer, and, as "deceived," one sincerely avows that one is not afflicted. I have placed 'deceived' in scare quotes because I take it that the grounding of the knowledge that *p*, such as evidences from physicians' reports, has sunk in. But, like a deceived person who is aware of evidence against *p* but discounts it, qua speaker—at the conscious level—one either puts the evidence out of mind or, for example, explains it away. Explaining away evidence for *p*, however, and exposing oneself to counterevidence, may produce believing *p*. When this happens to a self-deceiver, I speak of delusion: the person now ceases to know that not-*p* and believes that *p*.

When self-deception passes into delusion, there need not be the marks of delusion Mele cites from the psychiatric literature. This depends on the case. But the typical kinds of things about which people are self-deceived or (to cite a quite different causal element) a person's moving along a common kind of route from self-deception to delusion (or both together) may produce beliefs whose possession is an instance of delusion. The cancer case and certain romantic delusions confirm this.

Given how much agreement exists between us on both self-deception and delusion, one may wonder why Mele must reject my account? He says:

> Although I have never offered a conceptual analysis of self-deception, I have suggested the following proto-analysis: people enter self-deception in acquiring a belief that *p* if and only if *p* is false and they acquire the belief in a *suitably biased way*. (166)

It is significant that he does not work from an account of what self-deception *is*. His focus is on the process of entering self-deception (of deceiving oneself, in a significant use of that phrase). He perhaps takes it as obvious that success in the process as he describes it

results in self-deception. But that is really not obvious. Success must, to be sure, result in being deceived; but just as succeeding in self-annihilation or self-criticism need not result in being in states of self-annihilation or self-criticism (it is not clear what these would be), succeeding in self-deception need not result in being in a state of self-deception. It entails having *been* deceived, just as the other cases must result in having been annihilated or criticized. But the natural description of the process, 'deceiving oneself', need not transfer, in the way needed for the problem before us, to the resulting state, condition, or state of affairs. *Self-caused deception need not be self-deception*; and even an analysis of self-caused self-deception may not suffice for an adequate account of self-deception.

This observation is not meant negatively. I consider Mele's proto-analysis plausible if taken to be an account of "deceiving oneself" in the *sense* of producing a condition of delusion of the kind that sometimes looks like the state of self-deception. His "suitably biased way" of coming to believe p is highly consonant with my conception of the genetic and sustaining role of the motivation underlying self-deception, and I would stress that when the agent's responsiveness to counterevidence causes *delusion*, it may also eliminate or greatly reduce the (doubtless sometimes painful) dissociation characteristic of self-deception. I also find much of what he says in clarifying the notion of delusion instructive. But I have avoided commitment to the kind of psychological notion implicit in, say, "delusional jealousy." My anchor here is instead the commonsense notion of being deluded in believing something. This often does not rise to the *delusional* in the rich, psychiatric sense. Self-deception can, however, lead to either kind of delusion.

Two Models for Understanding Self-Deception

Given our wide agreement on data, why do we differ on self-deception? I credit the influence of a kind of difference more common in philosophy than many people realize: a difference in orienting models. Mele's model is apparently the *act* of deceiving, in which the deceived forms a false belief and so does not see the truth in question; mine is the *state* of being deceived—altered to apply, so far as possible, to a person who is both deceived and deceiver and hence, apart from inconsistent beliefs, must see the truth.

One might wonder why it is not as reasonable to take the deceiver as primary in our orienting model, hence to posit, as Mele does, a false belief that p as an element in self-deception with respect to p. In part, my reason for not doing so is that just as, in ordinary two-person deception, the deceiver has the upper hand, informationally speaking, in self-deception the *relevant* evidential elements bearing on p (those affecting the self-deceiver) favor *not-p*, say the proposition that one will not recover from cancer. But whereas on my view the evidence is dominant in producing true belief, on Mele's the motivational or other biasing elements dominate over any evidential ones present and S acquires a false belief.

