Ineffability and Philosophy

André Kukla

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

Ineffability and Philosophy

This book provides a fascinating analysis of the idea of what can't be said. As the author states, he cannot tell us directly what it is that can't be said, but he can tell us a lot *about* what can't be said.

The book ascertains whether the notion of there being a truth, or a state of affairs, or knowledge that can't be expressed linguistically is a coherent notion and the author distinguishes different senses in which it might be said that something can't be said.

The first chapter looks at the question of whether ineffability is a coherent idea. Chapter 2 evaluates two families of arguments regarding whether ineffable states of affairs actually exist: the argument from mysticism and the argument from epistemic boundedness. Chapter 3 looks more closely at the relation between mystic and non-mystic stances. In the fourth and final chapter the author distinguishes five qualitatively different types of ineffability.

Ineffability and Philosophy is a significant contribution to this area of research and will be essential reading for philosophers and those researching and studying the philosophy of language.

André Kukla is a professor in the Departments of Philosophy and Psychology of the University of Toronto. He is the author of *Studies in Scientific Realism* (1998), *Social Constructivism and the Philosophy of Science* (2000) and *Methods of Theoretical Psychology* (2001).

Routledge Studies in Twentieth-Century Philosophy

1 **The Story of Analytic Philosophy**
Plot and heroes
Edited by Anat Biletzki and Anat Matar

2 **Donald Davidson**
Truth, meaning and knowledge
Edited by Urszula M. Zeglén

3 **Philosophy and Ordinary Language**
The bent and genius of our tongue
Oswald Hanfling

4 **The Subject in Question**
Sartre's critique of Husserl in *The Transcendence of the Ego*
Stephen Priest

5 **Aesthetic Order**
A philosophy of order, beauty and art
Ruth Lorland

6 **Naturalism**
A critical analysis
Edited by William Lane Craig and J. P. Moreland

7 **Grammar in Early Twentieth-Century Philosophy**
Richard Gaskin

8 **Rules, Magic and Instrumental Reason**
A critical interpretation of Peter Winch's philosophy of the social sciences
Berel Dov Lerner

9 **Gaston Bachelard**
Critic of science and the imagination
Cristina Chimisso

10 **Hilary Putnam**
Pragmatism and realism
Edited by James Conant and Urszula Zeglen

11 **Karl Jaspers**
Politics and metaphysics
Chris Thornhill

12 **From Kant to Davidson**
The idea of the transcendental in twentieth-century philosophy
Edited by Jeff Malpas

13 **Collingwood and the Metaphysics of Experience**
A reinterpretation
Giuseppina D'Oro

14 **The Logic of Liberal Rights**
A study in the formal analysis of legal discourse
Eric Heinze

15 **Real Metaphysics**
Edited by Hallvard Lillehammer and Gonzalo Rodriguez-Pereyra

16 **Philosophy After Postmodernism**
Civilized values and the scope of knowledge
Paul Crowther

17 **Phenomenology and Imagination in Husserl and Heidegger**
Brian Elliott

18 **Laws in Nature**
Stephen Mumford

19 **Trust and Toleration**
Richard H. Dees

20 **The Metaphysics of Perception**
Wilfrid Sellars, critical realism and the nature of experience
Paul Coates

21 **Wittgenstein, Austrian Economics and the Logic of Action**
Praxeological investigations
Roderick T. Long

22 **Ineffability and Philosophy**
André Kukla

23 **Kant, Cognitive Metaphor and Continental Philosophy**
Clive Cazeaux

24 **Wittgenstein and Levinas**
Ethical and religious thought
Bob Plant

Ineffability and Philosophy

André Kukla

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2005 by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Simultaneously published in the USA and Canada
by Routledge
270 Madison Ave, New York, NY 10016

Routledge is an imprint of the Taylor & Francis Group

This edition published in the Taylor & Francis e-Library, 2005.

"To purchase your own copy of this or any of Taylor & Francis or Routledge's collection of thousands of eBooks please go to www.eBookstore.tandf.co.uk."

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data
A catalog record for this book has been requested

ISBN 0-203-32507-9 Master e-book ISBN

ISBN 0–415–33970–7 (Print Edition)

For Eli

Contents

Preface

An author who writes a book on what can't be said owes the reader an apology, at least in the Platonic sense of the word. It's true that I can't directly tell you what it is that can't be said. But I might be able to tell you a lot *about* what can't be said. For starters, I might be able to ascertain whether the notion of there being a truth, or a state of affairs, or knowledge that can't be expressed linguistically is a coherent notion. I might distinguish different senses in which it might be said that something can't be said. (Some of these varieties of ineffability might be coherent and others might be incoherent.) I might even be able to establish that some varieties of ineffability aren't merely coherent possibilities, but that they actually exist. To be sure, I will never be able to *display* an ineffable truth; but I might be able to establish its existence by an indirect argument, just as mathematicians sometimes prove that there exists a number having a certain property without being able to say what the number is. In any event, these are the sorts of questions that I seek to answer in this book.

The first and longest chapter of the book is devoted to the question of whether ineffability is a coherent idea. I note that ineffability comes in several degrees of stringency, and argue that the weaker grades are demonstrably coherent, while the stronger grades have not been demonstrated to be incoherent. These claims contradict the conclusions of influential analyses by Donald Davidson and William Alston, whose arguments I undertake to refute.

Granted that the notion of ineffable states of affairs is not incoherent, the question arises whether they actually exist. In Chapter 2, I evaluate two families of arguments to that effect. The argument from mysticism infers the existence of ineffabilities from the premise that some people (mystics) report having insights that

can't be put into words. The argument from epistemic boundedness infers the existence of ineffabilities from the premise that there are hypotheses that no one can entertain. These arguments run in opposite directions: the mystical argument supports the ineffability thesis on the basis of certain extraordinary capacities of the human mind, whereas the argument from epistemic boundedness supports the ineffability thesis on the basis of certain of the mind's *in*capacities. I conclude that these arguments jointly sustain at least some relatively weak forms of the ineffability thesis. I also note that there the question of whether there exist even quite strong forms of ineffability are amenable, at least in principle, to an empirical resolution.

Granted that mystics have ineffable intuitions, what are the conditions, if any, under which non-mystics should further accept that the intuitions are veridical? In Chapter 3, I maintain that the relation between mystic and non-mystic may be regarded as a special case of the relation between a putative expert and a novice who is unqualified to form an independent judgment in the expert's field. Alvin Goldman has delineated the sources of evidence available to novices for accrediting (or repudiating) an experts' claim. I apply Goldman's general analysis to the specific case of the mystic and the non-mystic. I conclude that the evidence both for and against accrediting the mystic is very weak. Once again, my more important conclusion is that there are possible data that could settle the issue.

In the first three chapters of the book, the ineffability of a state of affairs is identified with its unrepresentability in some language or class of languages. This is, I think, the prototypical variety of ineffability. But there are other senses in which it may be said that something "can't be said." In the fourth and final chapter, I distinguish five qualitatively different types of ineffability. A state of affairs *S* is *unrepresentable* if there is no sentence (in a given language or class of languages) for which *S* is the truth-condition. *S* is *unabducible* (by a person or class of persons) if, despite the fact that it may be representable, no suitable sentence for it ever comes to mind for consideration as a possible speech act. *S* is *unselectable*, if, despite the fact that a suitable sentence for *S* comes to mind, the speaker always evaluates it as an inappropriate thing to say. *S* is *unexecutable* if, despite the fact that a suitable sentence for *S* is selected for saying, the speaker finds herself incapable of performing the requisite speech act. Finally, *S* is *unreportable* if speaking a

suitable sentence for *S* is incompatible with that sentence's being true. Each of these varieties of inexpressibility comes in various degrees. Each of them also provides a different interpretation of the religious mystic's claim to have experienced an ineffable insight.

I have written these four chapters in such a way that each of them can be read independently of the other three. To attain that end, I've had to present some of the material more than once. For example, each of the four chapters contains at least a brief and simplified description of the grades of representational ineffability.

The writing of this book was supported by a research grant from the Social Sciences and Humanities Research Council of Canada. Much of the content of the section on McGinn's argument for cognitive closure in Chapter 2 appeared previously in "Mystery, Mind, and Materialism" (*Philosophical Psychology*, 1995, 8, 255–264, www.tandf.co.uk). I would like to thank Alvin Goldman, Rebecca Kukla, Richard Manning, and Joel Walmsley for their valuable comments on earlier drafts of this material.

1 Ineffability—the very idea

Are there truths or facts or states of affairs in the world that cannot be expressed or represented by linguistic means? The issue arises in a number of disparate fields of speculative thought, including religious studies, the philosophy of mathematics, and contemporary cognitive science. Religion is of course the most familiar source of ineffability claims. The idea that there is an ineffable being, or truth, or experience, and that this ineffability profoundly matters to our assessment of the human condition, is one of the meanings (the non-pejorative one) of the much-abused term "mysticism." There have been articulate and widely influential mystics in virtually every religious tradition. In contrast, mathematical ineffabilism is a distinctly heterodox point of view concerning some relatively arcane foundational issues in set theory and metamathematics. The mathematical ineffabilists include Cantor, Bolzano, Dedekind, and more recently, A. W. Moore (1990) and Rudy Rucker (1982). Despite their very different social situations, there are substantive similarities between religious and mathematical ineffabilisms. Both of them involve the claim that we are able to understand or come to know certain truths which it is beyond the power of language to express. As will be seen immediately below, this is not an invariable basis for ineffability claims. Mathematical and many religious ineffabilists also agree on the more specific proposition that the attempt to express the ineffable must systematically embroil us in contradictory assertions. Moore and Rucker are wont to emphasize these and other parallels by referring to their doctrine too as a species of mysticism.

There is also an influential line of thinking in contemporary cognitive science which has consequences for the ineffability thesis, although these consequences have not yet been fully spelled out.

Fodor (1983), Chomsky (1975), and McGinn (1989) have argued, each in his own way, that there are bound to be some hypotheses that the human mind is incapable of entertaining. In Fodor's terminology, human minds are "epistemically bounded." If the argument for epistemic boundedness is sound, it seems at least prima facie plausible that it will also sustain the closely related thesis that there are facts which cannot be expressed in any humanly accessible language.

The foregoing discussion serves to locate the topic of ineffability in a broader geography of ideas. However, I will postpone the evaluation of the religious, mathematical, and cognitive-scientific arguments until Chapter 2. In Chapter 1, I seek to clarify the import of the ineffability thesis itself. I hold in abeyance the question of whether there are good reasons for believing in the existence of ineffable facts, and ask the logically prior question: what does it mean to say that there are facts that cannot be expressed? Are there ways to construe the ineffability thesis which are at once philosophically interesting and coherent?

Indescribable entities

Let's begin with the most familiar version of the ineffability thesis: that God, or the Tao, or the soul, or mystical experience, or some other religious object is indescribable. Here are representative claims from Asian, ancient European, and modern mystical sources:

> Brahman has neither name nor form, transcends merit and demerit, is beyond time, space, and the objects of sense-experience. . . . Supreme, beyond the power of speech to express.
>
> (Shankara, quoted in Alston, 1972, 76)

> No form belongs to Him, not even one for the Intellect. . . . What meaning can there be any longer in saying: "This and this property belongs to Him"?
>
> (Plotinus, quoted in Alston, 1972, 76)

> To say that God is ineffable is to say that no concepts apply to Him, and that He is without qualities. . . . And this implies that any statement of the form "God is x" is false.
>
> (Stace, 1952, 33)

> Thus to the intellect He is blank, void, nothing. You cannot attach any predicate to Him.
>
> (Stace, 1952, 42)

The classical—and nearly definitive—analysis of this claim is William Alston's (1972). Alston's discussion is directed to the assertion that "you cannot attach any predicate" to God (except, of course, for His masculine gender). Alston doesn't inquire into what it might mean for a predicate to be inapplicable to an entity. There are two candidates: the predication may be false, or it may be neither true nor false. Stace evidently interprets the inapplicability of all predicates to God in the first manner ("any statement of the form 'God is *x*' is false.") This interpretation of the ineffability thesis suffers from the disadvantage of conflicting with the law of the excluded middle (since "God is *x*" and "God is not-*x*" are both going to be false). But this is not an argument that Alston deploys. His arguments make use only of the principle that for all predicates *x*, "God is *x*" is *not true*. This principle is a consequence of both the view that all predications are false when applied to God and the view that all predications are neither true nor false when applied to God. Thus Alston's arguments do not require us to choose between the two interpretations of ineffability. The availability of the two interpretations will come up again in my discussion, however.

Now there are two rather different circumstances that might impel one to claim that no predicates attach to God. One possibility is that we can't attach any predicates to God because His properties are all so extraordinary that they cannot be formulated. The second possibility is that we can't attach any predicates to God *because He has no properties*. Only in the first case does it follow that there are facts which cannot be expressed by linguistic means (the inexpressible facts being that God possesses the attributes which we are unable to formulate). If, as in the second case, God has no properties, then the fact that nothing can be predicated of Him doesn't entail that there are any facts which we are unable to express. On this second account, we might say nothing about God and still manage to express every fact, for there is nothing to *be* said about Him. It would be quite a challenge to tell a coherent story about entities that are devoid of any and all properties. I bring up this second gloss on indescribable entities only because many passages in the mystical literature suggest it. The second quotation from Stace, which says that God is nothing *to the intellect*, is pretty clearly

a claim of the first type—the implication is that the intellect is blind to what is patently there. But the first passage from Stace stipulates that we can form no concept of God *and that God has no qualities*, which seems to point to the second claim: we can form no concept of God because God has no qualities which might serve as candidates for representation. The same idea is also at least faintly suggested by Shankara's assertion that Brahman has no form *as well as* no name, and by Plotinus' implicit distinction between God's having no form "even . . . for the Intellect" and His having no form *tout court*. My inclination is to use "ineffability" to refer to the thesis that there are facts that defy expression. Thus I want to say that the second gloss, according to which nothing can be said of God because He has no properties to represent, doesn't have anything to do with ineffability. This is merely a nomenclatural point. In any event, Alston's critique of the concept of indescribable entities applies equally to both renderings.

Alston presents two arguments against indescribable entities. The more complex and less effective argument is that we can't be said to understand the term referring to the supposedly indescribable entity (say, "God") unless we have some way of identifying it. But any means of identification automatically provides us with a true description:

> I ask, "What do you mean by 'God'?" You might reply, "The first cause," or "The necessary being," or "The supreme mind holding moral relations with mankind," or "He who revealed Himself to the prophets," or "The father of Jesus Christ," or "The judge of our sins." If you were unable to give *any* such answer, wouldn't I be justified in concluding that you didn't understand the word "God" in any way? This means that a condition of your understanding any statement containing "God" is your capacity to supply some such identifying phrase, and any such phrase would constitute a predicate which could be attached to God. Hence "God is ineffable" asserts that an essential condition of its meaningfulness does not hold.
>
> (1972, 82–83)

One shortcoming of Alston's discussion of this point is that it is conducted in terms of an out-of-date, pre-Kripkean theory of reference (Alston's paper originally appeared in 1956). This is a minor point, however, since it's clear that the essence of the

objection is unaffected by the supposition that "God" is a rigid designator: any means we might be able to devise for dubbing God "God" is going to provide Him with a history, and that history is going to provide us with material for predicating various properties to Him. A more significant shortcoming of Alston's argument is that it presupposes that if a sentence (such as "God is indescribable") cannot be understood, then it cannot be true. Quantum mechanics might be cited as a prima facie counterexample to this principle. Cushing (1991), for instance, maintains that the principles of quantum mechanics provide us with a formalism that succeeds in explaining a broad range of phenomena, but that we nevertheless have no understanding of what that formalism says. I will not try to unravel this complex knot of ideas here (but see the subsection in Chapter 2 "Does epistemic boundedness entail ineffability?" p. 55). My point here is only that Alston would need to say more in order to secure the premises of his first argument. The third and most telling critique of the first argument is that it doesn't apply to the Ramsey-fied version of "God is indescribable," namely "Something is indescribable." This bloodless formula may not satisfy the religious fervor of the mystics, but it provides us with a species of ineffability claim which is immune to Alston's first criticism.

Alston's second, simpler, and stronger argument is that "God is indescribable" is self-defeating, for to ascribe the property of indescribability to God is already to give a description of Him. The contradiction is particularly obvious in the following passage by Stace:

> He is, in His very nature, unconceptualizable . . . His Mystery and incomprehensibility are absolute attributes of Him.
>
> (1952, 48–49)

Here Stace tells us directly that God possesses certain attributes, namely "His Mystery and incomprehensibility," for which perfectly adequate English terms exist. How, then, can it be maintained that He is "unconceptualizable"? The theological twist to the discussion is supernumerary. The argument in no way depends on the subject's being God or any other religiously significant entity. (The same can be said of Alston's first argument.) The charge of self-refutation applies to any claim to the effect that some being or process or feature or event or thing is indescribable. For example, A. W. Moore

makes the same argument against the putative ineffability of the absolutely infinite:

> if we cannot come to know anything about the infinite, then, in particular, we cannot come to know that we cannot come to know anything about the infinite; if we cannot coherently say anything about the infinite, then, in particular, we cannot coherently say that we cannot coherently say anything about the infinite.
>
> (1990, 12)

It was noted above that Alston's first argument against indescribable entities has no force against the Ramsey sentence, "Something is indescribable." The Ramsey sentence falls squarely within the scope of the self-refutation argument, however. The last sentence of Alston's paper suggests that the author may have missed this point:

> There may be something in the world which can't be talked about in any way, but if so we can only signalize the fact by leaving it unrecorded.
>
> (1972, 92)

To say this is already to say too much. Alston's own self-refutation argument shows us that there *cannot* be something in the world which can't be talked about in any way—for to say that something can't be talked about in any way is already to talk about it in some way.

Alston notes that proponents of indescribable entities may be able to deflect the force of both of his arguments by weakening their ineffability thesis. It's incoherent to claim that nothing can be said of God, but perhaps it can be maintained that there are certain *kinds* of things that can't be said of Him. Alston mentions several candidate-hypotheses which, he says, are logically unproblematic, yet may still qualify as a form of ineffability:

> you might wish to say, "We can speak only of extrinsic features of God, not of His intrinsic nature," or "God can never be characterized with the precision we can attain in science," or "We can speak of God only in a highly abstract way." None of these utterances need be self-defeating; for (1) in each case

> the sentence itself does not fall within the class of those declared impossible, and (2) a speaker or hearer can use a criterion for identifying God which does not involve attributing to Him a predicate of the sort which is ruled out.
>
> (1972, 91)

But what, exactly, does it mean for us to be incapable of speaking of God in any but an abstract way? This question brings us back to the two interpretations of ineffability presented near the beginning of this section. Let's agree that

(*J*) God was on the side of Joan of Arc against the English

is an example of the concrete sort of predication to which God is *not* susceptible. What, exactly, is being claimed about the status of *J*? Obviously, it's not going to be true. Does that mean that the negation not-*J* is true?—that God was *not* on the side of Joan of Arc? The only other possibility is to say that neither *J* nor not-*J* is true. This is evidently how Kellenberger (1979) would interpret Alston's limited-ineffability theses. Commenting on another proposal of the same type, he illustrates the idea that certain predicates are inapplicable to certain subjects as follows:

> The mathematical truth that the square of five is twenty-five is neither green nor not-green. Color concepts do not apply to mathematical truths.
>
> (1979, 310)

By the same token, Alston's limited-ineffability thesis would presumably imply that God was neither on the side of Joan of Arc nor not on her side. I'm not so sure that it isn't better to say that it's necessarily true that mathematical truths are not green and that God was not on Joan's side. The point I wish to emphasize, however, is unaffected by this choice. My point is that either way, there is nothing about the so-called limited-ineffability theses that serves to distinguish God from the mathematical truth that the square of five is twenty-five, or indeed from tables and chairs and every other run-of-the-mill entity in the universe. Under either interpretation, the limited ineffability thesis that non-abstract predicates don't apply to God is akin to the thesis that color predicates don't apply to mathematical truths, that psychological predicates

don't apply to tables and chairs, or that only noses can be snub. Everything in the universe is such that some predicates don't apply to it. There's nothing about this state of affairs that warrants the invocation of ineffability.

One might, I suppose, justifiably speak of an ineffability thesis if the class of inapplicable predicates were uncommonly broad. For example, it would surely be appropriate to say that God is ineffable if *all* predicates were to God as green is to mathematical truths or happiness is to chairs. But this is the hypothesis with which we started, and which Alston found to be self-defeating. Alston discusses another case of massive predicative inapplicability which might warrant being labeled an ineffability thesis: "God cannot be positively characterized in literal terms" (1972, 90–91). Kellenberger notes that this suggestion of Alston's corresponds closely to the mystical theology of Dionysius, according to which the only possible characterizations of the mystical object are negative: we can say what It is not, but we can't say what It is (1979, 310). The problem with this thesis, as both Kellenberger and Stace (1960, 134) point out, is that there is no principled way to make the distinction between positive and negative attributes. Rest is the absence of movement; but so is movement the absence of rest.

A recent proposal by John Hick (2000) might also be sufficiently restrictive to qualify as an ineffability thesis. Hick concedes that God is amenable to certain types of descriptions. For example, "in referring to anything, including God, we are attributing to it the characteristic of being able to be referred to" (41). Thus God is not absolutely ineffable. But, Hick continues, this conclusion

> is in itself a trivial truth in that nothing significant follows from it. It does however prompt us to distinguish between at least two kinds of attributes. There are what we can call substantial attributes, which would tell us something about what the Godhead in itself is like—for example, that it is personal or that it is impersonal. And there are what I have called formal attributes, which do not tell us anything about what the Godhead itself is like. Thus, for example, that it can be referred to does not give us any information about its nature. Formal attributes are thus trivial or inconsequential in that nothing significant follows from them concerning the intrinsic nature of the Godhead.
>
> (2000, 41)

In brief, Hick's limited ineffability thesis is that we can't ascribe any substantial attributes to God. This is, of course, itself an attribute of God. God is asserted to be the kind of being to whom nothing substantial can be attributed. Is this attribute in turn substantial or formal? It's clear that Hick is committed to saying that it's formal, for if it were substantial, we would have revived the self-defeating paradox that Hick's distinction was designed to avoid: God would be characterized by the substantial attribute of not being characterizable by any substantial attributes. But it's by no means obvious that the property of not being characterizable by any substantial attributes is merely formal in Hick's sense—i.e., that it does "not tell us anything about what the Godhead itself is like." The fact that the intrinsic nature of the Godhead is such as to defy substantial formulation surely tells us *something* about that nature. Hick might retort that unlike such characterizations as "personal" or "benevolent," the attribute of "not being characterizable by any substantial attributes" fails to describe an intrinsic feature of the object of discourse. But this is to beg the question: what is the difference between benevolence or greenness on the one hand, and the property of not being characterizable by any substantial attributes on the other hand that warrants their disparate treatment? This is not a rhetorical question. There may be a difference. But it wants spelling out.

In sum, the case for ineffability is in even greater disarray than Alston makes it out to be. Alston is right in claiming that the idea of an indescribable God or mystical experience is incoherent. But he concedes too much when he remarks that there may yet be something in the world which can't be talked about in any way. Moreover, the weakened ineffability theses that are designed to avoid self-stultification are either too mundane or too indeterminate in scope to count as ineffability theses. Are we already at the end of the story? Is it time to conclude that there is no ineffability? By no means. The analysis has shown that there is no entity that cannot be described. This result rules out a particular type of ineffability. But it doesn't rule out the more general possibility that there are, as my opening sentence speculated, truths or facts or states of affairs in the world which cannot be expressed or represented by linguistic means. Contrary to William James' famous discussion, the mystical experience, like everything else, is readily describable—as a "state of insight into depths of truth unplumbed by the discursive intellect" (James, 1902/1985, 380), as an experience, as a life-changing

experience, as an experience reported by Saint Theresa of Avila, etc. What the mystic cannot do, according to her own report, is to state the content of her insight. This point has been made by Moore:

> Some *insights* cannot in principle be expressed; they are ineffable, and they sustain ineffable knowledge. This formulation . . . allays the following natural worry about the whole idea of ineffability: that the very application of the concept is self-stultifying. There is nothing self-stultifying about describing someone as having ineffable knowledge, nor indeed about knowing and putting into words what is involved in their having it. What cannot be put into words is what they know.
>
> (1992, 427)

By the same token, there are true descriptions of every logically possible God (one such description is "logically possible"). But this doesn't rule out the possibility that there are facts about God that cannot be stated. I don't claim that the idea of an unstatable fact or insight is coherent—that remains to be seen. But at least it can be said that Alston's arguments against indescribable entities leave it unscathed.

The Tarskian approach

I approach the subject of unstatable facts from a broadly Tarskian perspective. I take a language to be an abstract system of syntactic and semantic rules that delimits a class *C* of sentences that are either true or false, and I assume that the semantic rules of the language associate a truth-condition *X* to each sentence *S* in *C* such that *S* is true in the language if and only if the condition *X* is satisfied. For example, the truth-condition for "Snow is white" in English is snow's being white. I take a *state of affairs* to be the same sort of things as a truth-condition, except that the former explicitly leaves it open, provisionally upon the result of further analysis, that there might be some states of affairs that are not associated to any sentence by any semantic rule. Thus we may say that "Snow is white" is true if and only if the state of affairs obtains that snow is white. A *fact* is a state of affairs that obtains. It's obvious that the existence of ineffable facts entails the existence of ineffable states of affairs. The converse proposition is not as entirely obvious—it

requires us to concede that to every ineffable state of affairs, there corresponds a complementary ineffable state of affairs such that either the initial state or its complement obtains. Be that as it may, the question whether there are ineffable facts subsumes the question whether there are ineffable states of affairs.

Tarski himself tentatively endorsed a thesis about natural languages which may, at first glance, seem to deny the possibility of ineffable states of affairs:

> A characteristic feature of colloquial language . . . is its universality. It would not be in harmony with the spirit of this language if in some other language a word occurred which could not be translated into it; it could be claimed that "if we can speak meaningfully about anything at all, we can also speak about it in colloquial language."
>
> (1956, 164)

Contrary to what may be suggested by the reference to the "universality" of natural languages, this thesis does not conflict with the hypothesis that there are ineffable states of affairs. The universality thesis claims that anything that can be expressed in any artificial language that we might contrive can also be expressed in natural, or "colloquial," language. This leaves it open that there might be truths that cannot be expressed in any of our natural languages (hence, if the universality thesis is right, also in any of our artificial languages).

Moreover, the universality thesis is not a necessary accompaniment to what I have been calling the Tarskian approach. One can without contradiction adopt the Tarskian approach while at the same time endorsing Whorf's (1964) view that some of the sentences of Hopi have no English equivalent. Indeed, it's clear that the Tarskian approach provides ample conceptual space for a notion of ineffability *within a specified language*. A state of affairs X is ineffable in language L if X fails to be a truth-condition for any sentence of L. Isn't it question-begging to suppose that such an X exists? Not, at least, when the orphaned state of affairs is a truth-condition of some sentence in *another* language. The fact that 2+2=4 is straightforwardly ineffable in the fragmentary language of arithmetic that contains only expressions for the natural numbers and the identity relation. It remains to be seen whether the Tarskian approach can accommodate the notion of a state of affairs to which

no sentence in any language, or in any humanly accessible language, corresponds.