Mele's model, then, does justice to the idea that self-deception appears to entail deception; mine does justice to the dissociation *self*-deception apparently entails and places the element of deception in the tendency to make sincere avowals of propositions one unconsciously knows to be false yet—in misleading but natural terms—"consciously believes" to be true. By virtue of the dissociation, the sincere avowal, which would normally imply belief, manifests a kind of deception. We both

take the speaker or would-be speaker to be the dupe, but I would say that the genuinely believing dupe is deluded, whereas the self-deceived one is dissociated.

This apparent difference in orienting models is related to the historicity question. To see this consider Mele's thesis that

> [A] true judgment about whether particular people do or do not enter self-deception . . . depends, among other things, on what roles motivation or emotion plays in the production or sustaining of these beliefs. Self-deception, as Audi understands it, "is not a *historical* concept. If I am self-deceived, so is my perfect replica" (1997b, 104). . . . As I see it, asking what self-deception would be like given the stipulation that the concept is not a historical one is analogous to asking what remembering that *p* would be like if "remembering" were not a historical concept. (173)

Let me respond first to the analogy. Certainly, we cannot remember unless we have *retained* something for some time.[4] But *re*membeing wears its historicity on its sleeve; 'self-deception', by contrast, has a process-state duality. Granted, 'self-deceived' is historical—when it does not mean simply 'in self-deception'. But 'in self-deception' is not historical. I can agree, then, that Mele is not stipulating; but nor am I.

The difference is apparently in our orienting models: an act model for him, a state model for me. Is there a reason to prefer the latter? Must we deny that if I am perfectly duplicated, then, if I am now in self-deception, a full, purely psychological description of me (apart from listing self-deception) would ground (or even entail) my being in self-deception? (I here ignore complications concerning objectual beliefs; accommodating those would require at least the duplicate's being in the same external circumstances.) Even someone who denied this would grant that the *only* relevant difference would result from different answers to the question of how the person got that way. Would anyone need to know that to understand me *now*? We normally distinguish what something is from how it arose; this seems natural here. In addition, there are normally no acts of self-deception analogous to acts of other-deception, such as momentary (successful) lying; and Mele rightly focuses on a *process* of entering self-deception. That process is important. I continue to hope that my account of self-deception does justice to the state in which that process terminates.

Reasons and Motives in Individuals and Groups

So far, my main focus has been on individuals. But we properly attribute both psychological and normative properties to groups. It is important to explore the application of individual concepts to social cases. One would expect that inasmuch as groups are constituted by individuals standing in certain kinds of relations, those concepts would apply to them. But the applications are difficult to articulate. Raimo Tuomela's far-reaching essay makes many important connections between the two cases.

Reasons, Motivation, and Normativity

Many kinds of things are cited as reasons for action. Philosophers who think about reasons—whether explanatorily or normatively—rightly seek economy in accounting

for them. One extreme is Procrustean regimentation, another excessive latitudinarianism. My theory of reasons aims at avoiding both defects. It does this by distinguishing reasons proper (say, *that* sending a condolence letter is necessary) from reason states (say, *believing* this proposition), by identifying different cases under each heading, and by systematically connecting all of these elements. A full review of the theory is impossible here, but in responding to Tuomela I can extend the account of how cognitive reasons differ from conative ones.

In contrasting our accounts of reasons, Tuomela says, "I have also added (ii) *belief-related* reason-cases, where a belief plays a salient motivational role. For Audi, belief contents can also be reasons, although not in a primary sense. In his 2001 book (see, e.g., 121) he does accept that also beliefs may play the role of reason (although subordinated to wants), so perhaps my approach also concurs with his recent views" (183–4). Let me explain my position. Unlike some writers on reasons,[5] I do not identify reasons with facts, though I take them to be *expressible by citing a proposition* or—when the relevant proposition is true—a fact. This propositional expressibility of practical reasons should not, however, mask something easily missed. The propositional expression of the reason is, in a sense, less basic than the infinitival one—for instance 'to acknowledge her grief'. This point reflects the difference between theoretical and practical reason. A supporting point is that the practical attitudes and, correspondingly, the contents of infinitives, are not appraisable as true or false, whereas propositions are so appraisable. Thus, reasons for action, propositionally expressed, can be appraised in a way that does not apply to them as expressed infinitivally, and misleadingly invites attempts to construe intentions and desires as having truth value.

A more important contrast between Tuomela's views and mine may be this: on my view, that A-ing will bring about a goal, G—say, reducing pain—counts as a reason to A, *in virtue of* the desirability of G (G itself may be desirable in virtue of a desirability-characteristic, a property that grounds some degree of G's desirability). The desirability may be instrumental, but most important here is the case in which G has value "in itself." Consider this idea. That A-ing will bring about G is not intrinsically normative but owes its normative authority to a normative property of the relevant state of affairs or act-type, G (G may be an act-type, say reducing pain). By contrast, the normative authority of bringing about G (of reducing pain) seems to be intrinsic to it and basic (or at least more basic). To reduce pain is desirable in itself, in a sense implying that there is reason for *any* action that does this. Reducing pain might be called a constitutive aim of practical reason.

Given this objectivist (partial) account of reasons, we can say that even for rational actions the agent sometimes has only an apparent objective reason to act. But suppose I rationally yet falsely believe that something is highly desirable, for instance that viewing certain sculptures would be enjoyable. I can still act in order to realize this end, where this is acting for a reason, in the normative sense in which grounding in a reason counts toward the rationality of the action. I can give my reason for going to the relevant museum by saying 'to view the new sculptures'; and I can act *for* this reason. But my action may fail to achieve the good I hoped for, just as a rational belief can fail to be true. When all goes well, however, we can act in the light of justified true beliefs, for reasons that are objectively good, and achieve practical success.

The distinction between objectively good reasons and reasons that only seem objectively good suggests a more important contrast between us. For Tuomela,

> [A]ll motivational reasons are taken to have . . . *normative force* . . . I also include mere "technical" or instrumental normativity, while Audi might have in mind only substantive (e.g., rational or moral) normativity . . . I may want to open the window and have formed the intention to do that. The (so-far nonexistent) state that the window is open conceptually and "hermeneutically" requires in this merely technically normative sense that I act appropriately in relation to the reason state to be tokened. Audi would hardly object to the requirement, basically involved here, that if you intend that a state obtain you have to act appropriately. (184)

I agree that a desire *on* which the agent acts or can act is a *motivational* reason and can explain action—this is a "hermeneutic" point. But this does not imply normativity—except insofar as explainability of action entails a kind of intelligibility of it that implies whatever limit on irrationality goes with instrumental behavior. Perhaps, however, this kind of connection is all that is required by what Tuomela calls "'technical' or instrumental normativity." In any case, the kind of rationality required for (motivational) explainability of action seems at best a minimal kind. It must be granted, however, that the kinds of instrumental patterns Tuomela has in mind are necessary for being a rational *agent*. This is an important normative point in itself.

If we go beyond the technical normativity in question, as the last quoted sentence appears to in suggesting an *intentions-as-normative-grounds view*, we arrive at a position that (as I think Tuomela would agree) is too strong. Granted, if I intend to A, I normally have a predominant want to A, may act *in order to A*, and may thus render my A-ing as a goal of instrumental action, minimally intelligible. But suppose the want to A is irrational, perhaps implanted by brain manipulation. I need not then have any reason *to A* (a normative one), even if I, in doing something in order to A or in A-ing, can be intelligibly said to have acted *for* a (motivational) reason.

To support this, let me indicate why the intentions-as-grounds principle is at once plausible and mistaken. It is plausible to hold that at a *given* time it is (at least normally) prima facie irrational to (1) intend to bring about G; (2) believe that A-ing is necessary for doing so (and that one can A); and yet (3) not A or even intend or try to A. But this is like the principle that one should not, at a given time, believe that *p*, that *p* entails *q*, and that not-*q*. The latter does not imply that believing *p*, and believing that it entails *q*, constitutes a normative reason to believe *q*. For one thing, *p* might be obviously false and believed irrationally; for another, *q* might be so obviously false that when one considers it, one should *disbelieve p* even if *p* is not obviously false, taken alone. (I have discussed such cases in 1995a and 2006b, ch. 8). By contrast, if *p* is independently plausible, it is, on *that* score, a (normative) reason, even if a weak one, for any proposition it obviously entails; similarly, if G is independently worth realizing, then bringing about G is a (normative) reason for instrumental action that will produce G. The general lesson is that from the negative point that, at any given time, certain *combinations* of propositional attitudes (of the kind in question) are prima facie irrational, it does not follow that any proper subset of them provides an *independent* reason to form the remaining attitude(s) or to act.[6]