The conceptual framework underlying the Tarskian approach is liable to a number of criticisms. Here are eight objections, roughly in order of how deeply they cut.

1 Languages as Tarski conceives them do not exist.
2 Natural languages are open-ended. Unlike Tarskian languages, they have no fixed repertoire of expressive devices.
3 Natural languages are deictic. This gives them expressive powers that Tarskian languages lack.
4 The Tarskian conception of effability applies only to the literal use of language. But the expressive power of natural languages is augmented by poetic, metaphorical, and other non-literal modes of speech.
5 The Tarskian conception of effability applies at most to the literal use of language. But it is impossible to distinguish literal from non-literal language.
6 Tarskian languages play no role in real-life communication.
7 Describing states of affairs, whether literally or figuratively, is merely one unprivileged language game among many others. Its analysis is not particularly important.
8 The very idea of describing states of affairs is incoherent.

I will try to deal with these objections below. A preliminary point: regardless of how well or poorly the Tarskian approach weathers its critique, a proper analysis of ineffability conducted on its terms will at the very least tell us how the ineffability issue plays itself out in the Tarskian arena. Even if the critic of the approach is unmoved by the impending softening-up operation, he or she will surely concede that the analysis can produce this kind of conditional information. The conditional results might even be of philosophical use to unreconciled critics of the antecedent assumptions. For example, if a consequence of Tarskian assumptions were found to be unacceptable on independent grounds, this result could be used as a reductio against the approach that produced it.

Here goes.

1 Tarskian languages do not exist.

The Tarskian approach equates languages with systems of syntactic and semantic rules abstracted from their actual users. On this account, it's meaningful to talk about languages that never have been and never will be spoken by any language-user. Richard Montague (1974), the prototypical practitioner of the Tarskian approach, specifies formal rules for a class of languages that includes languages with infinitely long sentences. It may be argued that languages conceived in this manner are devoid of philosophical interest —more on this below. Be that as it may, Tarskian languages exist as surely as numbers exist. One may interpret existence claims for numbers along either Platonist or formalist lines. But either way, there are true things to be said about them. Who would deny that 2+2=4? Moreover, these truths don't depend on anyone's stating or even entertaining the corresponding hypothesis. The product of two very large numbers is what it is even if no one ever tries to find it.

The same can be said about systems of syntactic and semantic rules. It's a truth about the system of rules that a Tarskian wants to call "English" that the sentence "Snow is white" is true in English if and only if snow is white. Now let "Frenglish" be the system of rules which is identical to the rules that comprise English, except that the nouns are all as in French. Then it's a truth about Frenglish that "Neige is white" is true if and only if snow is white. The fact that no one ever used Frenglish before and never will again doesn't change or invalidate the aforementioned fact. To be sure, the language of Frenglish has something of a history—see the previous three sentences. But it would still have been a fact about Frenglish that "Neige is white" is true if and only if snow is white even if this paragraph had never been written. In fact, one of the true things that can be said about Tarskian languages (a.k.a. systems of syntactic and semantic rules) is that there are some of them that never have been and never will be conceived by any sentient being. Frenglish just missed being one of them, as did Swenglish. Systems of rules are something—they're systems of rules. If Tarskian languages are going to succumb to philosophical criticism, it won't be on the ground of non-existence or incoherence. It will be on the ground that they're irrelevant or inadequate for the task at hand.

2 Natural languages are open-ended.

One can (the argument goes) introduce new expressive devices into English without its ceasing to be English. Moreover there is no definite, sharply marked degree or type of novelty beyond which a linguistic innovation results in our speaking a different language. This capacity for indefinite expressive extension is a characteristic of natural languages that sets them apart from Tarskian languages. You can't pin down a natural language to a determinate set of rules. Therefore there is no case of ineffability relative to a Tarskian language that entails ineffability in English. This objection was mounted by Tarski himself:

> [Natural] language is not something finished, closed, or bounded by clear limits. It is not laid down what words can be added to this language and thus in a certain sense already belong to it potentially. We are not able to specify structurally those expressions of the language which we call sentences, still less can we distinguish among them the true ones.
>
> (1956, 164)

Evidently, Tarski was not as wholehearted a Tarskian as Montague.

It seems to me that there is nothing more at stake here than a definitional issue. Let's concede that "natural languages" refers to open-ended systems. Let L be such a language. Now take a snapshot L_t of the indefinitely extensible language L at a moment in time t. By definition, L_t comprises all and only the expressive devices that have been explicitly introduced into L by time t. Since it isn't open-ended, L_t is not a natural language. Call it and other snapshots of L taken at different times natural *shmanguages*. I propose now to investigate the phenomenon of ineffability in various shmanguages and classes of shmanguages. To be sure, there is a particular philosophical interest in ascertaining whether there are ineffabilities in natural languages like L. But this question translates into the question whether there are ineffabilities in the set of all shmanguages that can be attained from the starting point of our current shmanguage. What does it mean to say that one shmanguage is attainable from the starting point of another? The same thing as Tarski means when he says that an as yet unexploited expressive device already belongs to the language "potentially." Both descriptions are amenable to a variety of interpretations. The aim of the next section

is to distinguish different grades of ineffability based on the extensiveness of the class of shmanguages which lack a sentence for the truth-condition in question. One of these classes, for example, is the class of shmanguages which it's nomologically possible for a human being to use (thus ruling out shmanguages with infinitely long sentences). The same project may also be described as an investigation of the grades of ineffability with respect to various more or less liberal senses in which either (1) one shmanguage may be said to be attainable from the starting point of another shmanguage, or (2) an expressive device belongs to a language [sic] potentially. One more thing: I hereby announce my intention to use the slightly shorter and much more euphonious "language" as an abbreviation for "shmanguage." This makes the term "language" ambiguous, but it's an ambiguity with which I've found it easy to live.

I will have more to say about the open-endedness of natural languages in the section on unrepresentability in Chapter 4.

3 Natural languages are deictic. This enables them to express facts that are beyond the range of any system of syntactic or semantic rules.

The idea is that indexical sentences like "This book is red" have no determinate truth-value—they're true on some occasions of their use and false on other occasions of their use. Thus what they express can't be captured by any Tarskian semantic rules of the form "Sentence *S* is true if and only if *X*." This is undoubtedly correct. But it doesn't yet show that we can describe states of affairs by means of indexicals that we can't describe without them. In order to obtain that result, we would additionally need to show that at least some indexical sentences on at least some occasions of their use are ineliminable, in the sense that they describe states of affairs which fail to be the truth-conditions of any non-indexical sentence in the language. Depending on our theory of the nature of time, the sentence "It is now three o'clock" might be ineliminable in this sense. If such ineliminabilities do exist, then indexical expressions will obviously extend the range of the effable beyond what can be captured by rules of the form "Sentence *S* is true if and only if *X*."

I propose to deal with this objection by granting it in its entirety and noting that it requires nothing more from us than a natural

extension of the notions of a semantic rule and of a truth-condition. Instead of saying that the semantic rules of a language specify truth-conditions for each sentence, we need to say that they specify truth-conditions for ordered pairs consisting of a sentence and an occasion for its use. Moreover, if some indexicals really are ineliminable, we might have to use them in describing the truth-conditions. In any case, the enterprise of devising formal rules for the use of indexical expressions has been actively pursued along these as well as other lines (e.g., Kaplan, 1989; Montague, 1974; Przelecki, 1983). So the core idea of pursuing the analysis of ineffability in languages conceived as abstract systems of rules isn't threatened by indexicals. The required liberalization of the notions of a semantic rule and of a truth-condition may result in a more generous assessment of the limits of effability. But that wouldn't be a problem for the analysis—it would be a result. The result might even be that nothing is ineffable in languages that contain certain sorts of indexicals. However, it will be shown in the section "Is the Tarskian criterion of ineffability vacuous?" p. 48, that the several quick and easy arguments to that effect are inconclusive.

4 The Tarskian conception of effability applies only to the literal use of language. But the expressive power of natural languages is augmented by the availability of non-literal modes of speech.

(I stipulate that proponents of the fourth objection are willing to grant that a distinction can be made between the literal and the non-literal use of language and that the Tarskian analysis deals adequately with the literal cases.) The exemplary formula

> "Snow is white" is true if and only if snow is white

does not, on the face of it, seem to provide an appropriate model for the interpretation of poetic, metaphorical, analogic, or ironic language. For example, consider the following sentence:

> Howard's relationship with Moira is an albatross around his neck.

It is doubtful that this sentence can be considered true on the condition that there is an albatross *A* such that *A* is around

Howard's neck, and Howard's relationship to Moira is identical to *A*.

The analysis of this objection turns on whether there are rules that govern the truth-conditions of figurative speech. To be sure, rules of the form

> "P" is true if and only if *P*

are not going to be forthcoming. But there is nothing in what I've been calling the Tarskian approach that restricts semantic rules to this format. The Tarskian format in its most general form is

> Sentence *S* is true if and only if *P*

or, with the liberalization contemplated in the discussion of indexicals,

> Sentence *S* on occasion *O* is true if and only if *P*.

There is no a priori constraint on the relationship between features of the sentence *S* and features of the metalinguistic expression *P* which is used to describe the truth-conditions for *S*. For example, if we can specify the occasions which call for an ironic interpretation of the sentence, then we can formulate the following semantic rule governing non-literal speech:

> "P" is true on ironic occasions if and only if it is not the case that *P*.

This is a perfectly acceptable Tarskian semantic rule. In brief, if there are rules that assign truth-conditions to figurative language, then the Tarskian notion of effability will not underestimate the range of the effable.

But what if there are no such rules? It may be claimed that there are no systematic conventions governing the truth-conditions of figurative language. "He is a warrior," said of someone who doesn't literally engage in warfare, may refer to his courage, or to his ruthlessness, or to the single-pointedness with which he pursues his objective. Moreover (on this view), what associates "He is a warrior" on a given occasion with the idea that he is courageous isn't a conventional rule—it's the fact that on this occasion, this

particular choice of words conveys the idea that he is courageous to the particular audience being addressed. This notion of "conveying" an idea is clearly a matter of probabilities and degrees: a given metaphor on a given occasion may very reliably succeed in conveying the intended message, while another metaphor may be chancier. There's no saying what a metaphor conveys on the basis of linguistic knowledge alone. A quirky association in one hearer's mind may make a metaphor apt for her that wouldn't work for anyone else. In principle, anything in the universe may potentially affect what is conveyed by non-literal speech.

A reminder: we're in the midst of considering the objection that a Tarskian analysis of ineffability will fail to take into account the possibility that some states of affairs may be effable in figurative language. But if figurative language is conceived of as in the previous paragraph, the kind of effing that can be done with it is radically different from the literal variety. On this account, a metaphor expresses a state of affairs *X* if it succeeds in conveying the message that *X* to the intended audience, where "conveying" means something like "causing to entertain." So a metaphor's effing of *X* is a causal fact, whereas a sentence's literal effing of *X* is a logical or conventional fact. It's difficult to resist the conclusion that the introduction of figurative effing is not so much an addition to literal effing as it is a change of topics. In fact, the notion of conveying a message slides seamlessly into nonverbal causings-to-entertain. If the aptness of a figure of speech is determined by the probability of its conveying the right thought, then the same job can be done by a judiciously timed slap in the face, or by the right sort of brain surgery. The fact that the speaker achieves her aim by availing herself of the conventional machinery of language is incidental to the proceedings, like scaring someone by employing the conventional "boo!" instead of an inarticulate scream.

If conveyability is allowed to count as a type of effability, then the issue of ineffability becomes the issue of what we can't be caused to entertain. This is a significant issue in its own right. The possibility that there are states of affairs the idea of which we cannot entertain has been discussed by several philosophers and cognitive scientists (e.g., Chomsky, 1975; Fodor, 1983; McGinn, 1989). This kind of "epistemic boundedness" (Fodor's term) is not unrelated to literal ineffability. But they're not the same issue. Ineffability is essentially about language, whereas language is only incidentally implicated in conveyability. When mystics claim to have an ineffable

insight, they must be drawing a distinction between effability and conveyability, for no human insight can be unconveyable—the fact that the mystic entertains it is already proof that humans can be caused to entertain it. Indeed, mystics do routinely attempt to convey their ineffable insight by figurative language, sudden yells, and even slaps (think of the carryings-on of the classical Zen masters). The mystic's claim to have an ineffable insight may be false or confused, but approaching the question with a conveyance notion of effability is a non-starter.

In sum, if figurative statements have truth-conditions, then the Tarskian approach includes their contribution to expressive power right from the start; and if figurative statements merely convey truths, then their causal efficacy doesn't count as a contribution to expressive power, any more than the comparable efficacy of judiciously timed slaps in the face. These two options—the *conventional* and the *causal* accounts of figurative meaning—strike me as satisfyingly complete. However, I have to admit that they're not logically complementary. There could be a third sense in which language might be said to express truths. In order to have an impact on the topic at hand, this third type of expressibility would have to be similar enough to literal description that we are willing to accept it as an extension of the notion of literal description, as opposed to a change of topics. This decision would to some extent be a matter of personal taste. I'm open to suggestions.

5 The Tarskian conception of effability applies at most to the literal use of language. But it is impossible to distinguish literal from non-literal language. Alternatively, there is no such thing as literal language.

This gambit presents us with the same pair of options as the previous objection. Sentences (or sentences-cum-occasions) either have truth-conditions or they don't. If they have truth-conditions, then presumably the claim that there's no such thing as literal language is a claim that the format of the rules associating sentences with truth-conditions are non-standard in some respect. I have no vested interest in making sense of this notion. I need only the conditional postulate that if the idea makes sense, then the Tarskian approach has no problem accommodating itself to it.

On the other hand, if sentences don't have truth-conditions, then, absent a tertium quid, what sentences express is what they convey,

whereupon the issue of ineffability evaporates in a cloud of inarticulate screams and slaps in the face. In the context of the previous objection, this conclusion applies only to figurative language. But if the objection is that there's no such thing as literal language, the evaporation leaves nothing behind. Unless some sentences have truth-conditions, there is no issue about ineffability to resolve. The apparatus of Tarskian semantics is revealed to be a hopeless attempt to achieve the doubly incoherent goal of providing a literal description of literal language.

Does this mean that it's a waste of time to engage in Tarskian semantics? Compared to what? If objectors of the fifth type are on the right track, then the notion that physics provides us, or even attempts to provide us, with a literal description of the physical world is just as misguided as the view that Tarskian semantics provides us with a literal description of literal language. It might be argued that physics, so conceived, isn't *doubly* incoherent—that at least its subject-matter (the physical world) exists. But this is small consolation at best. One incoherence is just as debilitating as two. Yet no one would want to derive from this state of affairs the moral that everybody should stop doing physics. It's just a certain *metastory* about physics—that it provides us with a literal description of the physical world—that has to be abandoned. The same can be said of Tarskian semantics. The fifth objection is sometimes formulated as the thesis that all writing is literature. The moral of this putative insight isn't that we should cease to write Tarskian semantics. It's that Tarskian semantics is literature. On this view, Tarskian semantics is a poetic allegory of language. Perhaps this comes to the same thing as saying that Tarskian semantics provides us with an idealized model of language. In any case, it's on a par with physics, which is good enough for me.

6 Tarskian languages play no role in real-life communication.

I have in mind Donald Davidson's provocative thesis that "there is no such thing as a language, not if a language is anything like what many philosophers and linguists have supposed" (1986, 446). It's Davidson's view that the meaning of an utterance is determined by the convergence of a joint process of mutual interpretation conducted by speaker and hearer. What this view denies is that meanings are given by pre-established semantic rules upon

which both speaker and hearer rely for mutual comprehension. For Davidson, any word can mean anything, given the right occasion. If this is right, then English, conceived as a system of semantic and syntactic rules, plays no role in the process of communication. But it doesn't follow that *languages*, conceived as systems of semantic and syntactic rules, play no role in communication. It's just that the language upon which speaker and hearer rely for mutual comprehension is an idiolect, the rules of which are valid only for a particular speaker on a particular occasion. Davidson grudgingly concedes this:

> We could hold that any theory on which a speaker and interpreter converge is a language; but then there would be a new language for every unexpected turn in the conversation, and languages could not be learned and no one would want to master most of them.
>
> (1986, 445)

A language of this type is indeed a very different affair from languages as hitherto conceived by "many philosophers and linguists." But what is the import of these differences to the topic of ineffability? The import is that ineffability-in-English, where English is conceived to be a fixed system of syntactic and semantic rules, is an artificial and philosophically uninteresting notion that doesn't tell us anything about what is ineffable in actual conversational practice. The more significant notion is ineffability relative to idiolects. To be sure, the discovery that a particular state of affairs is ineffable in a particular idiolect would, in itself, be of limited theoretical import. But it's prima facie possible that one might be able to establish that some states of affairs are ineffable in a theoretically interesting *class* of idiolects, such as the class of idiolects that are accessible to human beings. In sum, Davidson's thesis has a bearing on which specific systems of semantic rules it's important and interesting to think about; but it doesn't affect (or at least it isn't known to affect) what there is to be said about the relationship between systems of rules and states of affairs in general. There are other aspects of Davidson's philosophical views that create substantial problems for the concept of ineffability. These will be discussed in due time. The present point is only that Davidson's interpretative theory of communication doesn't add to these problems.

7 Describing states of affairs, whether literally or figuratively, is merely one unprivileged language game among many others. Its analysis is not particularly important.

This is, of course, the Wittgensteinian objection. Language is viewed here as a vast and heterogeneous collection of social practices that are designed to achieve a dazzling variety of ends. One of these (more likely, a family of these) is the game of stating facts. But it's a myth about this game that it has some sort of centrality that makes it especially deserving of philosophical attention. Stating facts is no more special than making promises or delivering insults.

I find myself in agreement with this sentiment. I wish only to add the following pair of observations. First, this Wittgensteinian perspective doesn't render the analysis of ineffability worthless. It merely shears it of overweening pretentions. I would find a good account of the game of delivering insults to be just as fascinating as a good account of the game of stating facts. Having said that, I do wish to lay claim to a limited sort of specialness for my topic. It isn't privileged by any absolute, transcultural standards of philosophical significance. But it is privileged by history. We're talking about the game of *scientific discourse*, in the broadest sense of the term. There was a time in the history of the West when the (de facto) privileged language game was the game of biblical citation and commentary. At that time, considerations pointing to the limits of what can be accomplished within the parameters of that game would have made for particularly important and interesting reading. For the past few centuries, the (de facto) privileged game has been the scientific rationality game. I see Tarskian semantic theory as providing some of the ground rules for this admittedly historically contingent game. It seems to me that if a case can be made for ineffability within the context of this apparatus, we would have obtained a limitative result on the scope of scientific discourse that would be worth reading about at this particular time in the history of ideas.

8 The very idea of describing states of affairs is incoherent. The project of ascertaining whether there are states of affairs that can't be described is therefore doomed to failure.

This is the point of view taken by Davidson in his seminal paper on conceptual schemes (Davidson, 1974). Briefly stated, Davidson's objection is that talk of describing states of affairs presupposes an untenable distinction between conceptual *schemes* and the *content* that they represent. What is ultimately at stake here is the status of the correspondence theory of truth. Is this the same point as the "no literal–figurative distinction" point of the fifth objection? Maybe so. But I want to leave open the possibility that there are ways of characterizing the literal–figurative distinction that don't depend on the correspondence theory. In any event, Davidson's objection is so distinctive and so influential that it requires a separate and more extensive treatment. To do it justice, however, I've found that I need to make some crucial distinctions between the grades of ineffability.

Four or five grades of ineffability

Weak ineffability

A fact is *weakly ineffable* if there is no sentence for it in one or more of the languages which some human beings actually speak, or have spoken, or will speak. A fact may be weakly ineffable with respect to one language but straightforwardly effable with respect to another—at least so far as the definition goes. But the effability or ineffability of a given fact with respect to a given language is the same for everybody. There is a maximally strong form of weak ineffability—universal weak ineffability—wherein every human language that was ever or will ever be devised lacks a sentence for a particular fact. But this is still a case of merely weak ineffability.

An example of a weak ineffability claim is Whorf's (1964) thesis of the untranslatability of some Hopi sentences into English. As Davidson (1974) points out, Whorf undermines his own claim by using English to convey the supposedly untranslatable content of the Hopi sentences. Still, Whorf's thesis gives us an example of the sort of issue in which weak ineffability claims are implicated. If the Whorfian untranslatability claim were true, there would be states of affairs—those that comprise the truth-conditions of the

untranslatable Hopi sentence—to which no sentence of English corresponds. These "Hopi states" would be weakly ineffable in English. Kuhn's (1962) claim that the conceptual schemes of new scientific paradigms are incommensurable with the schemes of the paradigms they replace has the same logical form as Whorf's. It's also sunk by the same Davidsonian torpedo: "Kuhn is brilliant at saying what things were like before the revolution using—what else?—our post-revolutionary idiom" (1974, 6).

Are there any examples of weak ineffability that are not susceptible to this easy refutation? Sure there are. One example was discussed in the section on the Tarksian approach (p. 11): the fact that 2+2=4, while easily statable in the standard language of arithmetic, is utterly ineffable in the language which consists only of expressions for the natural numbers and the identity relation. The more interesting question, of course, is whether there are ineffabilities with respect to complete natural languages.

Suppose Whorf were right about the untranslatability of Hopi into English. Then he wouldn't be able to tell us, in English, what the untranslatable Hopi sentence says. How then could we monolingual English speakers ever come to know that the corresponding Hopi state is weakly ineffable in English? We would have to rely, of course, on the testimony of bilingual speakers of English and Hopi that the Hopi sentence has no English translation. But how do the bilinguals come to obtain this item of knowledge? One possibility is that they make an abductive inference to the best explanation. The data being explained are the failures of the bilinguals' persistent attempts to find an English equivalent for some Hopi sentence. The explanation endorsed by the bilinguals for their persistent failure is that there is no English equivalent to be found. Other explanations are also tenable—e.g., that English equivalents exist, but that the shortest one is longer than the longest sentence which the bilingual speakers are able to process. Still, the bilingual's inability to translate the Hopi sentence into English would provide a measure of support for the thesis that the corresponding Hopi state is weakly ineffable in English. It's one of several prima facie plausible explanations of why the Hopi sentence defies translation.

Human ineffability

A fact is *humanly ineffable* if there is no sentence for it in any language that it's nomologically possible for human beings to use,

regardless of whether that language is ever actually spoken by human beings. It goes without saying that human ineffability entails weak ineffability, but that weak ineffability doesn't entail human ineffability.

The weak ineffability of a state of affairs, if it's sufficiently pervasive, may provide us with defeasible evidence for its human ineffability. The route to human ineffability is somewhat convoluted, however. Suppose we can establish for some fact *F* that *F* is weakly ineffable in English, and in French, and in Chinese, and in Hopi, and in Old Church Slavonic, and in Sumerian. In fact, suppose that we check hundreds of current and extinct languages and find that there is no sentence for *F* in any of them. This would provide us with defeasible but conventionally reasonable inductive grounds for adopting the general proposition that all actual human languages lack a sentence for *F*. This is the proposition that was dubbed "universal weak ineffability" in the previous subsection. But what accounts for this universal property of all actual human languages? One ready explanation is that using a language in which *F* can be expressed requires cognitive machinery that the human mind simply lacks. Given that *F* is universally weakly ineffable, this is a very compelling hypothesis. Its denial would amount to saying that we humans *could* devise and speak a language in which *F* is expressible, but we have never chosen to do so and never will. This is a possible state of affairs, but it cries out for an alternative explanatory hypothesis.

But how can we establish that *F* has no sentence in any actual present or past language? The problem isn't our practical inability to make an exhaustive survey. The same inductive projection that carries us from past and present languages to languages yet to be devised will also carry us from the examined past and present cases to the unexamined ones. The problem is that the sort of evidence that might persuade us of the existence of a weak ineffability in a single language—the testimony of bilinguals—is unavailable for the hypothesis that there is a fact which is inexpressible in every language that we have examined. The bilingual evidence is that there is a sentence in one of the bilingual's languages which has no translation in the second of the bilingual's languages. This type of evidence requires that the weakly ineffable fact be expressible in one of the bilingual's languages. Hence it has no purchase on the hypothesis that there is a fact which cannot be expressed in any of the languages that we have examined.

There may nevertheless be possible data that can be brought to bear on the hypothesis that there is a fact that cannot be expressed in any human language. However, the circumstances for obtaining the relevant data are going to be difficult to bring about. We need to be in contact with a language-using race of extraterrestrials—call them Plutonians—and the Plutonians need to have mastered all the human languages that are going to be in our inductive base. These conditions are almost sufficient for our extraterrestrial multilinguals to play the same role vis-à-vis human ineffability as terrestrial bilinguals play vis-à-vis weak ineffability: they provide us with creditable testimony to the effect that there are sentences in Plutonian which have no equivalents in any of the examined human languages. This much evidence gives us rational grounds for believing that there are Plutonian facts that are weakly ineffable in all examined human languages—hence, by induction, in all actual human languages. But to conclude that the Plutonian facts are humanly ineffable, we need one further item of information. We need to know that Plutonian cannot be used by human beings—that there is a nomological incompatibility between the innate structure of the human mind and the structure of the Plutonian language. Presumably, a reasonable case could be made for this hypothesis if the best and brightest of humanity's minds strove mightily for centuries to master Plutonian, and they all failed.

On the other hand, there is a theoretical argument based on Chomskian linguistic theory which suggests that it may be nomologically impossible to obtain the data needed to make the case for human ineffability. Chomsky famously argues that the data which are available to language learners are insufficient to enable them to figure out the grammatical rules of the language. No being, however intelligent, could parlay the linguistic data available to language learners into knowledge of grammar by a process of rational inference. Now this claim can be interpreted in either of two ways. The *weak* sense of the claim is that children *as a matter of fact* don't receive enough linguistic data to effect a rational reconstruction of the grammar; the *strong* sense is that children *can't even in principle* obtain enough data to reconstruct the grammar. (For a discussion of these interpretative options, see Kukla, 2001, 181–190.) Chomsky himself almost always relies on the weaker claim (e.g., Chomsky, 1980, 1986). However, he does make the stronger claim on some occasions (e.g., Chomsky and Fodor, 1980). The strong claim is based on an application of Goodman's (1954) new riddle of

induction: there are infinitely many grammars that are consistent with *any* finite set of linguistic data that a language learner might possess; thus, unless infinitely many candidates can be ruled out a priori, the chances of a language learner hitting upon the correct grammar are nil. Chomsky's theoretical solution is to suppose that we are innately directed to restrict our search to a finite set of alternative grammars—those that satisfy the constraining principles of "universal grammar" (UG). Now the same strong Goodmanesque argument can be run on the Plutonians: they wouldn't be able to learn their language unless *their* search through the space of possible grammars were innately constrained by *their* UG. But there's no reason to believe that our UG is truly universal, as opposed to merely planetary. Formally speaking, our UG is an unexceptional member of an infinite class of grammatical schemata, all of which could have played the same constraining role in facilitating language learning. Suppose now that there are no biological reasons why our UG is better suited to playing this role than infinitely many of its competitors. Then it's going to be effectively impossible for it to turn out that members of evolutionarily unrelated species deploy the same grammatical schema, which means (if the strong argument is sound) that they will be unable to learn each other's language. It follows that we will never encounter extraterrestrials who can tell us what we need to know in order to make an inductive case for human ineffability.