Social Action

Tuomela has produced a sophisticated theory of social action that builds on the theory of action for individuals. In his application of my own individual theory to the social case I find welcome confirmation. Individuals acting in groups are still understandable in terms of their own motivation and cognition, but the scope and function of that cognition is essentially social. There is only one point on which I would qualify the social application Tuomela makes of my account.

I have maintained that we each normally have a certain noninferential access to our propositional attitudes and that when we act for a reason we are noninferentially *disposed* to attribute the action to the relevant reason state(s), say our motivating want. Does this extend to the social case? Take playing a game. Tuomela suggests so, indicating that in the social case my noninferential disposition obtains: "$x_1, \ldots, x_m$ are noninferentially disposed, independently of seeking reasons they have had, or might have had, at or before t, for joint A-ing, to attribute their A-ing to the joint want and (explaining) mutual belief(s)" (190). If, for instance, we are building a summerhouse together, I would be disposed noninferentially to attribute your nailing down floorboards with me to a desire, for instance to complete our construction, and a belief that so acting will help.

Here we find an asymmetry with the individual case. True, I make no inference that manifests itself in consciousness as, say, reasoning from the articulated ground that the relevant desire and belief best explain your behavior to the conclusion that you have those attitudes. This may indeed be what Tuomela has in mind in suggesting the condition. He is outlining an account of social action on which the norm is for the mind to be occupied with what is being *done* as opposed, for instance, to grounds for others' cooperation. This is consistent, however, with my belief that you (continue to) have the relevant attitudes being *based* on observing your behavior in the context of the joint nailing. This dependence on observation does not extend, however, to my self-ascriptions of reasons; these beliefs seem "directly" grounded in the reason states themselves or, in any case, not dependent on observation of my behavior.

In a common kind of case, then, some of my essential social knowledge of others' reasons—at least in many instances of joint action—may be inferential in being premise-dependent, even if not in being based on an episode of reasoning. Tuomela's proposed condition suggests, however, a third kind of case. Many social situations may be such that *given* our history of interaction with others, we form beliefs about their intentions and beliefs in a noninferential way that is like belief-formation in certain cases of pattern-recognition. There would still be a dependence on *behavioral observation* that yields a contrast with the case of self-knowledge; but there might in such a case be certain dispositions to attribute attitudes to others that are genuinely noninferential. The attributions would be observation-dependent, but, unlike inferential attribution, not premise-dependent. We might do well to explore, then, whether social action and other social phenomena divide into at least three cases: those in which we infer others' attitudes; those in which we attribute them on the basis of observational beliefs yet do not go through any inferential process; and those in which a kind of direct social recognition of intentions and the like yields presuppositions and beliefs about others.

The importance of this asymmetry should not be exaggerated. For on my view of testimony, others *can* know noninferentially that you want to do a joint project if,

in normal conditions, you say you do. Surely a great many of our social beliefs and knowledge, including our beliefs and knowledge about others' motivation and cognition, is based on what they say (whether to us or to others within our hearing). In at least many such cases, the attributions Tuomela is speaking of *are* noninferential—not even premise-based. Nonetheless, even for testimony-based knowledge, *perception* of the testimony is needed; and this perceptual requirement does not apply to one's own case. This difference is important, but it can be accommodated by qualifications of Tuomela's proposal that preserve its basic integrity.