If, however, the strong Chomskian argument is sound, doesn't it give us an alternative and even quicker route to the conclusion that there are human ineffabilities? For the argument that leads us to conclude that the Plutonians will never be able to learn English also leads us, mutatis mutandis, to the conclusion that we humans will never be able to learn Plutonian—and doesn't that conclusion establish the existence of human ineffability? No it doesn't. For one thing, there may not *be* any Plutonians. But suppose there are. Let's say that a language which cannot be learned by X is *indecipherable* for X. There are grades of indecipherability just as there are grades of ineffability. The strong Chomskian argument purports to show us that Plutonian is going to turn out to be "humanly" indecipherable—i.e., it's going to be nomologically impossible for any human beings to learn it. This incapacity is logically entailed by the thesis that there are Plutonian facts that can't be expressed in any humanly accessible language. For suppose that there are such humanly ineffable Plutonian facts, but that Plutonian is *not* humanly

indecipherable. Then there *is* a language that human beings can learn in which the Plutonian facts can be expressed: Plutonian. But the converse proposition doesn't hold. The human indecipherability of Plutonian doesn't entail that there are humanly ineffable Plutonian facts. The strong argument for indecipherability in no way depends on there being such ineffable facts. It may be assumed that all the sentences of Plutonian and all the sentences of English can be placed in one-to-one correspondence, assuring that there are no humanly ineffable Plutonian facts. What makes Plutonian indecipherable according to the strong argument is that regardless of how much linguistic information we may amass, there are always going to be an infinite number of such one-to-one correspondences from which to choose. In sum, human indecipherability doesn't entail human ineffability. In the present context, the moral is that the thesis of human ineffability can be confirmed by extraterrestrial data only if Chomsky's innateness hypothesis is false. This doesn't, of course, rule out the possibility that there's some other way to confirm the human ineffability thesis that isn't incompatible with Chomsky's theory.

Nomological ineffability

A fact is *nomologically ineffable* if there is no sentence for it in any language that any nomologically possible being can use. The nomologically ineffable is ineffable on Pluto and in Andromeda as well as on Earth. Hence the convoluted extraterrestrial scenario which would arguably provide empirical evidence for human ineffability is of no help when it comes to nomological ineffability. Is there any other probative scenario for nomological ineffability? How about the following? The Plutonians might be able to establish that there are Plutonian ineffabilities, i.e., facts that cannot be expressed in any language accessible to the Plutonian race. If (contra Chomsky) they can manage to master one of our human languages, they might be able to impart this item of information to us. If we humans had already established that there are human ineffabilities, we would then have the beginning of an inductive argument to the effect that every nomologically possible type of mind is afflicted with limitations that render some facts inexpressible in any language which that type of mind can use. The generalization would increase in plausibility with the number of extraterrestrial species who told us that they suffered from species-specific ineffabilities.

Is this, however, nomological ineffability? It might be argued that the inductive generalization stills fall short of establishing the desired result. If it is sound, the induction tells us that for every nomologically possible race *R*, there is a fact that can't be expressed in any language that's accessible to *R*. But nomological ineffability is the stronger thesis that there is a fact that can't be expressed in any language that's accessible to *any* nomologically possible race. To conclude that there are nomological ineffabilities, we would have to ascertain that *one and the same fact* is inexpressible in any language which is accessible to any nomologically possible being. This critique is correct as far as it goes. But if the inductive argument is sound, then we can arrive at the requisite conclusion in just another few steps. Suppose that the inductive argument is accepted. Then for every nomologically possible type of being, there is a fact which cannot be expressed in any language that it could possibly use. Consider now the enormous conjunctive fact *F* consisting of the fact that defies expression in any human language, together with the fact that defies expression in any Plutonian language, and so on for every nomologically possible type of being (relative to some exhaustive taxonomy of beings). *F* would defy expression in any and all nomologically possible languages. It would be nomologically ineffable.

Unfortunately, the inductive argument itself is defective. The problem is that the inductive base will not serve as a random sample of the population of nomologically possible language users. For us humans to learn that the Plutonians suffer from Plutonian ineffabilities, either the Plutonians have to be able to master a human language, or we humans have to be able to master a Plutonian language. The same is true for any extraterrestrial race whose data could make its way into the inductive base. We could not obtain the relevant data from races for whom the class of accessible languages had no overlap with the class of humanly accessible languages. We've already seen that there are Chomskian reasons for doubting that we and the Plutonians could find a common language in which to exchange the relevant information. But set that theoretical problem aside. Even so, the requisite inductive projection would be based on a skewed sample. If there are going to be races whose cognitive operating characteristics differ markedly from our own, it stands to reason that they're more likely to belong to the unrepresented races with whom there is no possibility of communicating.

In the end, I don't see how one might be able to make an inductive case for nomological ineffability. But that doesn't mean that there can't be another type of argument for it. What might such an argument look like? Well, suppose it could be shown that there are facts that can be expressed only by infinitely long sentences. Perhaps this would establish a nomological ineffability on the ground that no nomologically possible beings—not even the Andromedans—are going to be capable of using languages that contain infinitely long sentences. Something like this argument seems to underlie Nietzsche's (1979) claim that language is incapable of encoding the whole of the truth. Nietzsche argues that one leaf is never exactly the same as another. Thus when we bring all leaves under the umbrella of the term "leaf," we must discard the differences that make each leaf unique. These unique properties that set each individual leaf apart from all the others are therefore ineffable. Of course, some of these properties may be captured by other concepts. This is where the appeal to infinity comes in: if each individual leaf possesses infinitely many properties, then (the argument goes) no system of finitely many concepts will be able to exhaust them. There will always be an unexpressed residue, "an X which remains inaccessible and indefinable for us" (1979, 83). This conclusion is not yet the thesis of nomological ineffability, for to say that every finite system of concepts fails to capture the whole of the truth is not yet to say that there's a part of the truth that can't be captured by any finite system of concepts. The fact that we can't say *everything* that's true about leaves in any language doesn't mean that there is *anything* that's true about leaves that we can't say in *some* language or other. Or so one might suppose at first blush. But we've already seen (two paragraphs above) how this difficulty can be finessed: the part of the truth that can't be expressed in any finite system is the biggest part of all, namely the *whole* of it. If Nietzsche's argument works, the conjunction of all the facts about this leaf is a nomologically ineffable fact.

The main weakness with Nietzsche's argument is that the skeptic need not accede to the assumption that the properties of individual leaves cannot be exhausted by a finite set of concepts. Why should we believe that the world is literally infinitely variegated? Maybe a thousand or a million or 10^{100} concepts are all that's needed for an exhaustive description. So long as the number is finite, it becomes uncertain whether mastery of the requisite conceptual repertoire is beyond the reach of any nomologically possible being. It's possible

that Nietzsche intended his conclusion to apply only to human beings—that the "us" for which *X* remains indefinable refers only to humans. This would make his claim more plausible, although he would still owe us an argument for supposing that the number of requisite concepts is greater than we human beings are designed to handle.

Of course, the fact that the Nietzschean argument for nomological ineffability fails doesn't mean that there aren't any nomological ineffabilities, much less that the idea of nomological ineffability is incoherent. A considerably stronger variant of Nietzsche's argument could be run on the natural numbers instead of leaves. Gödel has shown us that the set of all truths about the natural numbers cannot be finitely specified. The infinitely enormous total truth about numbers is therefore nomologically ineffable. For mathematical Platonists, this argument for nomological ineffability is apodictic. However, non-Platonists may still be plagued by some qualms. They might, for example, balk at the suggestion that a purely mathematical example settles the issue whether there are truths *about the world in which we live* that cannot be represented in any nomologically accessible language. This is not a view that I'm inclined to endorse. I merely indicate it as a dialectical possibility.

Weak logical ineffability and logical ineffability tout court

A fact is *logically ineffable* (*tout court*) if there is no sentence for it in any logically possible language. It may prove useful to distinguish logical ineffability from the closely related notion of *weak* logical ineffability. A fact is weakly logically ineffable if it can't be expressed in any language that can be used by any logically possible being, regardless of whether there are sentences for it in some logically possible languages. Weak logical ineffability obviously entails nomological ineffability: if a fact can't be expressed by any logically possible being, then it can't be expressed by any nomologically possible being. (Any uncertainty that we may have concerning the boundary between the logical and the nomological—e.g., because of Quinean scruples—will of course be inherited by the weak logical ineffability/nomological ineffability distinction.) Just as obviously, weak logical ineffability is in turn entailed by logical ineffability—*tout court*: if a fact can't be expressed in any logically possible language, then it can't be expressed by any

logically possible being. The converse entailment may also be true —that is to say, it may be the case that any logically possible language can be used by some logically possible being. But this hypothesis doesn't strike me as obvious. Absent an argument for the equivalence of the two types of logical ineffabilities, it seems a good idea to keep them provisionally distinct.

Suppose, once again, that some fact *F* can only be represented by infinitely long sentences—only now suppose that this is so in any logically possible language. This would provide us with a good reason for believing that there are nomological ineffabilities. With a few additional premises, this example can be turned into a comprehensible (albeit fatally flawed) case for weak logical ineffability. The premises are (1) that it's a logical truth that only infinite beings are capable of using languages that contain infinitely long sentences, (2) that there can be no more than one infinite being, and (3) that Wittgenstein's private language argument is sound. Premises 1 and 2 together entail that there is at most one being who can express *F* in any language. Premise 3 entails that unless two or more people use a particular language, nobody uses it. The conclusion is that no being can express *F* in any language. This is weak logical ineffability rather than logical ineffability *tout court*, because the private language argument places no restrictions on the range of languages that exist as abstract structures. It tells us only that one person can't use a language to express or describe facts without there being at least one other person who also uses the same language. So far as the argument based on premises 1, 2, and 3 goes, there may be a language in which *F* is represented by an infinitely long sentence, but no logically possible being could ever speak such a language.

How does this argument for weak logical ineffability fare? Well, premise 3 is a live philosophical hypothesis. Premise 1 has at least the authority of St. Augustine behind it. Augustine believed not only that God is infinite, but also that, by virtue of His infinity, He can think infinite thoughts. He can, for example, entertain all the truths about numbers (1947, XII 18). The idea is that only an infinite being can be equipped with the infinitely capacious expressive machinery that would be needed to produce an infinite sentence. Premise 2 is the weakest link of the argument. There have been numerous arguments in the history of philosophy for the uniqueness of God. But none of these rules out the possibility of sub-deistic beings who have the infinite capaciousness needed for using infini-

tary languages. For example, material beings who are infinite in one spatial dimension might be able to inscribe infinite sentences by the parallel activity of infinitely many fingers. But in an unbounded three-dimensional Euclidean universe, there's ample room for infinitely many beings each of whom is infinitely long but only three meters in circumference. I present this refutation of a bad argument that no one has ever proposed because I think that its contemplation gives us reason to keep an open mind about the availability of *good* arguments for ineffabilities of a very high order.

It may be argued that we will never be able to establish that there are logical ineffabilities *tout court*, for the same reason that it's impossible to prove the Church–Turing thesis. The thesis that every effective procedure is Turing-computable can't be proved because there's no way of exhaustively characterizing the class of all effective procedures. We may have a proof that every effective procedure that we have *thought* of is Turing-computable. But it's always conceivable that somebody—perhaps a Plutonian—will come up with a new one. There's no way of knowing that we have the whole class of effective procedures covered. The class of all logically possible languages may be indeterminate in the same way. There may be no way to characterize all the devices that can be used to pull off the trick of representing the world. A similar argument can be run on weak logical ineffability and the concept of a logically possible mind.

I don't know how I feel about these Church–Turingesque arguments myself. I'm not even fully convinced that the Church–Turing thesis is beyond all hope of a proof. Be that as it may, we can still hope to show that there are facts which are ineffable in classes of languages that are larger than the class of all nomologically accessible languages. I distinguish a nested series of such classes below. The catch is that I can't characterize the series without availing myself of a concept that I'm not at all sure I have a grip on. So I present these closing thoughts about more-than-nomological ineffability as amusing speculations that I'm not prepared to defend. The problematic concept is that of a *conceivable* language. To say that a language is conceivable is to say less than that it can be used. I can conceive of a language with infinite sentences, although neither I nor any other nomologically possible being would be able to use such a language. There are three degrees of conceivability, each of which serves to delimit a class of languages which may be bigger than the class of all nomologically accessible languages but smaller

than the class of all logically possible languages. These are (1) the class of languages that nomologically *can* be conceived by a human being (human conceivability); (2) the class of languages that can be conceived by any nomologically possible being (nomological conceivability); and (3) the class of all languages that can be conceived by any logically possible being (logical conceivability). It strikes me as prima facie possible that we may encounter empirical or theoretical reasons for supposing that there are facts which are ineffable in all humanly conceivable languages. With a little help from our extraterrestrial friends, we might even be able to broaden the conclusion to all nomologically conceivable languages. But the third class—the class of all languages that can be conceived by any logically possible being—is coextensive with the class of all logically possible languages, for surely there is a logically possible God who can conceive of every logically possible language. In fact, unless His linguistic prowess is limited by the private language argument, He's probably fluent in every one of them. In any case, if God is able to conceive of every logically possible language, then an argument for the existence of ineffabilities relative to the class of all languages that can be conceived by any logically possible being is as little to be hoped for as an argument for logical ineffability. End of amusing speculation.

Untranslatable languages

Ask a contemporary Anglo-American philosopher to comment off the cuff on ineffability, and her first thought will probably be that Davidson's arguments in "On the very idea of a conceptual scheme" (1974) show it to be incoherent. There are two arguments at issue: (1) that there are no languages that cannot be translated into English, and (2) that the distinction between conceptual schemes and the unconceptualized content that the schemes organize is incoherent. The prima facie connection between these theses and the topic of ineffability is easy to see. The hypothesis that there are human or higher ineffabilities entails that there are states of affairs which we humans cannot express in any language; but such a state would, at least on the face of it, have to be considered an instance of unconceptualized content. One could endorse the thesis of weak ineffability (with respect to, say, English) without appealing to a scheme–content distinction—this would require no more than our being able to state a sentence in, say, Hopi, that has no English

equivalent. But the argument that there are no untranslatable languages blocks even this possibility.

Before subjecting this prima facie case to further analysis, I wish to clarify the place of arguments (1) and (2) in Davidson's philosophical agenda. Davidson has no independent interest in (1) (henceforth, the translatability thesis). The philosophical aim of his paper is to establish (2) (henceforth, the incoherence thesis). But Davidson claims that the translatability thesis *entails* the incoherence thesis, hence that the proof of the former gives us the latter as a corollary (1974, 185, 190). In addition, he advances another argument for the incoherence thesis that doesn't depend on the translatability thesis. So for Davidson, arguments (1) and (2) are both refutations of the scheme–content distinction.

My first critical point is that the argument that (1) entails (2) suffers from a debilitating lacuna. Davidson tells us that conceptual schemes are to be identified with equivalence classes of mutually translatable languages (1974, 185). It follows that the possibility of alternative conceptual schemes entails the possibility of untranslatable languages. Davidson claims that the putative proof of the translatability thesis thus secures the desired result about schemes by modus tollens. But for Davidson's purpose, the indicated modus tollens doesn't yet show enough. Granted that it shows there are no *alternative* conceptual schemes, there's still a gap between this conclusion and the thesis that the scheme–content distinction is incoherent. The result that there are no alternative conceptual schemes is also consistent with the hypothesis that English happens to be able to express everything that's expressible in any language, but that there are features of the world that are inexpressible in any language. In other words, a sound demonstration of the translatability thesis purchases no more than Tarski's universality thesis. Of course, Davidson thinks that the notion of an inexpressible feature of the world is empty. But that is what he's supposed to be proving as a consequence of the translatability thesis. My point is that the argument begs the question. The argument doesn't go through unless one adduces independent reasons for repudiating the possibility that there are inexpressible feature of the world. To be sure, Davidson does eventually adduce such independent reasons. But that's the *second* argument for the incoherence thesis, which hasn't yet been touched on. The *first* argument for the incoherence thesis, according to which this thesis is a consequence of the untranslatability thesis, doesn't work.

Note that this critique of Davidson's first argument also constitutes a refutation of the prima facia argument against weak ineffability that was given at the beginning of this section. The argument was that if every language can be fully translated into English, then nothing is weakly ineffable with respect to English. The foregoing remarks show that this is a non sequitur: the fact that every language can be fully translated into English doesn't rule out the possibility that there are states of affairs that can't be expressed in any language, a fortiori in English. So the translatability thesis alone doesn't establish the non-existence of weak ineffabilities with respect to English. Of course, if Davidson's second, as yet unbroached, argument against unconceptualized content succeeds, then there would still be a case against human or higher ineffability. Moreover, the *conjunction* of the incoherence thesis and the translatability thesis would once again provide us with grounds for denying the possibility of even weak ineffability—for if everything that's expressible in any language is translatable into English *and* if everything *is* expressible in some language, then everything is expressible in English. Conversely, if it can be shown that there *can* be untranslatable languages, we would have an argument for the coherence of weak ineffability which doesn't depend on the coherence of the scheme–content distinction. Either way, there's a lot riding on the status of the translatability thesis, to which I now turn.

Is it possible that intelligent extraterrestrials might employ a language which is untranslatable into our own? Davidson rejects this possibility on the grounds that finding a workable interpretation of a putative language is the only evidence we can have that a corpus of utterances (or tentacle-waggings) *is* a language. If we can't translate Plutonian speech into English, then we have no reason to believe that the Plutonians are communicating at all. It has been claimed —by Fodor (1983, 124), among others—that this argument is blatantly verificationist. This counterargument makes the soundness of Davidson's argument depend on the status of verificationism. I think the argument is defective for less contentious reasons. Granting any type of verificationism that Davidson requires, there still are good reasons for rejecting this particular verificationist argument. In fact, I will claim that some of Davidson's other views —notably, his anomalous monism and meaning holism—strongly commit him to the possibility that there can be evidence for the hypothesis that the Plutonians have a language which is untranslatable into our own.

Let's consider the hypothesis that Plutonians have a language, and that their language is utterly untranslatable into our own. Does it follow from this hypothesis that there can be no evidence of their using a language? Davidson seems to presume that the ascription of a language to a society of beings can only be justified by displaying the semantics of the putative language. But there is nothing in Davidson's theory of truth and interpretation that would rule out the possibility of *indirect evidence* for the hypothesis that the Plutonians have a language. This point has been made by Schick, who provides us with the following scenario:

> There are criteria other than translatability that we could appeal to to establish the existence of an alternate conceptual scheme. If creatures from another planet flew to the earth in sophisticated spaceships and proceeded to transform the planet by building complex structures, for example, then even if we could not translate their speech, we would still have good reason for believing that they possess an alternative conceptual scheme.
>
> (1987, 37)

In this passage, Schick refers to alternative conceptual schemes; but his remarks apply, *mutatis mutandis*, to untranslatable languages. The evidence to which Schick refers is not particularly strong. After all, his aliens might perform their technological feats without the benefit of a communication system, relying only on innate technological expertise, as terrestrial ants and bees do. For all that, his example is good enough for the task it was designed to accomplish: the possibility of weak evidence is sufficient to block the conclusion that a hypothesis is strictly unverifiable. Nevertheless, I prefer to base my arguments on a different scenario—one which Davidson is *required* to accept as probative, on pain of contradicting some of the most central aspects of his philosophical views.

Suppose we encounter both Plutonians and Saturnians when we get Out There. We can make no sense of Plutonian communication (hence, Davidson would say, we can have no reason to suppose that they are communicating), nor is there any evidence that they can make sense of our language. The Saturnians, however, are able to learn English without difficulty. It may be recalled that there is a Chomskian argument to the effect that such interspecific communication is nomologically impossible. But nothing here hinges on the

extraterrestriality of the Saturnians and the Plutonians. I might as well have supposed that we encounter Tasmanians of whom we can't make sense, and a race of Hungarians who learn English with no difficulty. I talk about Plutonians and Saturnians to make my discussion directly comparable to Davidson's, who also employs an extraterrestrial scenario involving Plutonians and Saturnians. Once again, it is supposed that we encounter a race of Saturnians who have no difficulty learning English. Having learned it, they seem to use it very much as we do. They agree with us on such fundamental questions as whether there is presently a book on the table before us, whether 2+2=4, and so on. But they also tell us a curious thing: they claim that the Plutonians also have a language. How do they know? Because they've mastered it—they speak Plutonian with the same fluency as they speak English. When we ask them to teach it to us, however, they protest that this would be impossible, for English and Plutonian are not mutually translatable. Certainly these events are logically possible—it is not self-contradictory to suppose that we will encounter Saturnians who tell us just these things, at least in the *acoustical* sense of telling. If they do tell us these things, what should we make of them?

I want to say that these events would provide us with very strong evidence for the proposition that the Plutonians speak a language that cannot be translated into English. In fact, I already *did* say so in my discussion of weak ineffability in the previous section: I noted that the untranslatability of Hopi into English could receive support from the testimony of bilinguals. There are, however, a few dialectical escape routes available to supporters of the translatability thesis which I would like to block.

The main Davidsonian escape route is clearly marked. In fact, Davidson has discussed this very scenario:

> We should have to ask how we recognized that what the Saturnian was doing was *translating* Plutonian (or anything else). The Saturnian speaker might tell us that that was what he was doing, or rather we might assume that that was what he was telling us. But then it would occur to us to wonder whether our translations of Saturnian were correct.
>
> (1974, 186)

In Davidson's example, the Saturnian is not bilingual, but rather speaks a language similar enough to English that we can understand

it, but different enough that Plutonian can be translated into it. Nothing turns on whether we consider beings who are bilingual and can communicate with us in only one of their languages, or whether we consider beings whose single language is partially translatable.

Let's refer to the language that the Saturnians use when they talk to us as "Saturnlish." Davidson's point is that it can't be assumed a priori that Saturnlish is identical to English. To be sure, the Saturnians say things that are acoustically indistinguishable from English sentences like "The Plutonians have a language," and "Plutonian can be translated into Saturnian." But how do we know that these sentences have the same truth-conditions in Saturnlish as they do in English? Perhaps the best theory of Saturnlish is one that ascribes the following truth-conditions:

> "X has a language" is true (in Saturnlish) if and only if X is not a Plutonian and X has a language, or X is a Plutonian and X does *not* have a language.

On this interpretation, the Saturnians are not telling us that the Plutonians have a language at all—quite the contrary! If we use enough gruish devices of this kind, it is undoubtedly possible to contrive an interpretation of Saturnlish that doesn't yield the conclusion that Saturnians are telling us that the Plutonians have a language that cannot be translated into English. Let's call such an interpretation a *Davidsonian* theory of Saturnlish. The contrast is between Davidsonian theories and the *standard* theory of Saturnlish, which is the one that says that when the Saturnians say "P," they mean that *P*.

How shall we decide whether the Davidsonian theory is better or worse than the standard theory? According Davidson's own principle of charity, our interpretations are constrained by the requirement to ascribe the maximal amount of truth to others' utterances. As Davidson often tells us, we make maximum sense of the words and thoughts of others when we interpret in a way that optimizes agreement. Let's provisionally assume that we have no prior opinion about whether the Plutonians have a language, or whether all languages can be translated into English. In that case, the principle of charity will not help us to make a choice, since both the Davidsonian and the standard theory agree in their interpretations of Saturnlish pronouncements on all matters not relating to the linguistic capacities of Plutonians. In a situation like

this, we must presumably make our choice on the basis of relative simplicity. Considerations of simplicity are also the reason why neither Davidson nor his critics take heed of the following resolution to the question of whether there are untranslatable languages: given any finite set of putative linguistic data, there is mathematically bound to be a totally charitable interpretation of the whole corpus. You can even translate finite segments of brook-babbling into sensible English. Therefore there are no untranslatable languages. Presumably, we aren't ready to allow just any mapping into English to count as a *proper* translation. The question here is whether there are any reasons to suppose that every language must have a proper translation into English. It's difficult to imagine how one might argue that the Davidsonian theory of Saturnlish is simpler than the standard theory. If Davidson wishes to resist the conclusion that the Saturnians tell us that the Plutonians have a language, he will have to reject the assumption that the standard theory and the Davidsonian theory are equally charitable. Since both theories agree on all matters *except* whether the Saturnians tell us that the Plutonians have a language which is incommensurable with English, the greater charitability of the Davidsonian theory requires the assumption that it's absurd to suppose that Plutonians *do* have a language which is incommensurable with English. This is indeed how Davidson replied when I presented my version of the Saturnian–Plutonian scenario to him personally at a symposium some time ago. He replied that he would use Hume's argument against miracles in this situation: the improbability of the Saturnians' claim about the Plutonians is so great that it would be more plausible (and more charitable) to suppose that they mean something other than what the standard interpretation stipulates (or that they are simply lying, or engaged in an elaborate hoax—Davidson can play with the Saturnians' communicative intentions as well as with their semantics —the arguments are the same in either case). It seems to me that this reply puts the philosophical cart before the horse. The implausibility of untranslatable languages is being used as an argument against accepting their possibility. But it is only Davidson's argument that makes untranslatable languages implausible in the first place. So far as I know, no one expressed great skepticism about untranslatable languages before Davidson's argument called them to question. But now it seems that the argument against them relies on their *prima facie* implausibility. The stage is set for one of those infamous clashes of intuition that stop discussion short: I maintain that it is only

Davidson's argument that makes us think of the idea of an untranslatable language as bizarre; and Davidsonians say—would *have* to say—that the idea is bizarre quite independently of Davidson's argument. What do we do now? Fortunately for my side, I have another argument—one which I think a Davidsonian will be unable to resist.

Let's change the scenario so that we humans play the role of the Saturnians. We discover that there are Saturnians and Plutonians, and that they both have languages. How do we know? Because we have no difficulty translating either language into English. We find, however, that Saturnian and Plutonian are incommensurable *with each other*. In this situation, we do not have to evaluate the testimony of anybody else. Appeals to Hume's argument are therefore not in order. We know that Saturnian and Plutonian are mutually incommensurable *directly*. The question, of course, is whether such a thing can actually happen. Also, supposing that it can happen, would that mean that Davidson's thesis is false?

The first question has been extensively discussed by Robert Kraut (1986). Kraut was interested in establishing the thesis that we can have a distinction between scheme and content even if Davidson is right about the impossibility of untranslatable languages. I wish to establish the reverse proposition—that there can be untranslatable languages even if we repudiate the scheme–content distinction. Our two arguments pass over some of the same philosophical territory, however. In particular, Kraut notes that it's easy to devise languages that have less *expressive power* than English. His point is that something very much like the scheme–content distinction gives us a way to describe the relation between these languages and English which is at once useful and unobjectionable from the viewpoint of Davidson's theory of interpretation. This point will be important in the next section. For the purpose of this section, what is important about Kraut's observation is the point that there can be a language the best interpretation of which maps all of its sentences into a proper subset of the sentences of English—i.e., that a *fragment* of English can also be a language. If this is correct, then there is no resisting the conclusion that there can be two sublanguages of English E1 and E2 such that the proper subset of English sentences corresponding to E1 has no overlap with the proper subset of English sentences corresponding to E2—i.e., that two sublanguages of English may be mutually untranslatable. And *that* means that our extraterrestrial scenario, wherein Saturnian and Plutonian

correspond to just such incommensurable fragments of English, is a conceptual possibility.

Is Kraut right, however? He notes that Davidson has addressed the subject of language fragments. The issue comes up in relation to the criticism, attributed to Solomon, that Davidson's own anomalous monism commits him to the existence of incommensurable languages, such as those of commonsense psychology and of mathematical physics—"for the predicates of the one framework are nomologically irreducible to those of the other, yet in some cases they characterize the very same events" (Kraut, 1986, 410). Davidson's reply to Solomon is that such fragments could not constitute the entirety of one's language: "We cannot conceive a language without psychological terms or expressions—there would be no way to translate it into our own language" (1980). In order to preserve the thesis that there can be *no* mutually untranslatable languages, it's clear that Davidson's retort in this case must be made across the board for any dichotomization of English: it's true that we can delineate incommensurable fragments of a language, but these fragments do not qualify as *languages* in their own right.

In responding to this argument, Kraut asks us to consider whether or not we can make sense of "an outlaw community of Rortyian eliminative materialists, who have systematically purged their language of mental predicates": "Perhaps, on closer scrutiny, it isn't clear that we really can envisage such non-standard and attenuated conceptual repertoires; but neither is it clear that we can't" (1986, 410).

In sum, Kraut regards the case for Davidson's reply to Solomon as unproven. Davidson's position on this issue seems to be based on his accepting the constraint that the interpretative scheme for a language *L* must provide counterparts in *L* for every English sentence. I agree with Kraut that this constraint is radically undermotivated by the rest of Davidson's views. Indeed, I would further claim that this rule has bizarre and unacceptable consequences. In order to assess the partitionability of English into sub-languages, let's change the extraterrestrial scenario in what are really inessential ways. Instead of talking about Saturnians and Plutonians, let's consider children learning English. We begin with the observation that learning to speak English takes quite a bit of time. The question is: what shall we say of the intermediate stages at which the child is talking, but has not yet mastered all of the language's expressive resources? What is happening when two five-year-olds seem to be

engaged in a conversation? Certainly one cannot claim that the five-year-old's putative language fails to be a genuine language on the grounds that it cannot stand alone. It does stand alone. It seems to me that we're compelled to recognize that children have a language before they have mastered all the expressive resources of English—which is to say that fragments of English can be languages.

There is, of course, a Davidsonian attack on the coherence of the scenario I have just presented. According to Davidson's thesis of meaning holism, it's impossible for children to learn the language in discrete pieces—first the physicalist vocabulary, then the mentalist vocabulary, etc.—because the meaning of each piece is determined by its relation to all the other pieces. Thus children have to learn the whole language before they can be said to have mastered any part of it. If holism is true, then the language of children will not be *English*. Indeed, if holism is true, there is a sense in which children do not even speak a *fragment* of English (since English cannot be regarded as a straight sum of a number of discrete parts). But there's nothing in these considerations that would impel us to deny that the child is using a *language*. Holistic considerations don't save Davidson's thesis about the impossibility of untranslatable languages. On the contrary, holism seems to provide us with a quick route to its refutation: if holism is true, then the language of the child is strictly incommensurable with that of the adult. QED.

What if, in order to avoid the quick refutation, we concede that the language of the child is a fragment of English? Then the only way to avoid the conclusion that there can be incommensurable languages is to deny that children can learn incommensurably *different* fragments of English. That is to say, one has to defend the view that the expressive resources of English are acquired in an invariable developmental sequence, so that all children go through exactly the same intermediate stages of language acquisition. Now it seems plausible that there is *some* order of precedence that must be followed with respect to the acquisition of linguistic resources. But Davidson has to defend the enormously stronger thesis that it's impossible, or at least overwhelmingly unlikely, that two divergent developmental paths leading to the mastery of English can ever go through incommensurable way-stations. Perhaps this is so. But it certainly isn't obvious that this is so. It's incumbent upon Davidson to persuade us that it is so.

In sum, we have found no reason to reject the possibility that we may be able to translate Saturnian and Plutonian into mutually

incommensurable parts of English. It follows that there can be mutually untranslatable languages. But do we have to admit that there can be a language which is incommensurable with our language? Kraut thinks that a Davidsonian *cannot* take this view, apparently on the grounds that the admission that there are languages that we cannot translate "threaten[s] the connection between the notion of translation and the notion of truth" (1986, 403). As far as I can gather, what Kraut has in mind is that to say that a language is untranslatable is to say that we cannot give Tarski-style truth-conditions for its sentences, and that this display of biconditionals *is* Davidson's definition of truth. If this were so, then the idea of a language that we cannot translate would indeed be incoherent for Davidsonians. But Davidson has himself warned against this prevalent misapprehension about his views. He tells us that the problem with the view that Tarski's truth predicate captures the concept of truth "is due simply to the fact that Tarski's definitions give us no idea how to apply the concept in a new case, whether the new case is a new language, or a word newly added to a language" (1990, 287). Thus "definitions like Tarski's . . . cannot capture the concept of truth" (1990, 288). Truth has to be taken as a "primitive" notion (1990, 308). As a consequence, the possibility can't be ruled out a priori that there are truths which are inexpressible in English—or in any other language, for that matter. That is to say, the hypothesis of weak ineffability has not been shown to be incoherent.

Inexpressible facts

Let's start with a recapitulation of the results of the previous section. Contra Davidson, the thesis that every language is translatable into English (the translatability thesis) does not entail the incoherence of the scheme–content distinction (the incoherence thesis). The translatability thesis also doesn't entail that there are no weak ineffabilities. However, the non-existence of weak ineffabilities is entailed by the *conjunction* of the translatability thesis and the incoherence thesis. But, contra Davidson, the translatability thesis is itself unsupported. This means that there are no arguments against weak ineffability on the table. Note that this result does not in any way depend on the coherence of the scheme–content distinction: there are no Davidsonian arguments against weak ineffability even if we grant Davidson his incoherence thesis. Without a

scheme–content distinction, we cannot equate weak ineffability in *L* with the existence of a fact that cannot be stated in *L*; but we can still equate it with the non-existence of a translation in *L* for *S*, where *S* is a sentence in a language other than *L*.

There's still Davidson's *second* argument for the incoherence of the scheme–content distinction to be considered—the one that doesn't depend on the translatability thesis. Without the assistance of the translatability thesis, the incoherence of the scheme–content distinction does not provide us with any reasons for skepticism about weak ineffability. But it would establish the incoherence of *nomological* ineffability (and perhaps of human ineffability as well, though I won't pursue this topic here). For the thesis of nomological ineffability is the thesis that there are facts which cannot be expressed in any nomologically possible language—and if we want to say that there are facts that are inexpressible in any language, then it would seem that these facts must constitute pure, unconceptualized content.

I have two things to say about Davidson's incoherence thesis. First, in mounting his first and already refuted argument for the incoherence thesis, Davidson inadvertently provides us with sufficient (and unrefuted) material for a proof that there *are* alternative conceptual schemes, a fortiori that the scheme–content distinction is coherent. Recall that Davidson had equated conceptual schemes with classes of mutually translatable languages. His strategy in the *first* argument against the scheme–content distinction was to show that there could be no untranslatable languages, thereby securing the non-existence of alternative conceptual schemes by modus tollens. This argument was refuted in the previous section—it was shown that the translatability thesis does not, after all, entail the incoherence thesis. But nothing was said that would call to question the entailment in the other direction—that if there *are* untranslatable languages, then there are alternative conceptual schemes. That is to say, the translatability thesis is not, contra Davidson, a sufficient condition for the incoherence thesis; but Davidson himself avers that it is a necessary condition for the incoherence thesis. Thus the refutation of the translatability thesis, which took up the bulk of the previous section, is also a refutation of the incoherence thesis.

Does this settle the issue? Not entirely. There's still Davidson's second argument for the incoherence thesis to be considered. This argument makes no use of the identification of conceptual schemes

with mutually translatable languages which stands at the center of the first argument. This leaves it open for Davidson to jettison the identity altogether. With it would go his own first argument for the incoherence thesis, which doesn't work anyway, as well as my counterargument *against* the incoherence thesis as outlined in the previous paragraph—a trade-off that Davidson would be well-advised to accept.

The second thing to be said about the incoherence thesis is that Davidson's second argument for it is inconclusive at best. What is the second argument? It's that the putative distinction between scheme and content is empty:

> The trouble is that the notion of . . . fitting the facts, or of being true to the facts, adds nothing intelligible to the simple concept of being true. . . . Nothing, . . . no *thing* makes sentences and theories true. . . . *That* experience takes a certain course, that our skin is warmed or punctured, that the universe is finite, these facts, if we like to talk that way, make sentences and theories true. But the point is put better without mention of facts. The sentence 'My skin is warm' is true if and only if my skin is warm. Here there is no reference to a fact, a world, an experience, or a piece of evidence.
>
> (1974, 193–194)

In sum, the scheme–content distinction stands accused of doing no useful work. It's a candidate for shaving with Ockham's Razor.

Davidson's remarks here and elsewhere about the vacuity of facts may succeed in showing that an appeal to facts doesn't advance our understanding of truth. To say that "My skin is warm" is true if and only if it's *a fact* that my skin is warm is not to say anything more than that the sentence is true if and only if my skin is warm. But to establish that appealing to facts doesn't advance our understanding of truth is not yet to show that they perform no useful function whatever (this point has been made by Richard Manning, 1998). Moreover, I have a suggestion as to what other theoretical function facts might serve: *they can be used to explain what's going on in the extraterrestrial scenarios of the previous section*. This point is an elaboration of Kraut's. Davidson tells us that "My skin is warm" is true if and only if my skin is warm—that no reference to facts is needed. Now it's already been established that we may have evidence to the effect that some sentence *S* in the Plutonian language

has no translation in English. In that case, there is no way to complete the following sentence:

S is true in Plutonian if and only if . . .

where the continuation is an English sentence, after the model of "My skin is warm." So what should we English speakers say about *S*? Well, the Saturnians, who have proven to be reliable informants, will feel strongly impelled to tell us that *S* expresses a fact in Plutonian which cannot be expressed in English. When they speak amongst themselves, the Saturnians don't need to make any reference to the facts relating to *S*, any more than we need to allude to facts when stating the truth-conditions of "My skin is warm." They can just state the truth-conditions of *S* directly in Saturnian. But they can't do that when they speak to us in English. What, then, can they say to us when we ask them about *S*? On the one hand, they can't provide us with a direct translation; on the other hand, it would be incorrect for them to tell us that *S* expresses nothing at all. Once again, they could tell us is that *S* expresses a fact that can't be stated in English. Perhaps there's something else they could say instead. But we've found at least a provisional theoretical role for facts to play. Davidson's argument that facts should be eliminated from our ontology because they don't buy us anything is inconclusive.

The foregoing considerations result in our (provisional) acceptance of the idea that there are facts that can't be expressed in English. This is not yet nomological ineffability. In fact, it's nothing more than the thesis of weak ineffability, against which there was no surviving argument anyway. What has newly been accomplished is that we now have the explicit license to formulate the weak ineffability thesis in terms of *inexpressible facts*, and not just in terms of untranslatable sentences. But once we have the concept of a fact that can't be stated in our language, there is nothing to stop us from hypothesizing that there may be facts that can't be stated in any nomologically possible language. To begin with, having admitted that there may be facts that are inexpressible in specified languages, we surely have to concede that there may be facts that can't be expressed in any language that ever was or will be used by any sentient beings—the latter are, after all, only a very large list of specific languages. This is not yet nomological ineffability. But on what grounds would we refuse to take the last step of averring the

coherence of facts that can't be expressed in any *possible* language? Such a refusal would be completely unmotivated. These considerations don't constitute a positive argument for the coherence of nomological ineffability, but they do de-fuse the main argument for its *in*coherence. I'll settle for the verdict of not proven either way.

This completes my defense of the coherence of the idea of ineffability against pre-existing objections. In the next section, I take up a new objection, generated in-house by the details of my own analysis.

Is the Tarskian criterion of ineffability vacuous?

Pre-analytically, a state of affairs is said to be ineffable if no sentence "expresses" or "formulates" or "states" or "represents" that state of affairs. I've been using all these locutions interchangeably. When I've had to be precise, however, I've relied on the following explication: *a state of affairs* A *is ineffable in a language or class of languages* L *if there is no sentence* S *in* L *such that* A *is the truth-condition of* S. Let all this be the Tarskian criterion of (in)effability.

Here's a problem for the Tarskian criterion. Suppose I get an insight into an allegedly ineffable fact. There is nothing to stop me from referring to my insight as "this insight." But then the sentence "This insight is true," if used on the occasion of having the allegedly ineffable insight, has the same truth-condition as the insight itself. And then, by the Tarskian criterion of effability, the insight has been effed—a sentence has been found the truth-condition of which is the alleged ineffability. The conclusion is that there are no ineffable insights in English or in any other language that allows for the appropriate demonstrative construction.

When confronted with this argument, the first thought of a partisan of the Tarskian criterion will be to make the criterion more restrictive by discounting the use of indexicals. The defensive idea is that referring to a state of affairs by means of an indexical sentence is not to count as effing that state of affairs. But the dilemma for the Tarskian criterion doesn't depend on the use of indexicals. Consider the following non-indexical sentence:

> (S) The first ineffable insight to have occurred after midnight GMT on February 18, 1576 is true.

The sentence *S* is true (relativistic complications aside) if and only if the state of affairs which is the object of the aforementioned insight obtains. But then that state of affairs is the truth-condition of a perfectly good sentence of English—namely *S*—which in turn means, according to the Tarskian criterion, that the insight isn't ineffable after all.

At this point, a defender of the Tarskian criterion might try to retrench. It might be admitted that the argument shows that there cannot be any ineffable *insights*. But this admission doesn't yet entail that there can't be any ineffable *states of affairs*. The latter claim is weaker than the former: if there are ineffable insights, then there must be ineffable states of affairs—namely the states of affairs which are the objects of the ineffable insights. But the converse doesn't hold. There may be ineffable states of affairs which are totally beyond our cognitive horizons, so that we're incapable not only of representing them linguistically, but also of having a wordless intuition of them. Constructions like sentence *S* may establish that there aren't any ineffable intuitions, but they leave open the possibility that there are ineffable states of affairs.

In support of the foregoing defensive maneuver, it may be noted that the demonstrative device that produced a problem in the case of ineffable insights doesn't seem to work as well in the case of ineffable states of affairs. Confronted with an allegedly ineffable insight, the mystic can say apparently eff it by saying "This insight is true." The corresponding construction in the case of an allegedly ineffable state of affairs would be "This state of affairs obtains." Once again, the argument would be that "This state of affairs obtains" is true if and only if the indicated state of affairs obtains, and thus that the indicated state is a truth-condition of an English sentence, which in turn means that the state isn't ineffable after all. But which state of affairs is the "indicated" one? In the case of "This insight is true," the identity of the insight is relatively unproblematic—it's the insight enjoyed by the utterer of the sentence at the time of utterance. But which state of affairs is "this" one, in "This state of affairs obtains"? What the demonstrative is pointing to in this case seems entirely indeterminate.

However, as was noted above, the dilemma for ineffable insights doesn't depend on indexicals. The next question is whether some non-indexical sentence analogous to sentence *S* can be constructed which would be true if and only if an allegedly ineffable state of affairs obtains. The pursuit of the answer quickly embroils us

in tiresome technicalities which turn out to be inessential to my ultimate position on these problems. My ultimate position is that even the case against ineffable insights is fallacious. And if this circle of considerations is unable to vanquish the stronger thesis that there are ineffable insights, it's a fortiori not to be expected that it will be able to unseat the weaker thesis that there are ineffable states of affairs. So let me cut to the chase.

Let's refine the statement of the problem. The apparently problematic Tarskian criterion of ineffability tells us that a state of affairs is ineffable if there is no sentence for which the state of affairs is a truth-condition. Suppose that the state of snow's being white is ineffable in a language which is just like English, except that it lacks the means for representing colors. Suppose now that person P's one and only language is this truncated English, and that her one and only ineffable insight is that snow is white. Now speakers of the language can't say "Snow is white"; but they *can* say "P's ineffable insight is true." But for the sentence "P's ineffable insight is true" to be true, snow just has to be white. So snow's being white is the truth-condition for the sentence "P's insight is true." But then snow's being white is effable after all—by "P's insight is true." The conclusion is that there are no ineffable insights in this truncated English.

What about insights that are ineffable in English? In that case, we obviously can't say what the ineffable insight is. But we can parallel the reasoning. Let P's one and only ineffable insight be that an ineffable state of affairs *K* obtains. Once again, we can say "P's insight is true." But for P's ineffable insight to be true is for *K* to obtain. Therefore *K*'s obtaining is the truth-condition of an English sentence after all, namely "P's insight is true." But that means that *K isn't* ineffable. Therefore there are no insights that are ineffable relative to English.

If these arguments were to be valid, they would show either that there are no ineffable insights or that the Tarskian criterion of ineffability is inadequate. My intuition would strongly incline me in the second direction. Be that as it may, the arguments are not valid. I will run my refutation on the argument purporting to show that snow's being white is still effable in the truncated English that's bereft of color words. The problem with the argument is that snow's being white is not the truth-condition of "P's insight is true," even if P's insight is that snow is white. The truth-condition of "P's insight is true" is arguably the infinitely long disjunctive condition that

snow is white and P's insight is that snow is white, or that grass is green and P's insight is that grass is green, or that the sky is blue and P's insight is that the sky is blue, or. . . . But that's by no means equivalent to the condition that snow is white. It happens to be true that "P's insight is true" is true if and only if snow is white—but that's a contingent, material equivalence, not a conventional semantic rule. Here's an analogy that makes the point clear. The sentence "Box A contains 9 marbles" is true if and only if the number of marbles in the box is equal to the number of planets. But the number of marbles being equal to the number of planets isn't a truth-condition of "Box A has 9 marbles."

The same story can be told about ineffability in English: while it's true that "P's insight is true" is true if and only if *K* obtains, it doesn't follow that *K*'s obtaining is the truth-condition for "P's insight is true." The Tarskian criterion of ineffability isn't vacuous after all.

2 Mysticism, epistemic boundedness, and ineffability

Most of those who advance the thesis that there are ineffable states of affairs base their view on the putative fact that we (or some of us) possess knowledge that can't be put into words. I call this the *argument from mysticism*. Mysticism is generally thought of as the doctrine that some people have ineffable knowledge of specifically religious matters, such as the nature of God. But there is a broader usage according to which any knowledge that can't be represented linguistically qualifies as mystical. Rucker (1982) and Moore (1990), for instance, regard the thesis that we are able to grasp certain inexpressible set-theoretical truths as a species of mysticism. In any case, so far as the argument from the possession of inexpressible insights to the existence of ineffable states of affairs goes, the subject-matter of the insight is irrelevant.

The argument from mysticism will be evaluated in the second section of this chapter. The first section will be devoted to an entirely different route to ineffability. The mystical route traverses a terrain of materials and issues that are exotic to the scientific sensibility. In contrast, the route to be discussed in the first section runs through the heartland of contemporary cognitive science. The point of departure is the familiar cognitive-scientific thesis of "epistemic boundedness" (Fodor, 1983) or "cognitive closure" (McGinn, 1989). The destination, after a not very arduous journey, is the thesis of ineffability. Very roughly, the idea of this *argument from epistemic boundedness* is that human minds have limitations on what they can think; and what we can't think, we can't say.

It's noteworthy that the two arguments for ineffability run in opposite directions. The mystical argument supports the ineffability thesis on the basis of certain extraordinary *capacities* of the human mind: the reason we're supposed to believe in the ineffable

is that we (or some of us) are capable of grasping truths that defy formulation. In contrast, the argument from epistemic boundedness supports the ineffability thesis on the basis of certain of the mind's *in*capacities: it's because there are purported to be facts that the mind *can't* grasp that we should believe in the existence of an ineffable realm. Both arguments need the premise that the idea of an ineffable state of affairs is coherent; but the argument from mysticism *additionally* needs the premise that the idea of ineffable *knowledge* is coherent. Now if there is or can be ineffable knowledge, then there are or can be ineffable states of affairs —namely, those states which are the objects of the ineffable knowledge. But the coherence or existence of ineffable states of affairs doesn't ensure the coherence or existence of ineffable knowledge. In this respect, the argument from mysticism takes on a greater argumentative burden than the argument from epistemic boundedness. The coherence of the weaker thesis that there are ineffable states of affairs is defended in Part 1. In this part, it's *presupposed* that there can be ineffable states of affairs, and the stronger thesis is defended that there can also be ineffable states of knowledge.

The argument from epistemic boundedness

Here's the structure of the section. First I'll expand a bit on the concept of epistemic boundedness. I'll base my discussion primarily on Fodor's (1983). Second, I'll argue that epistemic boundedness entails ineffability. Finally, I'll evaluate three arguments for epistemic boundedness itself. I'll conclude that at least one of these arguments gives us good reasons to suppose that the mind is epistemically bounded. If this is right, then the ineffability thesis follows by modus ponens—the argument from epistemic boundedness goes through.

The concept of epistemic boundedness

What, exactly, does it mean to say that a mind is epistemically bounded? According to Fodor, "a psychological theory represents the mind as *epistemically bounded* if it is a consequence of the theory that our cognitive organization imposes epistemically significant constraints on the beliefs that we can entertain" (1983, 120). Fodor

distinguishes epistemic boundedness from the "boring" constraints due to parametric limitations:

> One could imagine that the computation that would select the right hypothesis is too long for the system to perform given available resources of memory, attention, etc.; or that . . . the best hypothesis contains too many clauses (in canonical notation) for the device to parse; or that the critically relevant data base is too complex for the device to represent. . . . Perhaps solving the riddle of the universe requires one more neuron than, de facto, anyone will ever have.
>
> (121–122)

The intended difference between boring constraints and genuine epistemic boundedness is that the latter is conceived to be a *de jure*, nomologically imposed limitation. The fact that solving the riddle of the universe requires one more neuron than any *actual* human being will ever have does not yet entail that the solution is beyond human epistemic bounds. For a hypothesis to be beyond human epistemic bounds, it has to be unthinkable by any nomologically *possible* human being. There is, of course, some conceptual slack in what counts as a possible human being. Can a being with a brain as big as a house still be a human being? Maybe so. But a being with a silicon-based nervous system, or with no recognizable nervous system at all—one where all information processing is carried out by chemical reactions occurring in a sack of solvent—is surely beyond the human pale. In any case, there's nothing in this section or in Fodor's analysis which depends on where, exactly, the boundaries of humanity are drawn.

Fodor illustrates how the class of accessible hypotheses may be nomologically constrained with Hume's theory of mind:

> the class of beliefs that can be entertained according to Hume's account is perhaps more sharply delimited than any modularity theorist has ever proposed. This is because the class of accessible *beliefs* is determined by the class of accessible *concepts*; and, for Hume, the class of accessible concepts is determined by the Empiricist principle; there are no concepts except such as can be constructed from sensations. So, in particular, if the hypotheses of the best science were to be such as to make reference to God, or to electrons, or to triangles, or to mental faculties, or to any

> other unobservables, then the best science is humanly inaccessible on Hume's account; it is beyond the epistemic bounds that Hume posits.
>
> (123)

Of course, this isn't the way Hume saw it. As Fodor notes, Hume derived the aforementioned empiricist *psychological* principle from the empiricist *semantic* principle according to which all logically *possible* concepts are constructible from sensations. Thus for Hume the empiricist psychological principle doesn't impose any epistemic constraints after all: if semantic empiricism is right, there is no logically possible concept that the Humean mind can't entertain. But even if the empiricist theory of meaning were to be false, there might still exist beings whose minds work in accordance with the empiricist psychological principle. Such minds would be epistemically bounded.

Does epistemic boundedness entail ineffability?

Does epistemic boundedness entail ineffability? The connection seems obvious: if we can't *think* that *P*, then surely we can't *say* that *P* either. Equivalently, if we *can* say that *P*, then surely we can think it. But is this so? It would trivially be so if thinking and talking were identical processes—if, as J. B. Watson (1925) believed, thinking were literally a matter of subvocally talking to oneself in a public language. Even if thinking and talking were distinct processes, ineffability would still be an immediate consequence of epistemic boundedness so long as the thinking process and the talking process were obviously guaranteed to have the same range of possible outputs. This theoretical option is nicely illustrated by David Cole's (1999) suggestion that thinking that *P* is a matter of entertaining an auditory or kinesthetic image of saying that *P*. Having an image isn't the same thing as talking, but it's clear that we can have an image of ourselves saying anything that we can *actually* say, and that we can say anything that we can imagine ourselves saying. If Cole's view of the relation between thinking and talking is correct, then any endogenously determined incapacity to entertain a particular thought would directly entail our incapacity to express it, and vice versa.

But there are theories of language and thought relative to which the equivalence of unthinkability and unsayability is not so

immediate. One of them is Fodor's (1975) own theory that thinking is conducted via the quasi-linguistic medium of Mentalese. Now Fodor has never suggested that the range of what can be thought in Mentalese isn't identical to the range of what can be said in natural language—but neither has he ever argued that the two ranges *are* identical. The hypothesis of *non*-identity between the range of the thinkable and the range of the sayable is an essential requirement for the argument from *mysticism*. If either Watson or Cole is right, then the argument from mysticism is a non-starter. It may be that the Mentalese theory is also inadequate to underwrite the argument from mysticism. But at least it gives that argument a fighting chance by providing a conceptual space for mystical insights: prima facie, a mystical insight might be a thought in Mentalese that can't be translated into a public language.

The topic here isn't the argument from mysticism, however—it's the argument from epistemic boundedness—and what's a conceptual *requirement* for the former is a conceptual *obstacle* for the latter. If thought is inner speech (whether subvocal or imagistic) in a public language, then the mystical argument is a non-starter, but epistemic boundedness trivially entails ineffability. If thinking is something other than inner speech, then the mystical argument gets off the ground, but the inference from epistemic boundedness to ineffability is no longer entirely trivial. When the time comes to assess the merits of the mystical argument, I'll have to worry about whether thinking really is something other than inner speech. In relation to the argument from epistemic boundedness, however, this possibility is a worst-case scenario. So let's assume that thinking is not just inner speech.

Then the realm of hypotheses can be divided into four classes:

1 The class of hypotheses that can be both entertained and stated. This is the realm of ordinary discursive knowledge, a sample inhabitant of which is the hypothesis that some apples are red.
2 The class of hypotheses that can be entertained but not stated. This is the realm of ineffable mystical insights. It remains to be seen whether there are or can be such hypotheses. The mystical argument is an argument to the effect that this category isn't empty. But the population statistics relating to the second class have no bearing on the argument from epistemic boundedness.