***

This chapter and the previous two have defended my views on many important philosophical questions and, at some points, extended them. My work to date has developed a systematic intuitionism in ethics, a moderate rationalism in epistemology, and an objectivist account of both theoretical and practical reason. The ethical intuitionism includes an integration of Rossian and Kantian elements. The moderate rationalism includes a moral epistemology that contains an account of the self-evident on which certain moral principles have that epistemic status. On my view, justification for believing self-evident principles can be noninferential even though they may admit of inferential justification or even of proof. The view provides both for defeasibility in our justification for believing self-evident principles and, on the other side, for groundability of our justification for holding such principles. The objectivist account of practical reason that goes with this view includes a theory of reasons for action and a multifaceted analogy between these and reasons for belief. I have also defended an epistemological internalism about justification side by side with an externalist conception of knowledge. In doing this, I distinguish different kinds of beliefs and, in that connection, of content. These positions have wide implications both for general epistemology and for specific topics, such as perception, memory, and the status of testimony as a basis of knowledge and justification.

Justification is not the only normative notion I have tried to clarify; I have taken rationality to be both broader and subject to less stringent criteria. Given my account of belief, and of the distinction between dispositional beliefs and dispositions to believe, I have framed a conception of faith that—in and outside the philosophy of religion—widens the field in which it may be viewed as rational. In all of the philosophical areas in question, however, I have been able to bring only selected elements into this volume. The principle of selection has been the demands of responding to the excellent set of critical papers in this volume. This response, even taken together with my opening chapter in part 1, is far too short to do full justice to the ideas the authors have presented; but it may enable readers, as it will certainly enable me, to progress toward new and expanded theories in the fields of our engagement.[7]

Notes

1. This is argued in some detail in my 1991a. I grant that that one could 'plan *to* raise one's hand when the clock strikes', but here 'plan' seems equivalent to 'intend'; note that one need not have *a plan* to do this.

2. My account of the nature and possible roles of volition is provided in my 1993f. For Adams's position on volitions, see Adams and Mele 1992.

3. From my 1982, 155.

4. The short article that Mele cites did not make a critical distinction. It is one thing to say that the *concept* of self-deception is not historical and quite another to say that my perfect replica would necessarily be in self-deception *exactly* matching mine. For suppose that I am in self-deception about whether, in a certain article, I was correct on a point, where I unconsciously know or truly believe that I was not. My just-created perfect replica, being a psychological duplicate, will believe but not know or *truly* believe the counterpart proposition, since he did not write the article. He will, however, have the appropriate dissociation and be in a state *psychologically isomorphic* with mine (for further description of this state, see, e.g., my 1989, which portrays self-deception both phenomenologically and as a kind of evidence-responsive state). It remains true, moreover, that the *concept* of self-deception is not historical, since the phenomenon clearly can occur in my duplicate regarding nonhistorical matters. I would thus now explicitly distinguish instances of essentially historical self-deception from the other cases. I am grateful to Mario De Caro for objections showing the need to take account of cases in which self-deception cannot be assumed to be ahistorical.

5. A leading proponent of the facticity of reasons is Derek Parfit (1997).

6. This view is developed in *Practical Reasoning and Ethical Decision* (2006b). Also relevant are points I have made about the closure of justification in, e.g., 1993a and 1995a.

7. For making this chapter and the previous two possible and contributing immeasurably to their content, I am grateful above all to the authors of the critical chapters. The editors— John Greco, Alfred Mele, and Mark Timmons—have all provided insightful comments; the commentators on the drafts of the critical chapters have helped me as well; and specific beneficial comments on an earlier draft (or parts of it) were generously provided by Fred Adams, Laurence BonJour, Roger Crisp, Bernard Gert, Peter Graham, Hugh McCann, Michael McKinsey, Paolo Monti, Jeff Speaks, Raimo Tuomela, and, especially, Walter Sinnott-Armstrong. For numerous critical comments and many helpful suggestions on several drafts, I am enormously grateful to Mark Timmons.

References for Chapters 15 to 17

Adams, F., and A. Mele. 1992. The Intention/Volition Debate. *Canadian Journal of Philosophy* 22: 323–38.

Alston, William P. 1993. *The Reliability of Sense Perception*. Ithaca, N.Y. Cornell University Press.

Audi, Robert. 1982. Self-Deception, Action, and Will. *Erkenntnis* 18: 133–58.