3 The class of hypotheses that can neither be entertained nor stated. This is also a realm of ineffabilities, but these ineffabilities are not candidates for mystical knowledge. They posit states of affairs that are totally inaccessible to the mind. Let's call them *deep* ineffabilities. Deep ineffability is what you get if the argument from epistemic boundedness succeeds.
4 The class of hypotheses that can be stated but not entertained. This is the possibility that makes problems for the argument from epistemic boundedness. In fact, the thesis that epistemic boundedness entails ineffability just *is* the claim that the fourth class is empty. The claim that the fourth class is empty together with the claim that we are epistemically bounded entails that the third class, the class of deep ineffabilities, is non-empty.

What about this fourth category? Is it coherent to suppose that a being can state a hypothesis which lies beyond its own epistemic horizon? What about a three-year-old child parroting the sentence "Ontogeny recapitulates phylogeny" without comprehension? What she says may very well be true; yet it can be supposed that this truth is beyond her personal epistemic bounds. By the same token, all of humanity might be as little children in comparison to a race of mentally superior Plutonians. We might be able to learn all of the syntax and phonology of the Plutonian language; yet there might be some Plutonian sentences that express ideas which we're endogenously incapable of entertaining. Nevertheless, couldn't we still *speak* those truths? Well, we might be able to produce the requisite sequence of phonemes. But we couldn't *state* the requisite hypothesis; nor could the uncomprehending child state that ontogeny recapitulates phylogeny. The argument is the same in both cases. Let *L* be the fragment of English which expresses all and only those ideas that the child is able to entertain. English is an *extension* of *L*, in the sense that the semantic rules of English include all the semantic rules of *L* and more besides. One of the additional rules of English, for instance, is the rule that the sentence "Ontogeny recapitulates phylogeny" is true if and only if ontogeny recapitulates phylogeny. But English isn't the only extension of *L*. Another one is Schmenglish, the semantic rules of which are identical to the rules of English except that the sentence "Ontogeny recapitulates phylogeny" is true in Schmenglish if and only if existence precedes essence. But now, when the child utters the sentence "Ontogeny

recapitulates phylogeny," which language is she speaking? If she's speaking English, then she's saying that ontogeny recapitulates phylogeny; but if she's speaking Schmenglish, she's saying that existence precedes essence. What the child says seems to be indeterminate between the two hypotheses. Indeed, it seems to be indeterminate over indefinitely many hypotheses that can be associated with the sentence "Ontogeny recapitulates phylogeny" by different extensions of *L*. The appearance of indeterminacy is well-founded. It's due to the fact that what we state when we utter the sentence S depends in part on what we *intend* to convey by our utterance, externalism about meaning notwithstanding. Semantic externalism is committed to the view that the speaker's intention isn't *sufficient* to determine what's being said. But it's surely *necessary*. Suppose, for instance, that an involuntary sneeze happens to sound exactly like the English sentence "I chew." I take it as obvious that not even the most enthusiastic externalist would want to say that the sneezer has stated that he chews. You can't state that *P* without intending to state that *P*. But intending to state that ontogeny recapitulates phylogeny, or that existence precedes essence, certainly requires the capacity to entertain these hypotheses, which is just what the child can't do. Therefore the child isn't stating anything at all. The same considerations lead to the conclusion that no one is capable of stating any hypothesis that he can't entertain. Therefore epistemic boundedness entails ineffability.

Fodor's argument for epistemic boundedness

If the foregoing analysis is correct, the existence of deep ineffabilities is a consequence of epistemic boundedness. But what's the status of the antecedent? The boundedness thesis has certainly not commanded universal assent. Richard Bambrough, for instance, writes: "The philosophers and poets who speak of the limits of thought and the limits of understanding represent as an impossibility what is only a difficulty, as a barrier what is only a boundary" (1978, 213).

The only reason he gives for holding this opinion is that there are many and varied ways of understanding and expressing truths. This is, of course, inadequate to secure his conclusion. But what is the case for epistemic boundedness based on? I will discuss three arguments for the boundedness thesis: Fodor's, Colin McGinn's, and one of my own devising.

Here's almost everything that Fodor has to say on the subject:

> Suppose . . . that we . . . assume that the issues about epistemic boundedness are empirical (though, of course, very abstractly related to data). It then seems to me hard to see how the unboundedness view can be made empirically plausible. The point is that any psychology must attribute some endogenous structure to the mind (really unstructured objects—bricks, say —don't have beliefs and desires and they don't learn things). And it's hard to see how, in the course of making such attributions of endogenous structure, the theory could fail to imply some constraints on the class of beliefs that the mind can entertain. . . . But I don't suppose that such reflections are conclusive. Perhaps, after all, someone will some day make sense of an unboundedness thesis. Suffice it for present purposes to claim that nobody has been able to do so to date. All cognitive psychologies thus far proposed . . . imply boundedness. . . . To repeat: when unboundedness has been defended, in the historical tradition, it has typically been on semantic rather than psychological grounds; and the semantic assumptions from which unboundedness was inferred were, in my view, uniformly not good.
>
> (1983, 125)

In this passage, Fodor offers two rationales for the boundedness thesis. First there's the idea that the attribution of any "endogenous structure" to the mind is going to imply that there are some hypotheses that the mind can't entertain. This connection between structure and boundedness isn't argued for—Fodor merely tells us that it's "hard to see" how things could be otherwise. The intuition is that you can't make a machine that does everything. Whatever the operating characteristics of the cognitive machinery may be, they're going to enable the mind to entertain certain types of hypotheses and not others. My intuition agrees with Fodor's on this point. But like Fodor, I find myself with nothing to say to someone who doesn't share our view of the matter.

Fodor's second rationale for the boundedness thesis is that all cognitive theories that anyone has ever proposed have implied boundedness. Actually, this second rationale is entailed by the first. Any possible theory of mind, extant or not, is going to attribute some sort of endogenous structure to the mind—or else it wouldn't

be a theory of mind. Then if the attribution of endogenous structure ensures the mind's boundedness, it follows that every theory of mind (a fortiori every theory of mind that has ever been proposed) entails the mind's boundedness. If Fodor had presented a persuasive case for the first rationale, the second wouldn't add anything to the argument—it would just be an obvious corollary. But the first rationale isn't persuasive—Fodor has nothing to say to a skeptic who doesn't share his intuition that any structured mechanism is going to be bounded. So now he moves to the weaker claim that every theory of mind that has ever been proposed has represented the mind as bounded. There are two questions to be asked about this weaker claim: (1) is it true?—and (2) so what if it is?

First, is it true that all proposed theories of the mind imply boundedness? Fodor doesn't try very hard to document his claim. In fact he doesn't try at all. But the documentation that he would give, if pressed, can be reconstructed from his other writings. In "The present status of the innateness controversy" (1981), which is almost contemporaneous with his discussion of epistemic boundedness, Fodor announces his intention to examine "a certain broad class of theories" which he calls *classical* theories of concept attainment. There are two types of classical theories: empiricist theories and nativist theories. Moreover, "just about every theory of concept attainment that any psychologist or philosopher has succeeded in taking seriously, barring only behavioristic theories, counts as classical" (258). So for Fodor, the claim that every cognitive theory implies boundedness is more or less coextensive with the claim that empiricism and nativism each imply boundedness. Fodor's disquisition on the Humean mind establishes that *empiricism* implies boundedness. A mind that works on empiricist principles is incapable of grasping non-sensory concepts. The only way to avoid the conclusion that such a mind is epistemically bounded is to make the dubious assumption that there are no non-sensory concepts. What about nativism? Nativists have no problem with non-sensory concepts. But they postulate a conceptual limitation of another sort. They believe that the only concepts which are accessible to us are those that are contained in a fixed repertoire of innate ideas. The only way for a *nativist* to avoid the boundedness thesis is to maintain that this innate endowment includes all logically possible concepts. But there's nothing in the nativist theory that would motivate such a view. In fact, when classical nativism is conjoined with modern biology, there are good reasons to believe that our stock of innate

concepts *is* smaller than the set of all possible concepts. Nativists in the era of Descartes and Leibniz might have been content to attribute the specifics of our innate endowment to the will of God. But modern nativists are going to prefer to cite evolution by natural selection. Now *God* might have chosen to supply us with all possible concepts, but it's extremely implausible to suppose that evolution would have produced this result. If our innate repertoire of concepts is shaped by our evolutionary history, it's compelling to suppose that a different evolutionary history would have led to a different conceptual repertoire. Modern nativists should expect that the innate conceptual system possessed by extraterrestrials would be at least partially incommensurable with ours. An interesting corollary is that nativists should be highly skeptical of the possibility for extraterrestrial communication. More relevant to the topic at hand is the conclusion that nativism, like empiricism, impels us to adopt the view that we're epistemically bounded. So Fodor is right: all the classical theories imply boundedness.

What about theories of mind that have been formulated since Fodor's discussion of epistemic boundedness? In particular, what about connectionism? Well, there are two ways of construing connectionist theories—implementationally and eliminativistically. The *implementational* interpretation is that connectionist models specify the manner in which fully symbolic activity—attaining concepts or entertaining beliefs—is implemented. On this interpretation, connectionist talk is at a lower level of abstraction than talk about concepts and beliefs. Now there are two ways in which this enterprise may be prosecuted: top-down and bottom-up. In the top-down approach, the theoretical goal is to *reduce* symbolic-level talk to talk about activation patterns among connectionist units. In this enterprise, you start with an intentional theory of some aspect of symbolic activity, and you try to construct a story about how the symbolic processes postulated by that theory are realized in neural (or near-neural) nets. This undertaking *presupposes* that we already have a symbolic story on hand for which we want to provide a reductive connectionist explanation. Obviously, this kind of connectionism isn't going to add new candidates to our stock of symbolic stories. But it's the symbolic-level theories that are relevant to Fodor's claim that all psychological theories have implied boundedness. For the purpose of getting clear on the boundedness issue, it doesn't matter *how beliefs are implemented*. What matters is *which beliefs may occur and which may not occur*, regardless of

how they're implemented. A specification of which beliefs may occur is a symbolic-level story, of which we still only have two varieties: the empiricist and the nativist. In sum, top-down implementational connectionism doesn't add any new theoretical options of the type that might affect the boundedness issue.

In bottom-up implementational connectionism, you let your connectionist theory dictate the story you tell about symbolic activity. In this case, the conceptual possibility exists that one might arrive at a new symbolic theory by doing connectionist research. It's possible, for example, that we might have independent grounds for supposing that human nets always have some property *P*, and then to show that a certain kind of symbolic activity can't be implemented in nets that have the property *P*. The result would be a new symbolic hypothesis arrived at from the bottom up. Moreover, a new symbolic hypothesis arrived at by this route might have a bearing on the boundedness issue. All this is, as I say, a conceptual possibility. But it hasn't happened yet. So Fodor's maxim still stands: all the proposed intentional theories of mind are classical theories, and all classical theories imply boundedness.

The *eliminativist* interpretation of connectionism is quickly dealt with. On this approach, connectionist theories don't *explain* how symbolic theories are implemented—they're intended to *supplant* symbolic theories altogether. The connectionist story about activitation patterns is intended to be the last word before the explanation of behavior. In other words, this style of connectionism is eliminativist with respect to the concepts of intentional psychology. But if you eliminate belief, you also eliminate the issue of whether the range of our potential beliefs is bounded. To ask whether the mind is epistemically bounded is to *presuppose* realism about the propositional attitudes. It's true that eliminativist connectionism is a theory that doesn't entail boundedness. But it doesn't entail *un*boundedness either. The issue simply doesn't arise. So Fodor's maxim *still* stands: every proposed theory of concept attainment has been classical. It's just that eliminativist connectionism isn't a theory of concept attainment. Similar remarks apply to Paul Churchland's militantly physiological eliminativism.

Now for the second question: so what? Exactly how does Fodor's dictum warrant acceptance of the boundedness thesis? Well, it might be construed as the premise of an enumerative induction: all examined theories of concept attainment entail boundedness; therefore we're warranted in supposing that the unexamined theo-

ries of concept attainment also entail boundedness; and if all theories of concept attainment entail boundedness, then boundedness is sure to be a fact—assuming, of course, that there are concepts and that their origin isn't miraculous. This is a very feeble induction. Aside from the undischarged presumption that eliminativism is false, there is the problem that the examined intentional theories which form the inductive base are not a random sample of the total population of intentional theories. They're the theories that we human beings thought of first, and as such are liable to the skewing influence of any cognitive bias that might afflict us.

Here's a better way to construe Fodor's implicit argument: if all hitherto conceived intentional theories imply boundedness, then we're *dialectically* permitted to presume boundedness in conversation and debate with any hitherto existing intentionalistic *theorist*. It would be nice for Fodor's position if boundedness could be presumed *tout court*, without having to specify the belief-states of the audience. But that presumption would be unwarranted if our interlocutor were an eliminativist, or if she were a cognitive scientist who rejects all extant theories of mind, believing that the right theory has not yet been formulated. So the most that can be said is that if any hitherto conceived theory of mind is true, then epistemic boundedness is a fact.

Even this last formulation is too strong, however. The most that can really be said is that if any hitherto conceived *naturalistic* theory of mind is true, then epistemic boundedness is a fact. That naturalism is a presupposition of Fodor's discussion is hinted at in the first sentence of the previously quoted passage ("Suppose . . . that we . . . assume that the issues about epistemic boundedness are empirical"). It's entirely explicit in the following footnote:

> The traditional way of [arguing against boundedness] is to infer the *universality* of thought from its *immateriality*—on the principle, apparently, that ectoplasm can do anything. Here is Geach's exposition of Aquinas' treatment (Fodor quoting Geach on Aquinas): "Aquinas . . . holds that a thought consists in the *nonmaterial* occurrence of a form of nature. . . . There can on this view be no special nature of the thought process to be discovered empirically; such a special nature might be expected to impose restrictions on what can be thought of, as a colored glass does on what can be seen through it—and Aquinas regards this sort of restriction as evidently impossible. Whatever nature

> of thing an A may be, if there can be an A there can be a thought of an A. . . . " (End of Geach quote—Fodor speaking again) The point here is not, of course, just that if A makes sense, so too does *the thought of A*. It's rather that, on the assumption that thought is immaterial, there are no empirical (no nonlogical) constraints on what we can think about. The question raised in the text is whether the universality of thought is plausible on any *other* ontological assumption.
>
> (1983, 138–139)

So there *have* been theories of mind that don't entail boundedness after all. What can't be found, according to Fodor, is a *naturalistic* theory of mind that doesn't entail boundedness. Thus we're free to presuppose boundedness in conversation with a lot of our colleagues—but not, inter alia, with Jerrold Katz or John McDowell.

I turn now to an issue that I've been skirting since the beginning of Chapter 2: ineffability comes in degrees. (The grades of ineffability are discussed at length in Chapter 1, p. 23ff.) When I say that a fact *F* is ineffable, the weakest claim that I can be construed as making is that there is no sentence for it in the language that I'm speaking. Call this *weak* ineffability. Weak ineffability leaves it open that the state of affairs might be expressible in a different language. Whorf's famous claim that there are Hopi sentences which have no English translation is a good example of a weak ineffability claim: according to Whorf, the states of affairs expressed by some Hopi sentences are weakly ineffable in English. Stronger than any form of weak ineffability is *human* ineffability: a fact is humanly ineffable if there's no sentence for it in any language that human beings are nomologically capable of using, regardless of whether that language has ever actually been used. If *F* is humanly ineffable, then we humans cannot devise a language in which *F* can be stated. Stronger still is *nomological* ineffability, which is inexpressibility in any language that any nomologically possible being can use. If there are facts that can only be expressed by infinitely long sentences, these facts would presumably be nomologically ineffable. Finally, a fact is *weakly logically* ineffable if there's no sentence for it in any language that can be used by any logically possible being, whereas *F* is *logically* ineffable (*tout court*) if there's no sentence for it in any logically possible language. It isn't obvious that weak logical ineffability is different from logical ineffability *tout court*—

but neither is it obvious that they're equivalent notions. This issue will come up in the subsection on the grades of mystical ineffability, p. 79.

The ineffabilities relating to epistemic boundedness are of the human and nomological varieties. When Fodor asserts that "the mind" is bounded, it's clear that he's talking about the *human* mind. If Fodor's claim about the human mind is correct, and if it's true that we can't express what we can't think, it follows that there are states of affairs that can't be expressed in any humanly accessible language, i.e., that there are human ineffabilities. But that isn't the end of the story. If, as Fodor tells us, it's the attribution of any type of endogenous structure to "the mind" that entails its boundedness, then any (naturalistic) theory of the Plutonian or the Andromedan mind is also going to entail boundedness. In fact, if Fodor's case for human boundedness is deemed to be adequate, then so is the parallel case for the thesis that all nomologically possible minds are bounded. Does this constitute a case for nomological ineffability? No it doesn't. The thesis of nomological ineffability says that there are hypotheses that can't be expressed by any nomologically possible beings. This thesis is a logical consequence of the principle that there are hypotheses that can't be entertained by any nomologically possible minds. But this principle claims more than the generalized version of Fodor's argument can hope to establish. The argument can hope to show that for every nomologically possible mind, there are hypotheses that that mind can't entertain, hence that it can't state—i.e., that there are Plutonian and Andromedan unthinkabilities as well as human unthinkabilities. But there's no reason to suppose that the realm of the humanly unthinkable (hence humanly ineffable) has any overlap with the realm of the Plutonian or Andromedan unthinkable (hence Plutonianly or Andromedanly ineffable). More generally, the fact that all nomologically possible beings have epistemic bounds is not a reason to suppose that one and the same hypothesis lies beyond the epistemic bounds of all nomologically possible beings. Fodor's argument is at most an argument for endorsing the thesis of human ineffability—more specifically, for deep human ineffability.

McGinn's argument for cognitive closure

Colin McGinn's concept of *cognitive closure* is the same as Fodor's "epistemic boundedness":

> A type of mind *M* is cognitively closed with respect to a property *P* (or theory *T*) if and only if the concept-forming procedure at *M*'s disposal cannot extend to a grasp of *P* (or an understanding of *T*).
>
> (1989, 350)

Like Fodor, McGinn presupposes naturalism:

> Minds are biological products like bodies, and like bodies they come in different shapes and sizes, more or less capacious, more or less suited to certain cognitive tasks.
>
> (350)

Like Fodor, he illustrates the concept of boundedness/closure with the Humean mind—a type of mind that can't grasp concepts referring to unobservables. Like Fodor, he maintains that human minds are bounded/closed. However, the two authors advance different types of considerations in support of their conclusion. Fodor claims that the bare fact of their having an endogenous structure already assures us that human minds are bounded. The details of the structure are irrelevant to Fodor's argument. McGinn, on the other hand, doesn't explicitly claim that the naturalized mind is cognitively closed solely by virtue of having a structure. What he maintains is that there's a particular theory *T* which is demonstrably beyond the cognitive grasp of human beings.

The two claims have significantly different consequences. If Fodor's rationale is sound, we can be sure that any (naturalistic) type of mind is going to be bounded. This sweeping conclusion can't be reached via McGinn's route. The fact that we humans can't grasp the specific theory *T* leaves it open whether some other nomologically possible mind can grasp *T*. McGinn tells us this himself, although he comments that he "would not be surprised" if *T* were ungraspable by any nomologically possible mind (350). In fact, McGinn's argument leaves it open whether there might be minds that don't have any epistemic bounds at all. So it's clear that the soundness of McGinn's argument would at most purchase the thesis of *human*, as opposed to nomological, ineffability. (As we've seen, the same is true of Fodor's more sweeping argument.)

The theory *T* that McGinn believes to be beyond human ken is the theory that gives the correct account of the intimate relation between consciousness and the brain:

> Let us say . . . that there exists some property *P*, instantiated by the brain, in virtue of which the brain is the basis of consciousness. Equivalently, there exists some theory *T*, referring to *P*, which fully explains the dependence of conscious states on brain states. If we knew *T*, then we would have a constructive solution to the mind-body problem.
>
> (361)

McGinn has two arguments purporting to show that we will never attain such a theory *T*. The first (and lesser) argument is that we will never need a theory that makes reference to *P*:

> To explain the observed physical data we need only such theoretical properties as bear upon those data, not the property that explains consciousness. We will never get as far away from the perceptual data in our explanations of those data as we need to get in order to connect up explanatorily with consciousness.
>
> (359)

As Flanagan (1991) has noted, this argument seems to be based on a simple oversight:

> McGinn's misstep comes from forgetting that consciousness has already been introduced. We are not looking for an explanation of "physical phenomena alone," at least not physical phenomena narrowly understood. There is a prior commitment to the existence of consciousness. Thus both facts about the brain and facts about consciousness are on the table to be explained.
>
> (338–339)

The second, more important, argument is that we could in any case never *construct* such a *T*, for "no coherent method of concept introduction will ever lead us to *P*" (McGinn, 1989, 358). This obstacle is due to a principle of *homogeneity* that "operates in our introduction of theoretical concepts" (358). According to the homogeneity principle, the formation of theoretical concepts proceeds "by a sort of analogical extension of what we observe": "Thus, for example, we arrive at the concept of a molecule by taking our perceptual representations of macroscopic objects and conceiving of smaller-scale objects of the same general kind" (358).

Evidently, McGinn's view is that the human mind is only slightly more capacious than the Humean mind. The latter is restricted to observational concepts, whereas the former has access to analogical extensions of observational concepts. The problem for human minds (as well as for Humean minds) is that *P* is neither an observational concept nor an analogical extension of an observational concept:

> This method [of introducing non-observational concepts] seems to work well enough for unobservable material objects, but it will not help in arriving at *P*, since analogical extensions of the entities we observe in the brain are precisely as hopeless as the original entities were as solutions to the mind–body problem. (358)

I am not inclined to dispute the view that analogical extensions of the entities we observe in the brain are useless for the task at hand. But what about the data of consciousness that are also on the table? Although McGinn doesn't emphasize the point, it's clear that he accepts the view that introspection generates its own brand of non-perceptual concepts, and that these may also, by analogical extension, yield theoretical terms that designate entities "of the same general kind." Furthermore, applying the homogeneity constraint to theoretical concepts based on introspection, we obtain the result that these concepts can *only* be analogical extensions of direct introspective concepts. Since both physical data and the data of consciousness are on the explanatory table, we will be able to avail ourselves of both types of theoretical concepts—those formed by analogy to perceptual concepts, and those formed by analogy to introspective concepts—in our attempts to explain the consciousness –brain link. Thus the conclusion that we will never be able to attain the concept *P* which is needed to solve the problem of consciousness does not yet follow from the point that *P* can't be a perceptually-based concept. McGinn also needs to maintain that *P* can't be an introspectively-based concept either. McGinn doesn't make this last point as sharply as one might wish. But his discussion reveals that he is cognizant of it. For instance, in considering whether the problem of consciousness is absolutely insoluble to all possible minds, he concedes that he hasn't shown it to be insoluble "for minds that form their concepts of the brain and consciousness in ways that are quite independent of perception *and introspection*" (361, emphasis added). In sum, it's McGinn's opinion that (1) the

analogical extensions of neither perceptual nor introspective concepts will be able to capture the elusive *P*, and furthermore, that (2) by the homogeneity constraint, these are the only concepts which are available to us for the purpose of theory construction. The conclusion is that we will never have a theory that captures *P*. I have reservations about both premises.

To begin with, McGinn's case for (1) is short of overwhelming. His discussion conveys no clear notion of what an introspectively-based theoretical concept might be like, much less what it might or might not accomplish. For that matter, even the claim that perceptually-based concepts will not enable us to contact consciousness is founded on nothing more than an appeal to intuition. It happens to be an intuition that I share. But there are others—Flanagan (1991) and Daniel Dennett (1991), for instance—who are not of the same mind. But these are quibbles in comparison to my misgivings about (2). Consider the manner in which the homogeneity constraint is described: "[T]here are reasons for believing that no coherent method of concept introduction will ever lead us to P. This is because a certain principle of *homogeneity* operates in our introduction of theoretical concepts" (358). There are plainly two ways to understand the claim that a principle "operates" in our introduction of theoretical concepts. The principle may describe an inherent limitation of our cognitive machinery, or it may be a normative account of how the machinery *ought* to be employed. McGinn manages to discuss the homogeneity constraint for several pages without committing himself to either reading. The plausibility of the argument depends on this equivocation, for the descriptive homogeneity constraint is easily shown to be false by example, while the normative homogeneity constraint is inadequate to underwrite McGinn's conclusion.

The descriptive homogeneity constraint is disproved by the intentional concepts of folk psychology. These concepts—belief, desire, intention, and so on—can't by any stretch of the imagination be regarded as analogical extensions of perceptual concepts. McGinn could conceivably claim that intentional concepts are obtained directly by introspection. However, to sustain this move, one would have to argue against the prevalent—and compelling—view that many intentional concepts refer to non-occurrent mental states, hence are unavailable to introspection. It doesn't matter how profoundly inadvisable the explanatory strategy of intentional psychology may be. The fact that it's entertained establishes that our

cognitive machinery is not restricted to analogical extensions of observational concepts. But if the claim is only that restricting theoretical concepts to analogical extensions of observational concepts is an advisable policy, then McGinn can't conclude that human beings will never be able to solve the mind–body problem. His (paradoxical) conclusion can only be that human beings won't be able to solve the mind–body problem so long as they're constructing concepts in accordance with the approved recipe. The possibility can't be ruled out that someone untutored in the proper style of concept-formation will stumble onto *P* by mistake. Either way, the argument from homogeneity fails.

The same counterargument can be run on the concepts of quantum mechanics in lieu of folk psychology. The quantum-mechanical concept of a *superposition* of states, inter alia, can hardly be regarded as an analogical extension of observational concepts. Flanagan comes close to making this criticism:

> To see that . . . the "homogeneity constraint" is overly restrictive, consider the case of the ubiquitous electron. If we assume a commitment to standard contemporary physics, it is the inference to the best explanation that certain observable processes in a cloud chamber are the traces of unobservable electrons. We never see the electrons directly while observing the process in a cloud chamber, nor for that matter do we see them anywhere else. Electrons are theoretical constructs whose postulation best explains certain observable data and whose postulation is in turn supported by certain (predicted) observations.
>
> (1991, 339)

Electrons do indeed pose a problem for the homogeneity constraint. But the problem isn't due to their unobservability. McGinn concedes that theoretical concepts referring to unobservable entities are permitted, citing molecules as an example. What the homogeneity constraint requires is not observability, but that the theoretical terms be obtained by analogical extension from the observation terms. This requirement would be met by "classical" conceptions of the electron such as Rutherford's. It's specifically the quantum-mechanical properties of the electron that undo the homogeneity constraint.

The argument from mediocrity

The third argument for ineffability hasn't, to my knowledge appeared in print. It's suggested, however, by the following isolated remark of Fodor's:

> This conclusion [that we're epistemically bounded] may seem less unbearably depressing if one considers that it is one that we unhesitatingly accept for every *other* species. One would presumably not be impressed by a priori arguments intended to prove (e.g.) that the true science *must* be accessible to spiders.
> (1983, 125–126)

This thought merits some elaboration. There are beings like spiders that are lower than us on the scale of epistemic capacity. It might be objected that spiders aren't epistemic agents at all, not even feeble ones. But at least it's clear that there are *nomologically possible* epistemic agents that are constitutionally lower than us on the scale of epistemic capacity. Fodor's and McGinn's Humean minds are an example. Now if some possible mind—call it M—is epistemically beneath ours, then there are minds that are *above* M (one of these being ours), even though M is a bona fide epistemic agent. But to say that there is a possible mind which is epistemically more capacious than M is to imply that M is epistemically bounded. So it's at least possible for some epistemic agents to be bounded. But then how do we know that we're not such bounded agents ourselves?

In sum, the existence of a range of (possible) epistemic capacities entails that some (possible) epistemic agents must be bounded. In fact, *all* varieties of epistemic agents are sure to be bounded except possibly those agents who are at the very head of the class. Minds that enjoy the very maximum of epistemic capacity may still be bounded; but at least their boundedness isn't assured by the simple fact of there existing (possible) minds that are more capacious than they are. Thus to claim that we're unbounded is to say that we occupy a unique place in the chain of possible beings: we're at the top of the epistemic heap! Epistemic agents come in various degrees of epistemic potency, and the theoretical maximum of epistemic potency is *us*. But isn't the ascription of such a privileged status to ourselves positively medieval? Four centuries after Copernicus and a century and half after Darwin, is it any longer possible for anyone

seriously to entertain the idea that we human beings occupy a privileged position in the chain of beings? If the answer to this question is "no," we have a new argument for ineffability.

I've presented this new argument for ineffability in the same format as the previous pair of arguments: a reason is cited for accepting the thesis of epistemic boundedness, and then the thesis of ineffability is derived by modus ponens from the lemma that boundedness entails ineffability. However, the present argument can be run on the thesis of ineffability directly, without altering its degree of persuasiveness. The new argument is that there are degrees of (possible) linguistic capacity, and that it's implausible to suppose that our species happens to realize the theoretical maximum of linguistic capacity. So the new argument isn't really an argument *from* epistemic boundedness. I discuss it here because it emerges from the same circle of reflections. I won't bother to keep the two forms of the new argument distinct.

Like the previous pair of arguments, this one yields at most human (as opposed to nomological) ineffability. For however improbable it may be that we are the maximally expressive of all nomologically possible species, it's nevertheless the case that some nomologically possible species is at the top of the heap. To be sure, this maximally expressive species may still suffer from ineffabilities; but there's nothing in the new argument that supports this hypothesis.

I call the new argument the *argument from mediocrity*, after a similar argument that goes by that name in the SETI literature. Is there any basis in probability theory for this argument? Well, there is this principle: given a dichotomous category system, more things belong to the more numerous category than to the less numerous one (does anybody want to disagree with that?); it follows that an element chosen at random from the universe containing the two categories is more likely to have come from the more numerous category than from the less numerous category. The argument is that if humanity were the maximally expressive species, it would belong to a very underpopulated category; therefore it probably doesn't belong to that category. More precisely, one may suppose that all nomologically possible types of language-using minds are rank-ordered by the degree of their expressive capacity, and that they're then divided into two categories: the maximally expressive, and all the rest. Then, since the second category is the more numerous, our best bet is that we humans belong to it rather than to the exclusive

club of the maximally expressive. To be sure, the ordering of minds by capacity may be only partial—there may be minds *M1* and *M2* such that the epistemic capacities of neither one are fully contained by the other. But if that's the case, then there are going to be hypotheses that *M1* can entertain but *M2* can't, and vice versa—and then *M1* and *M2* will be epistemically bounded anyway. So we need only consider completely ordered chains of epistemic capacity. The argument from mediocrity is that we're less likely to be at the very head of such a chain than from somewhere in the middle.

The probabilistic principle appealed to in this argument is undoubtedly valid—if all the elements in the two categories are equiprobable. (That's the import of saying that an element is chosen *at random* in the initial statement of the principle.) So, in the argument from mediocrity, we can't carve up types of minds in any way we wish. Given any two types *A* and *B*, the probability that *A* is realized in the world has to be the same as the probability that *B* is realized. But then, relative to such a special taxonomy, there can be no a priori guarantee that the set of maximally expressive types of mind is going to be a singleton. There might be a many-ways tie for first place. For all we know, the maximum of expressiveness may be routinely reached by most of the types of minds delineated by the equiprobabilistic taxonomy. If that were to be the case, then the better bet would be that we are among the maximally expressive. When we make contact with the Galactic Confederation, we may find that beings from every sector of the galaxy speak maximally expressive languages. In sum, the assumption that our linguistic capacity is equal to the nomological maximum doesn't entail the dubious, medieval hypothesis that we're unique. The argument from mediocrity for ineffability fails.

The grand conclusion of this section is quickly stated. Three arguments for human ineffability have been evaluated, and two of them have been found wanting. The third is Fodor's argument that all current naturalistic theories of mind entail epistemic (hence also linguistic) boundedness. To be sure, Fodor's claim doesn't enable us to *deduce* that our minds are epistemically bounded. But it provides an adequate rationale for *assuming* epistemic boundedness when we converse with anyone who endorses a naturalistic view of the mind.

The argument from mysticism

The mystical experience

The *locus classicus* for philosophical discussion of mysticism is William James' (1902/1985) *Varieties of Religious Experience*. In his seminal chapter on mysticism, James famously delineates "four marks which, when an experience has them, may justify us in calling it mystical" (380). Two of these marks—the *transiency* and the *passivity* of the experience—are "less sharply marked" than the other two, and are merely "usually found" in mystical states (381). Indeed, the other two marks by themselves "entitle any state to be called mystical"(381). These two defining characteristics are the *ineffability* of the mystical state and its *noetic quality*.

James characterizes the ineffability of mystical states as follows:

> The subject of it immediately says that it defies expression, that no adequate report of its content can be given in words. It follows from this that its quality must be directly experienced; it cannot be imparted or transferred to others. In this peculiarity mystical states are more like states of feeling than like states of the intellect. No one can make clear to another who has never had a certain feeling, in what the quality or worth of it consists. One must have musical ears to know the value of a symphony; one must have been in love one's self to understand a lover's state of mind. Lacking the heart or ear, we cannot interpret the musician or the lover justly, and are even likely to consider him to be weak-minded or absurd. The mystic finds that most of us accord to his experiences an equally incompetent treatment.
>
> (380)

Of the noetic quality he writes:

> Although so similar to states of feeling, mystical states seem to those who experience them to be also states of knowledge. They are states of insight into depths of truth unplumbed by the discursive intellect. They are illuminations, revelations, full of significance and importance, all inarticulate though they remain; and as a rule they carry with them a curious sense of authority for after-time.
>
> (380–381)

Evidently, the "ineffability" of mystical states entails the inability of mystics to convey the quality of their experience to non-mystics, whereas their "noetic quality" entails the inability of mystics to articulate the insight that they derive from the experience. Either one of these characteristics alone could justify our describing the mystical state as ineffable. But the sense of ineffability that's at issue here is only the second one. For me, ineffability is primarily a property of facts or states of affairs: a fact is ineffable with respect to a class of languages if it can't be stated in any language of that class. It remains to be seen whether we can derivatively talk of ineffable knowledge or belief as the knowledge of or belief in an ineffable state of affairs. At any rate, the occurrence of such an epistemic state is essentially James' second mark of the mystical. But the first mark, which James refers to as the "ineffability" of mystical states, has no bearing on the issue of ineffability as I understand it. If the unconveyability of an experience to those who never had it did count as an instance of ineffability, we would not have to advert to anything as exotic as mystical experience to establish the existence of ineffabilities. The ineffability thesis would follow directly from the unconveyability of love, or of the color red, to those who have never experienced it. But ineffability (as I understand this term) requires more than that some (or most) people can't successfully be told of some state of affairs; it requires that nobody can be told of it (in a given class of languages), because nobody can formulate the hypothesis that the state of affairs obtains. The unconveyability of the feeling of love to non-lovers is entirely compatible with the possibility of accurate and complete discourse about love amongst members of a community of lovers. That is to say, the ineffability of love in James' sense doesn't entail its ineffability in the present sense. By the same token, the unconveyability to non-mystics of the mystical experience doesn't establish that there are any facts about the mystical state that can't routinely be stated by one mystic to another. But a fact which one mystic can state to another mystic is not ineffable in the present sense. What I call the ineffability thesis requires that there be facts that nobody can state to anybody—not even to oneself.

But by this criterion, aren't *all* facts about mystical experience going to be effable? For suppose that the mystical experience has a quality Q. What's to stop the community of mystics from introducing a term "Q" for Q by ostensive definition, in the same way as the broader linguistic community ostensively defines

"red"? Well, it isn't obvious how the act of ostension is going to be performed in this case—literal pointing certainly won't work. But that's not the reply that I want to press. Let's assume that mystics *can* introduce a term for every quality that the mystical experience possesses. It follows that there is no fact about a mystical experience having or not having a particular quality that can't be expressed. But it doesn't follow from *that* that the mystical experience doesn't make us privy to an ineffable insight. It's just that the ineffable insight can't be about the qualities possessed by the experience; the insight has to be about some other state of affairs. But can't a community of mystics use demonstratives to refer to the mystical insight regardless of what that insight is about? Can't they call the content of *this* insight—the one that they receive while in a mystical state—by the name of "*T*"? And then can't they describe their mystical state as a state in which they have the insight that *T* obtains? And doesn't that mean that the insight is effable after all? The answers to the last four questions are: yes, yes, yes, and no. If *T* is defined as "the fact that was revealed to me while in a mystical state," then I can describe my mystical state as one in which I had the insight that *T* obtains. The fact that I had the particular insight that I did and not another is not ineffable. But to say that I had this insight is not yet to *state the fact that was revealed to me*. While there is a sense of "about" for which it may be said that all facts about the mystical experience are effable, it nevertheless remains possible that the mystical experience reveals a fact—not about the experience itself—that can't be effed. This topic is discussed at greater length in the section "Is the Tarskian criterion of ineffability vacuous?" on p. 48.

Two or three types of mystical claims

For the purpose of evaluating the argument from mysticism, a mystical state is any state that possesses the second of James' two marks of the mystical—it's an intuition that can't be put into words. Mystics are people who claim to have such an intuition. There is of course the question of what grade of ineffability is being claimed. This topic will be discussed in the next subsection. In this section, I distinguish two or three types of mystical claims along a different dimension. First, a mystic may claim to have an intuition that a state of affairs *X* obtains, and she may *judge* that her intuition is ineffable. Perhaps she comes to this opinion about *X* because she finds herself

unable to put it into words. Let's call this the *judgmental* claim. Second, a mystic may claim to have the conjunctive intuition X' which immediately informs her (1) that the state of affairs X obtains, and that X is ineffable. Let's call this the *intuitive* claim. Both the judgmental and the intuitive claim consist of two parts —the *substantive* claim that X obtains, and the *metaclaim* that the substantive claim is ineffable. They differ, however, in the grounds for making the metaclaim. If the opinion that X is ineffable comes after several years or moments of reflection on a mystical experience, the claim is of the judgmental variety. If the realization that X is ineffable comes like a flash as an integral part of the epiphany that reveals X, the claim is of the intuitive variety. For the sake of completeness, I should mention a third route to ineffability via intuition: one may have the intuitive realization *that there are ineffable states of affairs*. Let's call the claim that one has such a realization the *existential* claim.

Which of the three claims is the one made by the canonical mystics? It certainly isn't the existential claim. The canonical mystics of both East and West are univocal in their protestation that their insight can't be put into words. But the insight of the existential claimant is easily put into words—it's the insight that there are states of affairs that can't be put into words. Unlike the judgmental and intuitive claims, the existential claim doesn't entail that the claimant is in possession of ineffable knowledge. Hence existential claimants don't even qualify as mystics. But they stand in the same relation to the ineffability thesis as judgmental and intuitive claimants: if any one of these three claimants is right, the ineffability thesis is true.

Mysticism proper—the protestation that one's insight can't be put into words—is compatible with both the judgmental and the intuitive claim. Moreover, most mystical reports are noncommittal between these two option. The following passage by Tennyson is typical. He writes that his recurrent mystical experience would come upon him through repeating his own name silently,

> till all at once, as it were out of the intensity of the consciousness of individuality, individuality itself seemed to dissolve and fade away into boundless being, and this not a confused state but the clearest, the surest of the surest, *utterly beyond words*.
>
> (quoted in James, 1902/1985, 384, emphasis added)

Is the italicized phrase a description of a part of Tennyson's extraordinary noetic experience, or is it an opinion *about* the experience? There's no way to tell.

St. John of the Cross makes for an interesting case study in this regard. He writes:

> We receive this mystical knowledge of God clothed in none of the kinds of images, in none of the sensible representations, which our mind makes use of in other circumstances. Accordingly in this knowledge, since the senses and the imagination are not employed, we get neither form nor impression, nor can we give any account or furnish any likeness, although the mysterious and sweet-tasting wisdom comes home so clearly to the inmost parts of our soul.
>
> (quoted in James, 1902/1985, 407)

In this passage, St. John espouses an empiricist semantic theory: if "the senses and the imagination are not employed," then we can give no account of the matter. Hume combines the same semantic theory with the empiricist *psychological* theory that all our ideas come from sensation. In contrast, St. John adverts to knowledge which is "clothed in none of the kinds of images, in none of the sensible representations, which our mind makes use of in other circumstances." St. John's empiricism thus imposes a constraint on the hypotheses that can be encoded linguistically which doesn't apply to the hypotheses that can be entertained. This is how he provides conceptual space for ineffable knowledge. In the present context, the important point is that he justifies his belief in the ineffability of his non-sensory intuition by means of a philosophical argument. Thus his ineffability claim is of the judgmental variety. If the empiricist theory of meaning is rejected (as it must be), St. John's case for the ineffability of "the sweet-tasting wisdom" evaporates.

This isn't the whole story, however. St. John continues:

> The soul . . . feels as if placed in a vast and profound solitude, to which no created thing has access, in an immense and boundless desert, desert the more delicious the more solitary it is. There, in this abyss of wisdom, the soul grows by what it drinks in from the well-springs of the comprehension of love, . . . and recognizes, however sublime and learned may be the terms we employ, how utterly vile, insignificant, and improper

> they are, when we seek to discourse of divine things by their means.
>
> (407–408)

Here St. John seems to be telling us that the ineffability of the mystical intuition is directly given to him as part of the intuition itself. The recognition that any and all terms are "improper . . . when we seek to discourse of divine things by their means" is said to occur "*there*," in the same "abyss of wisdom" wherein the soul "drinks from the well-springs of the comprehension of love." That is to say, the substantive mystical insight and the meta-insight that the substantive insight is ineffable are both features of one and the same noetic experience. This would, of course, mean that St. John's claim is of the intuitive variety. These two passages don't necessarily betoken a confusion on St. John's part, for the intuitive claim and the judgmental claim are not mutually exclusive: one may have the intuition that *X* and at the same time have good reasons for believing that *X*. But each claim engenders its own dialectic.

The grades of mystical ineffability

Not surprisingly, most accounts of mystical states fail to specify the degree of ineffability being claimed. Among those that do address this issue, we find the full range of claims, from the weakest to the strongest. J. A. Symonds asserts nothing more than his personal inability to describe his experience:

> Suddenly . . . at church, or in company, or when I was reading, and always, I think, when my muscles were at rest, I felt the approach of the mood. Irresistibly it took possession of my mind and will, lasted what seemed an eternity, and disappeared in a series of rapid sensations which resembled the awakening from anaesthetic influence. One reason why I disliked this kind of trance was that I could not describe it to myself. I cannot even now find words to render it intelligible.
>
> (quoted in James, 1902/1985, 385)

This modest claim doesn't even amount to weak ineffability, since Symond's failure is entirely compatible with there being an English sentence that does the job.

According to Paul Henle, "religious mystics may be taken as asserting that their insights are ineffable with regard to all known symbolisms" (1949, 419). This is a philosopher's opinion rather than a claim by a mystic. The opinion, however, is that mystics claim that their substantive insight is *weakly* ineffable with respect to all the languages that we human beings have ever devised. This metaclaim is stronger than the claim that the substantive insight is weakly ineffable in the mystic's own language of discourse. But it still doesn't claim as much as *human* ineffability. An example of a human ineffability claim is provided by D. T. Suzuki:

> according to the followers of Zen its apparently paradoxical statements are not artificialities contrived to hide themselves behind a screen of obscurity; but simply because the human tongue is not an adequate organ for expressing the deepest truths of Zen.
>
> (Suzuki, 1949, 33)

Whereas Henle maintains that the mystical insight is ineffable in all *known* symbolisms, Suzuki makes the stronger claim that it's ineffable in all symbolisms that are humanly accessible.

There are examples of ineffability claims that are stronger than Suzuki's. However, I haven't found any that make a distinction between nomological and weak logical ineffability. The following passage, again by a commentator rather than a mystic, endorses either nomological or weak logical ineffability:

> It is of the very nature of the intellect to involve the subject–object opposition. But in the mystic experience this opposition is transcended. Therefore the intellect is incapable of understanding it. Therefore it is incomprehensible, ineffable.
>
> (Stace, 1952, 40)

The assertion that the mystical insight is incomprehensible to the intellect as well as ineffable complicates the story. Apparently Stace, like many other mystics and commentators on mysticism, wishes to distinguish two epistemic states: an "intellectual" understanding which is as incapable of comprehending the truths of mysticism as our linguistic apparatus is of expressing them, and a super-intellectual mode of apprehension which provides the vehicle for the mystical intuition. I will have more to say about this idea

below. For the moment, I will ignore it and comment only on the claim that no intellect is capable of effing the mystical truth. According to Stace, the cause of mystical ineffability is to be found in certain deficiencies of the intellect. Since these deficiencies inhere in "the very nature" of the intellect, we may suppose that they afflict every possible type of intellect. If the possibility at stake is nomological, the result is nomological ineffability. If the possibility is logical, the result is weak logical ineffability. Stace's claim falls short, however, of affirming the existence of logical ineffability *tout court*. The thesis of logical ineffability makes the claim that there are facts that can't be represented in any logically possible language. Citing limitations of the intellects of language users—even of all logically possible language users—doesn't immediately purchase this stronger thesis. An argument for logical ineffability has to be based on limitations that afflict all logically possible *languages*, regardless of whether there are any logically possible beings who may be able to use them.

Are there any claims in print for the existence of logical ineffabilities? Pletcher maintains that the ineffability claims of the religious mystics are routinely that strong: "Is the mystical experience coherently describable in *some* conceptual system? . . . mystics and mystic interpreters consistently answer this question in the negative" (Pletcher, 1973, 207). This passage is followed by selections from mystical texts which purport to illustrate Pletcher's historical thesis. These quotations don't sound like unambiguous endorsements of logical ineffability to me. For example, here's one from Lao-Tze: "The great everlasting infinite First Cause can neither be defined nor named" (207). This passage doesn't tell us whether the First Cause can't be defined or named in the language of discourse, or in any humanly accessible language, etc.

Finally, it should be noted that the enormous range of claims concerning the ineffability of the substantive insight doesn't necessarily betoken an inconsistency in the mystical literature. For one thing, there's no a priori reason to suppose that the substantive insights reported by various mystics are all the same—some mystics may have and correctly report an insight that's weakly ineffable, while others correctly report a human ineffability, etc. For another thing, each grade of ineffability entails all the lower grades. Thus the hypothesis that all mystical insights are logically ineffable would be consistent with the full range of metaclaims. One doesn't always have to assert the very strongest thesis to which one's entitled to.

Accepting the metaclaim: necessary conditions

The previous subsections were devoted to discussing what the mystics say. Now it's time to assess the merits of their claims. The first point to be made is that the substantive claim and the metaclaim are logically independent. One could coherently accept the mystic's claim to have had a veridical intuition into the workings of the world, but reject her assertion that the intuition is ineffable. Conversely, one might concede that the mystic has an ineffable intuition, but maintain that there are inadequate grounds for accepting that her intuition is veridical. Note that the concession that the mystic's opinion is ineffable is sufficient to conclude that there are ineffable facts. For suppose that the mystic's substantive belief is that *P*, and that *P* is an ineffable hypothesis. If *P* is ineffable, then so is its negation not-*P*. But either *P* is true or not-*P* is true. Whichever one is true, the fact that it is true will be an ineffable fact. So the acceptability of the ineffability thesis doesn't depend on the veridicality of the mystic's substantive opinion. It's enough that the opinion can't be put into words. The epistemic status of the mystic's substantive opinion is the subject of Chapter 3. At present, I deal exclusively with the metaclaim that the substantive opinion is ineffable.

There are certain basic conditions that need to be met if the metaclaim is to be a viable hypothesis. Most basically, *believing* that *P* and *saying* that *P* need to be two distinct processes, and it needs to be a coherent possibility that the range of the believable exceeds the range of the sayable. If believing that *P* requires that we tell ourselves that *P*, perhaps subvocally, then the argument from mysticism is undone. In the discussion of the argument from epistemic boundedness (pp. 53ff.), the assumption that thinking and talking are distinct processes represented a worst-case scenario. But relative to the argument from mysticism, it's a necessary condition. So in this section, we can't *assume* that thinking and talking are distinct processes—we have to *show* it.

Fodor's (1975) Mentalese hypothesis gives us what we need: Mentalese is a non-linguistic medium in which hypotheses can be formulated and beliefs can be entertained. Fodor gives two arguments for the existence of Mentalese. The first is that its existence is required if we are willing to attribute thoughts to non-linguistic organisms, and it's overwhelmingly plausible that some non-linguistic organisms like gorillas and human babies can think.

The second is that its existence is required to account for first-language learning. For a pre-linguistic child to learn the extension of the English term "dog," he has to hypothesize that "dog" refers to dogs; but if English is the first language that he's learning, he has no alternative but to frame this hypothesis in a non-linguistic medium. These arguments are not apodictic—one might contemplate biting the bullet and denying that babies can think or that learning a first language involves a process of hypothesis testing. But they certainly render the Mentalese hypothesis plausible.

To be sure, Fodor never tells us that there can be Mentalese thoughts that have no linguistic equivalents. But neither does he say anything that implies that all Mentalese thoughts *must have* a linguistic equivalent. In fact, granting that we can think in Mentalese, it necessarily follows that there are weakly ineffable human thoughts—i.e., thoughts that can't be expressed in a given language. Consider, for example, a fragment of English which (1) contains fewer expressive resources than English, but which (2) is large enough and closed under sufficiently many syntactic operations to qualify as a language in its own right. Perhaps this language—call it Prenglish—is just like English except that it lacks one indefinable English term. By hypothesis, there are some English sentences that have no equivalent in Prenglish. But every English sentence has a Mentalese equivalent. Therefore there are some Mentalese thoughts that can't be expressed in Prenglish. These thoughts are weakly ineffable relative to Prenglish. Can we be sure that the language of Prenglish exists—that there is a proper fragment of English which is a language in its own right? To deny the existence of Prenglish is to commit oneself to the utterly implausible view that children who have not yet mastered all the expressive resources of English don't speak a language at all. (This point is discussed at greater length in Chapter 1, in the section on untranslatable languages, p. 34.)

I have no equally compelling demonstration of the existence of humanly ineffable Mentalese thoughts—human thoughts that can't be expressed in any humanly accessible language. But it's easy to show that humanly ineffable thoughts are a coherent possibility. One can imagine a species whose linguistic capabilities fall systematically short of its epistemic capabilities. In fact, if we could build a machine that thinks like a human and speaks English, we could also build a machine that thinks like a human but can only speak Prenglish. All we'd have to do is to disable some of the Mentalese-

to-English circuits of the first machine. Equivalently, there might be a cognitively subhuman species whose prowess in Mentalese is equal to our own, but whose linguistic capabilities are nomologically fixed at the five-year-old human level. If this is a coherent possibility, then it's also a coherent possibility that we humans stand in the same relation to Plutonians as the subhumans stand to us. There would then be some Plutonian thoughts that have no equivalent in any humanly accessible language, but that humans can nevertheless entertain. In other words, there would be humanly ineffable human thoughts.

The same sorts of considerations won't purchase the coherence of nomologically or logically ineffable human thoughts. To secure this stronger thesis, it would have to be shown that there can be some human thoughts that can't be translated into a public language by any (nomologically or logically) realizable community of cognitive systems. I don't know of any persuasive arguments to that effect. But neither do I know of any persuasive arguments for the *in*coherence of nomologically or logically ineffable human thought. To be sure, there are arguments purporting to show the incoherence of nomologically and logically ineffable *states of affairs*, and the conclusion of these arguments trivially entails that there can't be any nomologically or logically ineffable thoughts either. But, as noted in the introduction, it's a presupposition of the present discussion that the notion of a (nomo)logically ineffable state of affairs is coherent. The question here is whether, given this presupposition, there are any additional difficulties that render the concept of a (nomo)-logically ineffable *thought* unusable. My answer is that I know of no reason to suppose that there are any such difficulties. This doesn't show that difficulties may not yet turn up—someone might come up with a proof of incoherence tomorrow. But the same can be said of any hitherto unrefuted thesis. When we have no reason to suspect that an idea is incoherent, it's appropriate to default to the view that it's not incoherent.

A final point: Stace, as well as many other mystics and mystic commentators, would strenuously object to my equating the substantive mystical insight with an untranslatable Mentalese idea. Recall from the subsection on the grades of mystical ineffability (p. 79) that Stace regards the mystical truth as being not only ineffable, but also incomprehensible to the "intellect." Other passages make it clear that for Stace, the uncomprehending intellect is an epistemic organ that works by forming concepts and composing

propositions out of them: "Thus to the intellect He is blank, void, nothing. You cannot attach any predicate to Him . . . because every predicate stands for a concept, so that to affirm a predicate of Him is to pretend that He is apprehensible by the conceptual intellect" (Stace, 1952, 42). Presumably, the ineffable insight of the mystic is the product of a non-intellectual epistemic process that doesn't rely on conceptualization. This takes Mentalese out of the picture, for Mentalese most certainly relies on concepts: for the pre-linguistic child to hypothesize that "dog" refers to dogs, he has to utilize his Mentalese concept of a dog. So if Stace is right, the mystical insight is as unrepresentable in Mentalese as it is in English. Does this notion of nonconceptual (not just nonverbal) knowledge make sense? I will discuss this issue in Chapter 3. For now let's simply suppose that it does make sense. Suppose, in fact, that Stace is right in maintaining that the mystical insight is inexpressible in Mentalese. The fact remains that if there is a Mentalese idea, conceptual though it be, that can't be put into words, then the ineffability thesis is true. The acceptability of the ineffability thesis doesn't *require* the existence of nonconceptual insights. If Stace is right, then there are two varieties of insights that can't be put into words: those that can be represented in Mentalese, and those that can't. Either one alone would sustain the argument from mysticism.