——. 1983. The Causal Structure of Indirect Justification. *Journal of Philosophy* 80 (7): 398–415.

——. 1986a. Acting for Reasons. *Philosophical Review* 95 (4): 75–205.

——. 1986b. Belief, Reason, and Inference. *Philosophical Topics* 14 (1): 27–65.

——. 1988. Self-Deception, Rationalization, and Reasons for Acting. In *Perspectives on Self-Deception*, ed. Brian McLaughlin and Amelie O. Rorty, 92–120 Berkeley and Los Angeles: University of California Press.

Audi, Robert. 1989. Self-Deception and Practical Reasoning. *The Canadian Journal of Philosophy* 19 (2): 246-66.

——. 1991a. Intention, Cognitive Commitment, and Planning. *Synthese* 86: 361–78.

——. 1991b. Structural Justification. *Journal of Philosophical Research* 26: 473–92.

——. 1993a. The Old Skepticism, the New Foundationalism, and Naturalized Epistemology. In *The Structure of Justification*, ed. Robert Audi, 353–77. Cambridge: Cambridge University Press.

——. 1993b. *Action, Intention, and Reason.* Ithaca, N.Y.: Cornell University Press.

——. 1993c. Ethical Naturalism and the Explanatory Power of Moral Concepts. In *Naturalism: A Critical* Appraisal, ed. Steven Wagner and Richard Warner, 95–115. Notre Dame: University of Notre Dame Press.

——. 1993d. Mental Causation: Sustaining and Dynamic. In *Mental Causation*, ed. John Heil and Alfred R. Mele, 53–74. Oxford: Oxford University Press.

——. 1993e. *The Structure of Justification*. Cambridge: Cambridge University Press.

——. 1993f. Volition and Agency. In *Action, Intention, and Reason*, ed. Robert Audi, 74-108, Ithaca, N.Y: Cornell University Press.

——. 1994. Dispositional Beliefs and Dispositions to Believe. *Nous* 28 (4): 419–34.

——. 1995a. Deductive Closure, Defeasibility, and Skepticism: A Reply to Feldman. *Philosophical Quarterly* 45: 181, 494–99.

——. 1995b. Memorial Justification. *Philosophical Topics* 23 (1): 31–45.

——. 1995c. Perceptual Experience. Doxastic Practice, and the Rationality of Religious Commitment. *Journal of Philosophical Research* 20: 1–18.

——. 1996. Religious Experience and the Practice Conception of Justification. In *The Rationality of Belief and the Plurality of* Faith, ed. Thomas D. Senor, 127–49. Ithaca, N.Y.: Cornell University Press.

——. 1997a. *Moral Knowledge and Ethical Character*. Oxford: Oxford University Press.

——. 1997b. The Place of Testimony in the Fabric of Knowledge and Justification. *American Philosophical Quarterly* 34 (4): 404–22.

——. 1998. The Axiology of Moral Experience. *Journal of Ethics* 2: 355–75.

——. 1999a. Doxastic Voluntarism and the Ethics of Belief. *Facta Philosophica* 1 (1): 87–109.

——. 1999b. Self-Evidence. *Philosophical Perspectives* 13: 205–28.

——. 2001a. *The Architecture of Reason*. Oxford: Oxford University Press.

——. 2001b. An Internalist Theory of Normative Grounds, *Philosophical Topics* 29: 19–46.

——. 2003. *Epistemology: A Contemporary Introduction to the Theory of Knowledge*. London: Routledge.

——. 2004a. The A Priori Authority of Testimony. *Philosophical Issues* 14: 18–34.

——. 2004b. *The Good in the Right: A Theory of Intuition and Intrinsic Value*. Princeton, N.J.: Princeton University Press.

——. 2006a. Ethical Generality and Moral Judgment. In *Contemporary Debates in Ethical* Theory, ed. James Dreier, 285–304. Oxford. Blackwell.

——. 2006b. *Practical Reasoning and Ethical Decision*. London: Routledge.

——. Forthcoming a. The Ethics of Belief: Doxastic Self-Control and Intellectual Virtue. *Synthese*.