Accepting the metaclaim: sufficient conditions

In the previous subsection, I argued that there are no good reasons for rejecting even the strongest grades of mystical metaclaims. Are there positive reasons for adopting any of them? Let's look at the intuitive claim first—the report that the substantive mystical insight as well as the realization that the insight is ineffable are both intuitively given in the mystical experience. As James notes, the epistemic situation is different for the mystic and the non-mystic. According to James, the mystic's epistemic relation to her own mystical experience is the same as the ordinary person's relation to ordinary perceptual experience. We suppose that ordinary people are entitled to believe in what they see; for reasons of parity, we should also grant the mystic the same privileges in regard to the revelations of her mystical experience. (James speaks only of the mystic's substantive insight in this context; but the same argument applies to the meta-insight of the intuitive claim.) In both cases, the belief in what one experiences is "invulnerable" (424). Nowadays,

most philosophers would say that the evidence of one's own senses may be overridden by other sources of evidence, as when we judge that a stick half-immersed in water is not bent, appearances to the contrary notwithstanding. There is, however, a broad consensus that accepting the testimony of one's own senses, or of one's memory, is the default option: we are right to accept such non-discursive presentations of facts as veridical so long as there are no reasons to reject them. The mystical experience is just such a non-discursive presentation of facts. Are there reasons to doubt the putative facts? The status of the *substantive* claim is discussed in Chapter 3. As for the *meta*claim which is at issue here, I refer the reader back to the first sentence of this paragraph: there are no good reasons for rejecting even the strongest grades of mystical metaclaims. If crediting her own intuition is indeed the mystic's default, the absence of reasons to reject her intuition is sufficient reason for her to accept it.

What should the non-mystic make of the mystic's metaclaim? A number of writers have argued that we also have a presumptive epistemic right to accept the testimony of *others* so long as we are not in possession of reasons for rejecting them. (Several variations on this theme are represented in Schmitt, 1994.) This view is defended in either of two ways. The first way is to argue that we have independent grounds for expecting that a testimonial is likely to be true, even if we know nothing else about it except that it is a testimonial. This might be based on an inductive argument (most past testimonials have been true), or on the view that human beings have an innate tendency to tell the truth, or on the view that we ourselves usually tell the truth and others are like us in most respects (this is also an inductive argument). The second way is to argue that testimony, like perception and memory, is a "basic" source of information—one whose reliability has to be assumed a priori, because it admits of no non-question-begging justification, and yet reliance on it is essential to the conduct of our epistemic affairs. These views are all variations on the theme that accepting the testimony of others is another default option. The claim is controversial, and it wouldn't be appropriate for me to try to adjudicate it here. My point is only that even if non-mystics could find no other evidence for the mystic's metaclaim than the mystic's say-so, it's at least arguable that her say-so would be sufficient to warrant the claim, so long as there are no contraindications. And there are presently no contraindications.

The foregoing analysis applies, mutatis mutandis, to the *existential* claim—the claim that one has the intuition that there are ineffabilities: the possessor of the intuition is entitled, ceteris paribus, to credit her own insight, and the rest of us are arguably entitled, ceteris paribus, to do the same. Now what about the *judgmental* claim to have had a substantive insight and to be in possession of good reasons for supposing that the insight is ineffable? If the grounds for adopting the metaclaim are discursive rather than intuitive, the mystic and the non-mystic ought to be able to discuss them and come to the same resolution; as regards the judgmental metaclaim, mystic and non-mystic are in the same epistemic boat.

What sorts of reasons might be adduced for the judgmental metaclaim? First and foremost is the fact that the mystic fails in her attempt to put her insight into words. Strictly speaking, the non-mystic has to be willing to credit the testimony of the mystic even for this very weak and unexciting claim. But just because it's so weak and unexciting, it's not very much for the mystic to ask to be taken at her word. Moreover, the evidence provided by the mystics' expressive failures might be parlayed into evidence for higher grades of ineffability. For example, the fact that legions of mystics assert their inability to state the content of their insight may be explained by the theoretical hypothesis that their insight is weakly ineffable in all known languages. To be sure, other explanations are also possible. But the weak ineffability of the insight receives some measure of confirmation by virtue of being one of the hypotheses that are able to explain the data.

There is, however, a palpable weakness in the proffered explanation of the data. It's only if the mystics' inexpressible insights are all, or at least predominantly, the *same* insight that an appeal to weak ineffability is explanatory. If every mystic fails to clothe a different insight in words, there is no need to appeal to weak ineffability, for the data are adequately explained by the obviously true hypothesis that most people are bound to fail some time or other in their attempts to state the ideas that they entertain. However, this obviously true hypothesis would not provide an explanation for the fact (if it is a fact) that there is a frequently entertained idea that nobody can state. But if the idea at issue can't be stated, how can either mystics or non-mystics know whether all or most of the mystics are talking about the same one? The problem could easily be solved if the idea were ineffable in one language *L1* but statable in another language *L2*—for then the evidence that one

and the same idea is ineffable in *L1* could be that every bilingual claimant chooses the same sentence of *L2* as an expression of it. However, this sort of evidence would not be available if the thesis being evaluated were the somewhat stronger grade of weak ineffability according to which there are ideas that can't be expressed in any known human language. If two people each have an idea that can't be expressed in any language, how can they or anyone else come to know that their ideas are the same? I will try to provide an answer to this question in Chapter 3. I will claim that there is nomologically possible evidence that would support the hypothesis that two universally inexpressible ideas are the same, but that we are not presently in possession of such evidence. This is, of course, merely a promissory note. But the current conclusion is the same whether or not the promissory note is accepted: the evidence concerning mystics' inability to put their insights into words does not warrant the conclusion that their insights are weakly ineffable in natural languages.

There may be other strategies for the argument from mysticism that don't depend on the data on subjective inexpressibility. I don't know of any that work, however. St. John's empiricist argument is a case in point. The premise of St. John's argument is that the mystical experience is non-sensory—it's the need for this premise that qualifies St. John's argument as a species of the argument from mysticism. Combined with the empiricist semantic principle that all linguistic terms must be definable in terms of sensations, St. John's premise directly yields the conclusion that the mystical experience is ineffable. The grade of ineffability appropriate to the conclusion depends on the strength of the empiricist assumption. If the stricture that all terms must be sensory is a limitation peculiar to the human mind, then the conclusion is the thesis of human ineffability. If the stricture applies to all nomologically possible minds, the argument concludes for nomological ineffability. And if the empiricist principle applies to all logically possible terms, as it does for Hume, then the conclusion is the thesis of logical ineffability. Unfortunately for the argument from mysticism, even the weakest of these versions of empiricist semantics is untenable (see Grover Maxwell, 1962).

I want to discuss one more argument for ineffability based on mystical experience. This one requires its own subsection.

The argument from contradiction

A. W. Moore (1990) and Rudy Rucker (1982) have independently maintained that we possess an understanding of certain mathematical truths that cannot be formulated in any logically possible language. The argument begins with the observation that we can't speak of certain sets—e.g., the set *R* of all sets that are not members of themselves—without falling into contradiction. The contemporary solution to this and related antinomies is to deny that the problematic set exists. But, say Moore and Rucker, we surely possess some understanding of some of the properties of that collectivity. For example, doesn't it capture or reflect or intimate or otherwise indicate some sort of understanding to say that the set of all natural numbers, not being itself a natural number, is a member of *R*? According to Moore and Rucker, the resolution is to acknowledge that we have an understanding of these matters, but to say that our understanding is ineffable. We may try to convey—and actually succeed in conveying—that understanding by talking about "the set of all sets that are not members of themselves." But the paradoxes show us that such talk cannot be understood literally—for strictly speaking, there is no set of all sets that are not members of themselves.

What shall we say of this argument? As both Moore and Rucker note, the form of the set-theoretical argument for ineffability is very similar to the form of the traditional, religious argument from mystical experience. Indeed, both mathematical authors tend to treat the set-theoretical argument as a species of mystical argument. In both the religious and the mathematical cases, an appeal is made to a *state of insight*, and reasons are given for supposing that the insight cannot be represented in language. Sometimes the reason is even the same: the attempt to state the insight results in contradictions:

> It stirs and it stirs not.
> It is far, and likewise near.
> It is inside all this, and it is outside all this.
>
> (Isa Upanishad, quoted in Stace, 1972, 255)

There are important differences between the set-theoretical and the mystical versions of the contradiction argument, however. In the mystical argument, we're asked to accept the testimony of the

mystics on two points: that they have an insight which most of us do not possess, and that the attempt to represent the content of the insight results in a contradiction. In the set-theoretical argument, the insights are readily and universally available—who could fail to see that the set of all sets that are not members of themselves has various mundane properties, such as containing the set of all natural numbers as a member? Moreover, the claim that a contradiction ensues from trying to represent the insight is based on a simple deductive argument that anyone can follow. Evidently, the set-theoretical argument stands on much firmer ground than its religious cousin.

On the negative side of the ledger, Moore and Rucker don't really discredit what I take to be the orthodox view, namely that our supposed insights about paradoxical sets like *R* are *illusory*—that since there is no set of all sets that are not members of themselves, there's no insight about "it" to be had. "It" neither contains the set of all natural numbers nor fails to contain the set of all natural numbers—because there is no "it."

I will have more to say about the orthodox view below. But first I want to pursue an objection to Moore and Rucker's thesis that begins by granting them the premise that we do have an understanding of the paradoxical sets. To accept this premise is to repudiate the orthodox view. But does the premise warrant the conclusion that our understanding is *logically* ineffable? The same question can be asked about the religious version of the contradiction argument: granted that the religious mystic really has an insight and that the attempt to put it into words produces a contradiction, does it follow that the insight is logically ineffable? Henle (1949) has proposed and Pletcher (1973) has elaborated an account of the mystical scenario which seems to provide us with an alternative conclusion. They try to refute what I called the orthodox view (that the mystics' contradictions show that the mystics have no coherent insight) by constructing an artificial language in which the attempt to express certain facts issues in a contradiction, even though these facts are readily expressible in our language. If the demonstration works, it weakens the case for the orthodox view: apparently the generation of a contradiction doesn't necessarily mean that there is no coherent thought being expressed—the contradiction could instead be due to limitations in the expressive resources of language. In addition, though this is not a point that

Henle or Pletcher makes, the demonstration would also refute the Moore–Rucker thesis that the generation of a contradiction by the attempt to state a coherent thought entails that the thought is logically ineffable. If Henle's and Pletcher's demonstration works, it shows that the generation of a contradiction in a single language doesn't mean that the thought can't be expressed coherently in any language—the contradiction could be due to inherent limitations in the expressive resources of the language in which it's generated.

Here is a streamlined version of Henle's demonstration which eliminates some inessential complications. Let *L* be a written language in which the non-logical symbols are superimposed over one another instead of being concatenated. Thus the sentence "Alfred is short and Alonzo is tall" might be written by superimposing a symbol for Alfred over a symbol for the property of being short, concatenating the result with a symbol for conjunction, and concatenating that result with the superimposition of a symbol for Alonzo and a symbol for the property of being tall. The superimpositions may make *L* very difficult to read, but it can be supposed that the symbols are chosen in such a way that there is always a unique way of resolving which symbols comprise a complex expression. Suppose now that users of *L* conceive of the relation of greater-than between numbers and introduce a symbol for it in their language. Look what happens when they try to express even so elementary a fact as that 3 is greater than 2, but 2 is not greater than 3. They come up with a four-part concatenation. The first part is composed of the superimposed symbols for 3, the greater-than relation, and 2; the second part is a symbol for conjunction; the third part is a symbol for negation; and the fourth part is a superimposition of symbols for 2, the greater-than relation, and 3. But the superimposition that comprises the first part turns out to be syntactically identical to the superimposition that comprises the fourth part! Letting *X* stand for that particular superimposition, and supposing that the symbols in *L* for conjunction and negation are "&" and "~", users of *L* will have expressed their elementary arithmetical insight with the following sentence: *X* & ~*X*. Contradiction.

The moral that Henle and Pletcher wish to draw from this demonstration is that a coherent idea—even a correct idea—may issue in a contradiction on account of limitations in the expressive resources of the language:

> What I want to suggest . . . is that we consider the possibility that the mystics have discovered that our conceptual system is deficient in some ways. It may be that what the mystic is experiencing is genuinely ineffable . . . relative to the conceptual scheme which our language and thought embodies. Hence the mystic's tendency toward paradox may be a fault of the descriptive concepts he has to employ, rather than some defect in the description of his experience.
>
> (Pletcher, 1973, 205)

Henle's and Pletcher's point is that we are not compelled to accept the orthodox view that the mystic's (or the set theorist's) contradictions reveal that the thought being expressed is incoherent. Their demonstration purports to show that a thought that generates a contradiction in one language may be representable in a noncontradictory way in another language. *My* point is that if this demonstration does what it purports to do, it also provides an alternative to Moore's and Rucker's thesis that if a coherent thought generates a contradiction in one language, then it can't coherently be expressed in any language—i.e., it's logically ineffable.

Does the demonstration work? I don't think so. I think there is a far more natural way to describe what is going on in Henle's and Pletcher's scenario. In the language L, the symbol X which is produced by superimposing the symbols for 2, 3, and the greater-than relation is clearly *ambiguous*. One and the same symbol, namely X, represents both the true proposition that 3 is greater than 2 and the false proposition that 2 is greater than 3. So it's not simply a consequence of arithmetic that $X \,\&\, {\sim}X$ is a true sentence of L—it's a consequence of arithmetic only when the first occurrence of X is given one of the two possible readings and the second occurrence is given the other reading. If they're both given the *same* reading, then $X \,\&\, {\sim}X$ is unproblematically false. To be sure, it isn't possible to state in L itself which reading we are ascribing to any token occurrence of X; Henle's and Pletcher's demonstration does reveal the existence of weak ineffabilities with respect to the language L. But what the demonstration fails to show is that the derivation of a contradiction may be due to deficiencies unique to the language in which the contradiction is derived. The *apparent* contradiction is adequately explained by the fact that the expression X is ambiguous. Indeed, there's no reason to suppose that this explanation couldn't be given in the language L itself (though it wouldn't be possible to

specify, within *L*, what the two senses of *X* are). *X* & ~*X* isn't a contradiction in *L* in any sense that has to be worried about. It isn't any more worrisome than the claim that John (Locke) is not the same person as John (Donne). The upshot is that it hasn't yet been shown that there's a third way to deal with contradictions besides dismissing them as nonsense (the orthodox view) or attributing them to logically ineffable insights (Moore's and Rucker's view).

But how can Moore and Rucker, as well as Henle and Pletcher, dispose of the orthodox view? All four of these authors wish to perform the same conceptual maneuver: confronted with a self-contradictory statement, they want to resist the inference that the speaker is talking nonsense, and attribute the contradiction instead to the speaker's having an insight that's inexpressible in her language (Henle and Pletcher) or in any language (Moore and Rucker). Let's call this maneuver a *henle-moore*. It's worth noting that henle-moores have been performed in contexts other than set theory and religious mysticism. I recently had occasion to discuss two instance in print (Kukla, 2000). One of them concerns Latour and Woolgar's (1979) social constructivist hypothesis that scientific facts are constituted by the negotiating activities of scientists. One difficulty with this view is that the scientific facts that are constructed routinely carry dates that are earlier than the dates of their construction. For example, Latour and Woolgar want to say that the fact that the hormone TRH occurs in mammalian brains was constructed in 1969; but they don't want to say that prior to the construction in 1969, mammalian brains were devoid of TRH. What they end up saying is that it became true in 1969 that TRH had existed for eons before 1969. This is as straightforwardly incoherent as saying that next week Tuesday will come before Monday. But in his review of their book, Ian Hacking suggests that Latour and Woolgar's temporally incoherent gloss is necessitated by the fact that "the grammar of our language prevents us" from speaking the truth in this case (Hacking, 1988, 282). According to Hacking, the strictly nonsensical formulation serves to point up the fact that "what logicians would call the modality and tense structure of assertions of fact is misunderstood" by our ordinary ways of speaking (281). In other words, Hacking pulls a henle-moore. The second henle-moore discussed in my book is pulled by Margolis (1991) on behalf of the thesis of relativism. Margolis concedes that the relativist thesis can't be stated in a coherent manner; but he holds

out the hope that it will turn out to be coherent when embedded in a radically different conceptual scheme, perhaps a three-valued logic.

As I noted in my 2000 discussion, the problem with henle-moores is that they provide us with altogether too facile a means for extricating ourselves from any and all conceptual difficulties. Whenever somebody discovers an incoherence in what you have to say, you can blame the language! It seems to me that a disciplined rationality must at a minimum regard the orthodox view of incoherence—the view that linguistic incoherence betokens incoherence of the underlying thought—as the default option. It's the default because it's the one and only way we have of picking out incoherent theses. If the derivation of a contradiction doesn't pick out the incoherent theses, what does? How would anything ever be found to be incoherent? If you want to resist the movement from linguistic incoherence to cognitive incoherence in some specific case, you owe us an account of what makes this case different from the general run of cases. The most persuasive account that I can imagine would be a demonstration that the linguistic incoherence can be made to vanish when we express the insight in an alternative language. This is the resolution that Hacking and Margolis *point* to, although they take no concrete steps toward providing it. It's one thing to claim that the temporal logic in current use isn't up to the task of expressing Latour and Woolgar's temporal insights; it's another thing actually to construct an alternative temporal logic that does the job.

Henle's and Pletcher's argumentative goal is less ambitious. They try to show that there are languages in which insights come out as contradictions, even though there is no problem expressing these insights coherently in English. Even if it were successful, this demonstration wouldn't show that the mystic's or the set-theoretician's or the social constructivist's or the relativist's linguistic incoherence cloaks a valid insight. But at least it would establish the general claim that a contradiction may *sometimes* be due to trying to squeeze a valid insight into too restrictive a linguistic mold. The demonstration fails, however: the contradiction that's supposed to have been generated isn't a contradiction in any challenging sense. So it hasn't yet been established that the general claim underlying Henle's, Pletcher's, Hacking's, and Margolis' specific claims is even available as a theoretical option. The default judgment of incoherence stands.

Maybe someone will yet succeed in providing the demonstration that Henle and Pletcher failed to provide. That would vindicate the henle part of the henle-moore—the part that says that a thought which comes out as a contradiction in one language may be coherently expressible in a different language. But it wouldn't help the moore part—the part that says that a thought which comes out as a contradiction in *every* logically possible language may still be a coherent thought.

In sum

What general conclusions have been drawn about the argument from mysticism? To begin with, no one has yet demonstrated that any grade of metaclaim is incoherent. Even the strongest metaclaim that the mystical experience is logically ineffable is still in the running. This absence of refutations provides us with the grounds for a weak presumption that ineffable insights of all grades are a coherent possibility. That's all I have to say about nomological and higher ineffabilities. The coherence of the hypothesis that there are humanly ineffable insights is more than a presumption. I provided a model for such insights based on Mentalese, thus showing that they are a coherent possibility if the concept of Mentalese is coherent. Weakly ineffable insights are not only a coherent possibility—they actually occur in children who have mastered only Prenglish. This observation leaves it open whether insights occur that are ineffable in complete natural languages. I expressed the opinion that this hypothesis is verifiable in principle on the basis of data relating to mystics' reported inability to express their insight, but that the requisite data are not in.

These conclusions can be supplemented somewhat by considering the conjoint consequences of the arguments from mysticism and from epistemic boundedness. One of the arguments from epistemic boundedness—Fodor's—provides adequate warrant for the thesis of human ineffability. This augments the mystical argument's conclusion that human ineffability is merely a coherent possibility. It also augments the mystical argument's conclusion that weak ineffability relative to natural languages is both coherent and verifiable—for if human ineffabilities actually occur, then so do weak ineffabilities relative to natural languages.

In sum, the series that comprises the grades of ineffability has a natural break at the level of human ineffability. In the balance,

it's currently rational to believe in the actual occurrence of weak and human ineffabilities. By contrast, nomological and higher ineffabilities are merely speculative possibilities. But it's not crazy to entertain them.

3 Believing the mystic

I call someone a *mystic* if she claims to have had an ineffable insight. This characterization of mysticism is roughly the same as William James' (1902/1985). In his seminal treatment of the subject, James describes mystical states as "states of insight into depths of truth unplumbed by the discursive intellect" (380). There are two differences between this description and the one with which I'll be operating. First, I don't presuppose that the mystic's insight is veridical. (Neither does James, though it sounds as if he does in the quoted passage.) Second, instead of equating a mystical insight with an insight into a (purported) truth which is inaccessible to the "discursive intellect," I will say that a mystical insight is one whose content can't be put into words. I leave it open whether the inability to state the content of an insight entails, or is entailed by, the inability to plumb it by the discursive intellect. (This topic is discussed in the subsections in Chapter 2, "Does epistemic boundedness fail?", p. 82 and "Accepting the metaclaim: necessary conditions," p. 85.)

Understanding mysticism to be a type of claim rather than a species of knowledge, it's uncontroversial that there have been legions of mystics in diverse geographical locations and historical eras. Ineffable insights have been reported in the Christian mystical literature of Meister Eckhard, St. Theresa, and St. John of the Cross, in the Zen Buddhist lore on the experience of satori, in the Taoist and the Vedantist literatures, etc. Nor does the experience occur only in a religious context. Bucke (1901) and Happold (1963) provide autobiographical accounts of mystical experiences that are discussed in entirely secular terms. Of course, to say that mystical experiences have been reported in all times and all places is not yet to say that the *same* mystical experiences have been reported in all times

and all places. This far more controversial thesis will be discussed below.

There are evidently two parts to the mystic's claim. The mystic reports both a *substantive* insight into the workings of the world, and the (perfectly effable) *meta*-insight that the substantive insight is ineffable. These parts are independent: there is no inconsistency in accepting that the mystic has had a veridical substantive insight while denying her claim that it's ineffable, or that she can know that it's ineffable. Similarly, there's no inconsistency in accepting that the mystic has been struck by an ineffable idea while denying that there's any reason to lend credence to that idea. We could inquire into the grounds, if any, for accepting either part of the mystic's conjunctive claim. This part of the book deals with the grounds for accepting or rejecting the substantive insight, granting that the insight is ineffable. What does it mean for one person to accept or reject another's insight, when the other protests that the content of her insight is ineffable? What is it that's being accepted or rejected? I wish to assure the reader that I'll get to that issue eventually. For the time being, let's pretend that this isn't a glaring lacuna in the presentation.

Two preliminary points. First, it's a presupposition of my discussion that the notion of an ineffable insight or belief is coherent. (This presupposition is defended in Chapter 2.) Second, ineffability may refer to any of several grades of expressive incapacity: an insight may be inexpressible in the language that the mystic happens to speak, or in all humanly accessible languages, or in all logically possible languages (see the subsection "The grades of mystical ineffability," p. 79). It's my impression that all these claims can be found in the mystical literature (see the section on the grades of ineffability, p. 23). But the degree of ineffability being claimed makes no difference to the issue with which I will be dealing. At every level of ineffability, the question remains the same: what are the grounds, if any, for accepting or rejecting someone else's claim to have a veridical insight when the claimant can't put her insight into words? Note that this question would still make sense even if the higher grades of ineffability—e.g., ineffability in all logically possible languages—prove to be incoherent. The presupposition required for my discussion is only that the weakest notion of ineffability—inexpressibility in the mystic's language of discourse—is coherent.

Should the mystic's ineffable substantive insight be accepted? James' answer to that question comes in two parts, for the mystic

and the non-mystic are (quite reasonably) ascribed different epistemic states. To begin with, he concedes that "mystical states, when well-developed, usually are, and have the right to be, absolutely authoritative over the individuals to whom they come" (422). I won't evaluate this claim here, except to note that virtually every epistemological theory on the current philosophical table entails that there are *some* noetic experiences that usually are, and have the right to be, authoritative over the individuals to whom they come (e.g., perception, memory—I've intentionally elided the "absolutely").

This part of the book deals mainly with the social-epistemological question of what the *non*-mystic should make of the mystic's claim, and what he *might* be able to make of it, given various possible states of the evidence available to him. James' position on this question isn't entirely clear. He writes that "mystics have no right to claim that we ought to accept the deliverance of their peculiar experiences, if we are ourselves outsiders" (424). This sounds like a clear and unambiguous denial. But elsewhere he writes: "No authority emanates from [mystical states] which should make it a duty for those who stand outside of them to accept their revelations uncritically" (422).The difference between the two formulations turns on the adverbial qualifier "uncritically" in the second quotation. Its employment suggests that James' claim is only that we (outsiders) are not justified in accepting the mystic's claim *purely on her say-so*. This leaves it open that the total state of the evidence may (critically) warrant a defeasible acceptance of the claim. This interpretation is supported by the continuation of the first passage: "The utmost [mystics] can ever ask of us in this is to admit that they establish a presumption" (424). If a presumption were to be established in the mystic's favor, then the epistemic status of her claim would differ only in degree from that of any successful theoretical hypothesis in science.

There are commentators on the mystical literature who hold the view that James' assessment of the case for mysticism is too generous. The severest critics do not merely maintain that the evidential grounds for accepting the mystic's claim are weak or non-existent. They argue that *there cannot be* evidence for accepting the mystic's claim. This is represented to be an a priori truth. Steven Katz, for instance, writes:

> There are major, perhaps insuperable, problems involved in the issue of trying to verify mystical claims, if by verification we

> mean the strong thesis that independent grounds for the claimed event/experience can be publicly demonstrated. Indeed, it seems to me . . . that it is not possible to provide "verification" of this sort. As a corollary of this view it also seems correct to argue that no veridical propositions can be generated on the basis of mystical experience.
>
> (1978, 22)

Katz's qualms about public observability have their roots in the views of some of the logical positivists (Hempel, 1952, 22; Feigl, 1953, 11; Popper, 1959, 44). Feigl explicitly makes the connection between mysticism and the issue of publicity:

> The quest for scientific knowledge is . . . regulated by certain standards or criteria. . . . The most important of these regulative ideas are:
>
> 1. *Intersubjective Testability*. . . . What is here involved is . . . the requirement that the knowledge claims of science be in principle capable of test . . . on the part of any person properly equipped with intelligence and the technical devices of observation and experimentation. . . . If there be any "truths" that are accessible only to privileged individuals such as mystics or visionaries . . . then such "truths" are not the kind that we seek in the sciences.
>
> (1953, 11)

Strictly speaking, this pronouncement doesn't rule out the possibility that we may be justified in accepting the mystic's claim—it's just that it can't be part of the special subset of justified beliefs that goes by the name of "science." While this is all that Feigl says *here*, acquaintance with his broader corpus of (early) writings, as well as the scare quotes around "truth" in the quoted passage, make it clear that his position is that mystical "truths," like all other non-scientific claims, can never justifiably be believed. If Feigl and Katz are right, there is no need to examine the mystical literature in detail in order to assess the case for accepting the validity of the mystic's claim; the claim can be rejected as unsupportable on a priori grounds. Clearly, the first order of business is to evaluate this view.