——. Forthcoming b. The Grounds and Structure of Reasons for Action. In *Intention, Deliberation and Autonomy: The Action-Theoretic Basis of Practical Philosophy*, ed. Christoph Lumer and Sandro Nannini. Hants, United Kingdom: Aldershot.

——. Forthcoming c. Skepticism about the A Priori. In *The Oxford Handbook of Skepticism*, ed. John Greco. Oxford: Oxford University Press.

——. Forthcoming d. Testimony, Veracity, and Credulity. In *The Epistemology of Testimony*, ed. Jennifer Lackey and Ernest Sosa, forthcoming. Oxford: Oxford University Press.

Burge, Tyler. 1982. Other Bodies. In *Thought and Object*, ed. Andrew Woodfield. Oxford: Clarendon.

——. 1986. Cartesian Error and the Objectivity of Perception. In *Subject, Thought, and Context*, ed. Philip Pettit and John McDowell, 117–36. Oxford: Oxford University Press.

Byrne, Alex, and Michael Tye. 2006. Qualia ain't in the head. *Nous* 40 (2): 241–55.

Chalmers, David. 2002a. The Components of Content. In *Philosophy of Mind: Classical and Contemporary Readings*, 608–33.

——. 2002b. *Philosophy of Mind: Classical and Contemporary Readings*. Oxford: Oxford University Press.

Devitt, Michael. 1990. A Narrow Representational Theory of the Mind. In *Philosophy of Mind*, 2nd ed., ed. William Lycan, 287–317. Oxford: Blackwell.

Goldberg, Sanford. 2002. Do Anti-Individualistic Construals of Propositional Attitudes Capture the Agent's Conceptions? *Nous* 36 (4): 597–621.

Goldman, Alvin I. 1999. Internalism Exposed. *Journal of Philosophy* 96: 271–93.

Graham, Peter J. 2000. Conveying Information. *Synthese* 123: 365–92.

——. 2007. Theorizing Justification. In *Knowledge and Skepticism*, ed. Michael O'Rourke. Cambridge, Mass.: MIT Press.

Heil, John, and Alfred R. Mele, eds. 1993. *Mental Causation*. Oxford: Oxford University Press.

Kant, Immanuel. 1985. *Groundwork of the Metaphysics of Morals*. Trans., H. J. Paton, London: Hutchinson University Library, 1948.

Kitcher, Philip. 1983. *Mathematical Knowledge*. Oxford: Oxford University Press.

Lackey, Jennifer. 1999. Testimonial Knowledge and Transmission. *The Philosophical Quarterly* 49: 471–90.

——. Forthcoming. Learning from Words. *Philosophy and Phenomenological Research*.

McKinsey, Michael. 1991. The Internal Basis of Meaning. *Pacific Philosophical Quarterly* 72: 143–69.

——. 1994. Individuating Beliefs. *Philosophical Perspectives* 8: 303–30.

——. 2003. *On Liberty*. Ed. David Bromwich and George Kateb. New Haven, Conn.: Yale University Press.

Parfit, Derek. 1997. Reasons and Motivation. *Proceedings of the Aristotelian Society*. Supplementary volume.

Pautz, Adam. 2006. Sensory Awareness Is Not a Wide Physical Relation: An Empirical Argument against Externalist Intentionalism. *Nous* 40 (2): 205–40.

Ross, W. D. 1930. *The Right and the Good*. Oxford: Oxford University Press.

Shoemaker, Sydney. 1980. Causality and Properties. In *Time and Cause*, ed. Peter van Inwagen, 109–35. Dordrecht: Reidel.

Stalnaker, Robert C. 1990. Narrow Content. Reprinted in *Context and Content*, 194–209. Oxford: Oxford University Press.

Steup, Matthias, ed. 2001. *Knowledge, Truth, and Duty*. Oxford: Oxford University Press.

Wagner, Steven, and Richard Warner, eds. 1993. *Naturalism: A Critical Appraisal*. Notre Dame: University of Notre Dame Press.

Williamson, Timothy. 2004. Philosophical "Intuitions" and Scepticism about Judgement. *Dialectica* 58 (1): 109–53.

Index