It's worth noting that a lot of my discussion is applicable to a broader question than that of mystic validity. Most of what I have

to say applies to any epistemic claim made on the basis of an experience which is not available to others, whether or not the claim is effable. For example, revelations of God's will through divinely selected prophets are often unproblematically effable: the prophets have no difficulty telling us what God said to them. I will call such an experience a *revelation*; and I'll use Feigl's other term, *visionary*, to refer to one who reports revelations. According to my terminology, every mystic is a visionary, but not every visionary is a mystic. Given these definitions, the aim of this part of the book can be exactly described as an attempt to assess the merits of Feigl's hypothesis—that the "truths" accessible to mystics or visionaries are not candidates for justified belief by others. Not surprisingly, the additional condition for mysticism that the revelation be ineffable makes additional difficulties for the mystic's claim as compared to the non-mystical visionary's. For example, if the mystic's insight is ineffable, what is there for the non-mystic to accept or reject? I haven't forgotten my promise to answer this question.

Here's a preliminary answer to the question whether we can ever have sufficient grounds for accepting somebody else's revelatory claim. Questions of ineffability aside, the relation between the visionary and the nonvisionary is an instance of the relation between a putative *expert* and a *novice* who is unqualified to pass an independent judgment on the matter at hand. There are, of course, many different and conflicting accounts of the expert–novice relation. But all of them (with the possible exception of Locke's) agree that there are nomologically possible states of affairs wherein a novice is justified in relying on a putative expert's putative expertise. Indeed, almost everyone would claim much more than this. It's commonly agreed that all of us *do* often and justifiably rely on expert testimony. This additional claim doesn't figure in the present argument, however. The argument is this: since there are nomologically possible states of affairs wherein novices may rely on (putative) experts, and since the visionary–nonvisionary relation is an instance of the expert–novice relation, it follows that there are nomologically possible states of affairs wherein nonvisionaries may rely on the testimony of visionaries.

To be sure, this argument isn't ironclad: there may be special features of the visionary–nonvisionary relation that make it impossible for any of the justificatory states of affairs ever to be realized. The main candidate for such a difference is the one to which Feigl alludes: putative knowledge based on revelation fails to satisfy

the criterion of "intersubjective testability." This is "the requirement that the knowledge claims of science be in principle capable of test . . . on the part of any person properly equipped with intelligence and the technical devices of observation or experimentation" (Feigl, 1953, 11). The idea is that even though I may rely on expert testimony in forming my beliefs about what happens when a lump of sodium is immersed in water, it is always possible for me to do the experiment myself and thereby confirm or disconfirm the expert's claim. When it comes to the "mystic or visionary," however, the possibility for independent confirmation is absent. There are at least three reasons why Feigl's proposal should not be accepted.

First, the proposal is unmotivated. On Feigl's account, when the novice *can* subject the expert's claim to independent test, he is sometimes justified in accepting it even though he *doesn't* perform the independent test. This has to be admitted, for otherwise testimonial evidence would always be redundant. But this admission makes Feigl's requirement very puzzling, for whether or not the novice's acceptance is justified is made to turn, not on anything that the novice observes or thinks or does, but on some things that he *could* do. It seems prima facie plausible to say that the bare possibility of an independent test is sufficient to elevate a hypothesis to the category of those which are *candidates in the running* for justified scientific beliefhood. But this isn't enough of an admission to underwrite the usefulness of testimonial evidence. One has to admit that testimony sometimes yields justified knowledge in the absence of independent confirmation. But then why should the fact that one *could* obtain independent confirmation make an epistemic difference?

Second, it isn't clear that the purported difference between revelatory claims and ordinary claims is a real difference. The mystic's answer to Feigl could be that the mystical experience which is the justificatory basis of the mystic's own belief *is* available to anyone who engages in the appropriate spiritual exercises. This view is explicitly espoused in Hindu and Buddhist mystical texts, where detailed instructions are provided for the obtaining of mystical states of consciousness. It's not as often encountered in the Christian or secular mystical literature, where mystical experiences tend to occur adventitiously, "by God's grace," rather than as a result of intentional activity directed toward that aim. On this view, only a chosen few are privy to revelations, and there's nothing that the spiritually second-class can do about it. In order to sustain

his exclusionary thesis, Feigl would have to say that mystical and other revelatory states are capricious in this way. For the sake of the argument, let's suppose this is so from now until the end of Chapter 3.

The third argument against Feigl's thesis is that modern science accepts as justified numerous results which it would be physically impossible for any individual scientist to test. Hardwig (1985), for example, discusses an experiment in particle physics that had 99 authors and took 280 person-years to complete. No single individual could *possibly* conduct an independent test of the conclusions of this study. So Feigl's choice is either to deny large chunks of modern science or to give up his thesis. In fact, in order to preserve his thesis, he would have to give up almost the whole of science. For it's physically impossible for any individual investigator to test more than a vanishingly small proportion of the current stock of justified scientific beliefs. In trying to bar revelatory testimony from entry into scientific discourse, Feigl's principle leads to an extreme (although not necessarily total) skepticism.

I can see only four moves that are available to Feigl in order to block this conclusion. The first move is to grant that a novice need not be able independently to test *all* the testimonial claims which he's entitled to believe, but to maintain that he should still be able to test *any one of them*. I can't perform the whole 280-person-year experiment by myself, but I can test any single investigator's part of the study—and that's what differentiates this case from the case of revelation. There are two reasons why this first move won't work.

1 It ignores the cumulative nature of scientific knowledge. The point can be made by use of the following drastically oversimplified but nevertheless telling account of this accumulation. Let P(1), P(2), . . . be the sequence of all justified scientific beliefs listed in the chronological order in which they were established. Since the P(i) number is in the millions (at least), there is no possibility of my subjecting each and every one of them to independent test. But (the claim is made) I can independently test any one of them, and this is enough to allow the whole series into science. The problem is that by the time we get past, say, P(10,000), the supposedly independent test of any fact in the series is going to presuppose the justifiability of legions of prior facts. Any investigation that employs a microscope, for example,

is going to rely on a complex of theoretical and practical facts about the proper construction and use of microscopes and the proper interpretation of microscope images. If I rely on testimony to include this requisite information in the conduct of my test of P(10,000), then it's not an *independent* test. Therefore I can't independently test any but the first few scientific facts.

2 Even if science weren't cumulative in this way, it's not the case that anyone can subject any scientific claim to independent test. Many people—perhaps most people—would be unable to subject an arcane hypothesis of modern physics to an independent test even if they were granted a thousand uninterrupted years to work on it, and even if conducting the test were their only value in life. This could be due to any of a variety of irremediable intellectual, perceptual, motivational, etc., deficiencies. It surely shouldn't count against a hypothesis that it physically can't be confirmed by the blind, the lazy, the stupid, sufferers of attention deficit disorders, schizophrenics, hopeless Romantics, inveterate comedians, etc. The general point is that we must admit, once again on pain of acceding to an extreme skepticism, that a scientific hypothesis may justifiably be accepted even if there exist people who are constitutionally incapable of independently testing it.

A review of where we stand: we're in the midst of discussing the third argument against Feigl's exclusionary principle—the argument that the principle leads to extreme skepticism. I noted that a Feiglian had four potential responses to this argument, the first of which is to claim only that *any* scientific hypothesis must be independently testable by anyone. I've just finished giving two reasons why that response won't work. Now it's time for the second potential Feiglian response to the skepticism argument. This is to stipulate that it merely needs to be *logically possible* for anyone to test the hypothesis. It may be nomologically impossible to live long enough to subject a modern scientific hypothesis to a completely independent test, but there is no contradiction in supposing that we *might* have lived long enough. By the same token, the blind might have been sighted, the schizophrenic might have been sane, etc. This criterion might very well succeed in excluding introspective reports of one's own mental states from the realm of science (though Goldman (1997) doesn't think so). But it clearly won't help Feiglian exclusionism with respect to revelations of objective states of affairs:

just as there is no contradiction in supposing of any given person that he might have had all the personal qualities and abilities needed to confirm a quantum-mechanical hypothesis by himself, so also is there no contradiction in supposing of any given person that God might have granted him the same revelatory experience.

Here is the third move against the argument that Feigl's exclusionary principle leads to skepticism: instead of requiring that other *individuals* be able to establish the knowledge claim for themselves, allow that it's sufficient that other *groups* be able to do it. But which other groups? We can't require that *all* other groups have the capacity for establishing the knowledge claim, for that would include groups composed of single individuals—the group requirement would then reduce to the earlier individual requirement. It won't help to rule out singletons by fiat, since groups of two or three are still going to be unable to recapitulate more than a minute fraction of our scientific knowledge. In fact, increasing the minimum size of the groups that have to be able to establish knowledge claims isn't going to help, no matter how large a minimum is considered. Suppose we require that knowledge claims be certifiable by all groups containing at least one thousand members. Besides being a bizarrely arbitrary epistemic principle, there's also the problem that there are bound to be groups of one thousand intellectually or perceptually or motivationally deficient individuals that won't be up to the task.

The fourth and last rejoinder to the skepticism argument: what if we require that knowledge claims be certifiable, not by *any* other person or group, but by *some* person other than the original claimant (or some group which doesn't include the original claimant as a member)? Goldman (1997) repudiates this epistemic requirement on the grounds that we can coherently suppose that scientific research is conducted by individuals who are all causally isolated from one another, each in her own light cone, or by an individual who is the only sentient being in the universe. I'm not entirely sure that these scenarios are coherent. For one thing, one might object that Wittgenstein's private language argument casts doubt on the possibility that an isolated individual can have a language in which to formulate scientific claims. Be that as it may, requiring that scientific claims be certifiable by two individuals (working independently) is as arbitrary as requiring that they be certifiable by seven, or thirteen. Revelations granted to twin prophets may be more convincing by a degree than monoprophetic revelations, just

as revelations granted to thirteen prophets might be more convincing than when n=2. Whether this is so will be discussed later on. But the defensive move here is to say that when we drop from n=2 to n=1, the claim not only loses another measure of credibility—it ceases to be even a *candidate* for acceptance. This idea cries out for some sort of justification. Why postulate such a drastic and unique discontinuity between n=1 and n=2?

So Feigl's a priori objection to mysticism and revelation is untenable. Moreover, I know of no other a priori objections on the table. Thus there are at present no known reasons to suppose that revelations can never be accepted by the nonvisionary. Of course, this falls short of establishing that there are nomologically possible states of affairs wherein nonvisionaries *are* justified in accepting revelations (much less that such states of affairs actually obtain). We've seen that the visionary–nonvisionary relation is a species of the expert–novice relation. Goldman (2001) offers a taxonomy of the sources of evidence available to novices for determining whether a putative expert is really an expert whose testimony should be accepted. I will now try to ascertain whether any of these sources of evidence is open to the nonvisionary in his assessment of a revelatory claim. Goldman cautions that his taxonomy is not necessarily exhaustive. So there would be hope for the pro-visionary position even if none of Goldman's sources of evidence were available. But it would be incumbent on anyone who maintains a pro-visionary stance to come up with another, non-Goldmanian source of evidence which is available to the nonvisionary.

Goldman discusses five sources of evidence on which the novice might rely. The first source is *argumentative justification*: the novice may independently judge the expert's arguments to be good, or to be better than a competing expert's. There are two senses in which this might be so. In *direct* argumentative justification, the novice becomes justified in believing an argument's conclusion by becoming justified in believing the argument's premises and their support in relation to the conclusion. This process would be counted by a Feiglian as an independent test of the conclusion, for the expert merely provides the necessary ingredients for the test—her role could as well be played by a nonintentional process, e.g., a random typing machine that happens to type out a good argument which one happens to read and understand. In effect, the consumer of the direct argumentative justification becomes himself an expert on the topic at hand. *Indirect* argumentative justification arises from the idea that

one speaker in a debate may demonstrate dialectical superiority over the other, and this dialectical superiority might be a plausible *indicator* for the novice of greater expertise. For example, in a disagreement between two experts E1 and E2, the novice may not be able to follow the argument and therefore fail to obtain any direct argumentative justification for either side. But he may notice that whenever E1 presents an objection to E2's views, E2 has a ready answer, while the reverse is not the case. This could be taken as evidence that E2 has investigated the matter more thoroughly, having already taken into account E1's objections, while the reverse is not the case. This form of evidence is, however, inapplicable to the case of revelatory experience, because the visionary doesn't use argumentation. Her claim is akin to an eyewitness perceptual report at least to the extent of being believed on a non-discursive basis. James: "mystical experiences are as direct perceptions of fact for those who have them as sensations ever were for us . . . they are face to face presentations of what seems immediately to exist" (423–424).

The second of Goldman's five sources of evidence for preferring the opinions of one expert over another is that one of the experts is appraised as superior by meta-experts. Clearly, this cannot be the *only* source of evidence for accepting experts, for its applicability depends on our already having identified a class of experts on experts. In the case at hand, there could be (prima facie) two sources of meta-expertise about which purported visionaries are true visionaries. First, the meta-expert may herself be a visionary, and she may be checking the reliability of the purported visionaries by seeing how their revelations compare with her own. If we nonvisionaries are to accept such a person as a meta-expert on visionaries, we have to be able to form a justified belief in her revelations. Such evidence could conceivably be based on the testimony of meta-meta-experts. But it's clear that some other form of evidence must eventually be brought into play, on pain of infinite regress. The second possibility is that the meta-expert is not herself a visionary. Then, if she is indeed a meta-expert on visionaries, she must have found a source of evidence for warranting visionaries that doesn't depend on having the revelation oneself. Either way, there has to be another way to tell the true visionaries from the impostors.

The third source of evidence is that one of two competing experts betrays more signs of potentially distorting interests and biases than

the other. If one commission salesperson in a store claims that the one and only brand of lawn mower they sell is the best on the market, while another commission salesperson in the same store confesses that their mower is the worst on the market, we would be justified, ceteris paribus, in believing the second salesperson. The rationality of this maneuver is based, I think, on the principle of inference to the best explanation. When a testimonial suits the interests of the testifier, we have at least two explanations for why the testifier asserted what she did: (1) she is sincerely saying what she knows to be the case, or (2) she's saying it just to suit her interest. (There are more potential explanations than that, but the addition of further alternatives doesn't affect the point I'm about to make.) When the testimony *doesn't* serve the testifier's interest, explanation (2) is eliminated. Other things being equal, explanation (1) therefore increases in probability by virtue of having fewer competitors with which to share the probability pie.

How does this apply to the case of visionary expertise? In the case of effable revelations, it depends on the precise nature of what is revealed. There would be good cause for suspicion if the visionary reports that it has been revealed to her that those who send her money will go to Heaven. But what if the revelation is that the world is run by an evil incubus who despises humanity? To be sure, this scenario might be a pleasing one to some pathological individuals. But suppose that the visionary is visibly upset by her revelation, and that she is extremely reluctant to convey its content. Then there might be cause for worry.

But what about ineffable mystic insights? The issue of ineffability can't be finessed any longer. To begin with, what is it to accept the mystic's claim? It isn't to accept the very fact about the world that has been revealed to the mystic, since the non-mystic never gets told what that fact is. (If the non-mystic achieves a wordless understanding of the mystic's insight, then he becomes a mystic himself.) One could say that the non-mystic accepts the fact that the mystic has had an ineffable and veridical insight into the nature of the world. By itself this is pretty thin—it doesn't add much to or effect many alterations in our prior stock of beliefs. Thin as it is, it seems to me that Goldman's third type of evidence, if it has any bearing at all, speaks against acceptance of the proposition that the mystic's ineffable insight is correct. For there is an interest in getting this proposition accepted: to be seen as a privileged recipient of ineffable truths is to cut an attractive and impressive figure. In contrast, I see

no benefit to the mystic if she takes her own purported mystical experience to be nothing more than a stress-induced hallucination. At the very least it can be said that there is no reason to expect support from Goldman's third type of evidence for the acceptance of the thin claim that the mystic has had a correct ineffable insight.

But there's another potential source of information that the non-mystic may derive from the mystic's claim. Let me begin by observing that even those mystics who most clearly insist on the ineffability of their insight have some things to say about what their insight teaches them. James:

> In spite of their repudiation of articulate self-description, mystical states in general assert a pretty distinct theoretic drift. It is possible to give the outcome of the majority of them in terms that point in definite directions. One of these directions is optimism, and the other is monism.
>
> (415–416)

> That doctrine . . . that eternity is timeless, that our "immortality" . . . is not so much future as already now and here . . . finds its support . . . which floats up from that mysterious deeper level.
>
> (422)

Isn't this inconsistent with the protestations of ineffability? Not necessarily. One way to reconcile the apparent inconsistency is to say that the mystic has an ineffable insight, and that this insight has some effable consequences. On this account, the mystic's monism, her optimism, and her view that eternity is timeless are all entailed by her ineffable insight; but the insight itself is not exhausted by these or any other effable consequences. Does it make any sense to talk about the effable consequences of an ineffable insight? At least with respect to the lower and weaker grades of ineffability, it surely does. Consider the lowest grade of ineffability: inexpressibility in a given language. It's clearly possible to take a language L, remove from it all sorts of expressive devices until we get a fragment of L—call it L'—such that there are propositions in L which are (1) not expressible in L', but which (2) have consequences that are expressible in L'. Here's a concrete example: let L' be obtained from L by excizing all ways of negating sentences, and let (PvQ)&-P be a sentence of L. This sentence is ineffable in L',

but it has consequences (e.g., Q) that may very well be expressible in L'. Of course this rationale depends on there being a more expressive language in which the ineffable fact can be stated. The rationale won't work for the highest grade of ineffability—inexpressibility in all logically possible languages. I concede that, for all I know, the idea of there being effable consequences of truths that *are* ineffable in this very strong sense may be incoherent. In fact, I concede that the idea of logical ineffability itself may be incoherent. If the worst comes to the worst for the mystic's epistemic fortune, she will simply have to settle for a less-than-maximal grade of ineffability claim. But this retrenchment would not alter the coming analysis of the insight which she finds herself unable to express.

Now let's return briefly to the third Goldmanian source of evidence for certifying experts: their having no special interests or biases that incline them in the direction of what they claim to know, or better yet, their having the *opposite* interests and biases. How does this apply to the case when the expert is a mystic and her expertise is ineffable? Given that ineffable insights may have effable consequences, it is possible for an outsider to ascertain whether the mystic is biased in favor of the effable consequences. In my view, this source of evidence runs *against* the mystic. For the effable consequences adduced by the canonical mystics are very much in accord with what the human heart desires to be the case. It's (altogether too) reassuring to be told on the best authority that if we understood the workings of the world aright, we would have cause for unbounded joy, or that we would realize that we are immortal. So far as this particular type of evidence goes, mystics would be much more credible if they announced that it was a consequence of their ineffable insight that if we understood the world aright, we would be filled with fear and loathing, and that when you're dead, you're dead and that's all there is to it. So the good news for mysticism is that we have found a source of nomologically possible evidence that could be used in warranting the mystic's claim. The bad news is that the actual evidence is negative.

The fourth Goldman source of evidence is *consensus* among the putative experts: if they all answer the question in the same way, it's (defeasible) evidence that they've got the answer right. This is by far the commonest ground on which those who are sympathetic to the mystic's claim and those who are antipathetic to it have argued. There's a voluminous literature in the field of religious studies in which *perennialists* contend against *constructivists* over the issue of

consensus among mystics. The perennialists include Aldous Huxley (1944), Rudolph Otto (1932), Evelyn Underhill (1911), Frithjof Schuon (1975), Alan Watts (1954), Huston Smith (1976), and Robert Forman (1993). The constructivists include Steven Katz (1978), Robert Gimello (1978), Hans Penner (1983), and Wayne Proudfoot (1986). The perennialists maintain that the mystical experience is essentially the same in all cultures and all historical eras; the constructivists argue that the mystical experiences of each culture and each era are soaked through and through with that culture's or that era's conception of the world.

There are two preliminary conceptual problems that need to be addressed before engaging in the main perennialist–constructivist debate. First, what counts as having "essentially" the same experience? Qualitative identity in every respect is surely too strong a requirement. But every experience is similar to every other experience in *some* respect. What, exactly, is the perennialist claiming and the constructivist denying? This crucial question has not been sufficiently addressed in the perennialist–constructivist literature. To my mind, the appropriate basis for individuating mystical experiences is their *noetic content*. Two mystical experiences may differ in their emotional or perceptual content—one may be ecstatic and the other serene, one may involve visions of Buddhas and the other of Christ—but, for the purpose of the perennialism dispute, they should still be considered the *same* mystical experience so long they engendered the same ineffable insight.

Second preliminary issue: how can you tell whether two ineffable insights are the same? The constructivist Katz has argued that you can't tell:

> the use of the terms "paradox" and "ineffable" [in describing mystical experience] do not provide *data* for comparability, rather they eliminate the logical possibility of the comparability of experience altogether. Consider the following example: (1) mystic A claims experience x is paradoxical and ineffable; while (2) mystic B claims experience y is paradoxical and ineffable. The only logically permissible conclusion one can draw in this situation is that both mystic A and mystic B claim their experience is paradoxical; *nothing* can be said about the content of their respective experiences x and y for there is no way to give content to experiences x or y in such a manner as to learn anything about them, apart, as we have said, from their

> both being paradoxical, which could then serve as the basis of a reasonable comparison.
>
> (1978, 54–55)

If Katz is right, then we can be sure on a priori grounds that the perennialists will never be able to marshal the evidence they need for victory. Neither, of course, could there be data that support the constructivist thesis. But most constructivists would be willing to settle for a verdict of not proven either way (see, for example, Katz, 1978, 22–23). My suggestion that ineffable beliefs can have effable consequences provides a way out of this dilemma for perennialists: ineffable insights may be individuated by their effable consequences. That is to say, the fact that two ineffable insights yield identical effable consequences is evidence that they are identical insights. To be sure, there are no grounds for denying that two different ineffable facts may have the same effable consequences. But suppose it were established that mystics the world over have independently derived the same effable consequences from their ineffable insights. What could account for this state of affairs? The hypothesis that they have the same ineffable insight would provide a ready explanation. In fact, it strikes me as extremely implausible to suppose that mystics the world over could have had *different* ineffable insights, but that all these diverse insights happened to yield the same effable consequences. Other explanations for the presumed consensus of effable consequences can, of course, be contrived. One of them is that the mystics have *no* ineffable insight at all—that the supposed effable consequences are not in any sense derived—that they're bare posits—and that they're all the same because they are what human beings are constitutionally prone to posit. Supposing that a consensus of effable consequences were found, I see no reason to prefer this skeptical explanation over the perennialist explanation. At the very least, the perennialist hypothesis must be reckoned to be in the running.

A background assumption shared by perennialists and constructivists alike is that the consensus issue is crucial to the disposition of the issue of mystical warrant: if a consensus exists, then we should accept the mystics' claim, and if a consensus doesn't exist, then the claim should be resisted. This assumption is enshrined in the very names of the contending parties. Going by their etymologies, it's appropriate to use the term "perennialism" to refer to the hypothesis that the mystics of all eras are in agreement; but "constructivism"

is not a transparent choice for its opposite. The hypothesis that mystical reports are socially constructed certainly doesn't *mean* the same thing as the thesis that there is no significant degree of agreement among mystical reports. It does, however, immediately entail that the mystics' reports are not to be regarded as accounts of an independently existing state of affairs. But this is a hypothesis about the epistemic status of the reports, not their degree of unanimity. The fact that the issue is described as a debate between perennialists and constructivists reveals the implicit assumption that consensus and epistemic status are one and the same issue. Now it's true that consensus is *evidence* for truth. It's evidence because it's a datum that's well explained by the hypothesis that all the mystics have twigged on to one and the same objective fact. The reasoning is the same as that which lends greater credence to eyewitness reports when two observers make the same report. But consensus or dissensus alone certainly doesn't settle the issue. For one thing, there are other explanations of consensus. Goldman discusses one alternative at length: the consensus among putative experts might be due to the fact that all of them slavishly follow the lead of a "guru" rather than making an independent assessment. This particular objection could be countered by perennialists with the claim that the same reports are found in cultures and eras between which there was little intellectual exchange. But there are still other plausible alternative explanations of consensus. One of them is that the hypothesis in question conforms to a universal human bias—all the mystics believe it to be true because they all want it to be true. This isn't to deny that consensus *is* evidence for truth—it's just not as decisive as the participants in the perennialist–constructivist debate make it out to be.

The converse point is also worth making: dissensus doesn't automatically mean that accepting any mystical report is unjustified. It may be that there are only a few qualified mystics surrounded by legions of incompetent or dishonest colleagues. It's also possible that the mystics of different cultures and eras are simply reporting different truths. Moreover, it's even possible to have evidence that states of affairs like these obtain, and for this evidence to override the negative evidence of dissensus (if dissensus is indeed to be counted as negative evidence). We've seen one potential source of other-than-consensual evidence already: the relative lack of interest or bias in one claim as compared to another. Unfortunately for mysticism, the evidence from this source is not supportive. But other

forms of other-than-consensual evidence are coming soon. In sum, *anti-perennialism* (the thesis that perennialism is false) isn't the same thing as constructivism, and *anti-constructivism* (the thesis that constructivism is false) isn't the same thing as perennialism.

So when all is said and done, who's right—the perennialists or the anti-perennialists? The main anti-perennialist argument is exemplified in the following passage by Katz:

> *There are NO pure (i.e., unmediated)* experiences. Neither mystical experience nor more ordinary forms of experience give any indication, or any grounds for believing, that they are unmediated. . . . The notion of unmediated experience seems, if not self-contradictory, at best empty. . . . A proper evaluation of this fact leads to the recognition that in order to understand mysticism it is *not* just a question of studying the reports of the mystic after the experiential event but of acknowledging that the experience itself as well as the form in which it is reported is shaped by concepts which the mystic brings to, and which shape, his experience. To flesh this out, straightforwardly, what is being argued is that, for example, the Hindu mystic does not have an experience of *x* which he then describes in the, to him, familiar language and symbols of Hinduism, but rather he has a Hindu experience, i.e., his experience is not an unmediated experience of *x* but is itself the, at least partially, pre-formed anticipated Hindu experience of Brahman. Again, the Christian mystic does not experience some unidentified reality, which he then conveniently labels God, but rather has the at least partially prefigured Christian experience of God, or Jesus, or the like.
>
> (1978, 26)

In brief, the argument is that there *cannot be* equivalent mystical experiences across cultures, because all experiences are shaped by our conceptual apparatus and our conceptual apparatus is shaped by our culture. This argument carries its pedigree on its sleeve. It's an application to mystical experience of the thesis of the theory-ladenness of all observation. If mystical experience is thought of as a kind of perception, the classical and widely accepted arguments of Maxwell (1962) and Kuhn (1962) can be imported into the new domain, where they yield the consequence that mystical experiences are always affected by the mystic's conceptual system.

The perennialist Forman has tried to fend off the force of this

argument by maintaining that the extension of the Kuhnian model of perception to mystical experience is unmotivated:

> the constructivist's model was taken over wholesale from ordinary sensory or perceptual experience, and so far no one has plausibly explained why or how a model *for perception* should be applied willy-nilly to mysticism, which bills itself as *non*-perceptual.
>
> (Forman, 1993, 34)

Apparently, Forman is unfamiliar with the work of William Alston, whose *Perceiving God* (1991) is an extended and at least plausible discussion of why a model for perception might be applied to mysticism. I will not review Alston's arguments here. I will only say that in calling to question the perceptual model of mysticism, perennialists forfeit the use of a powerful ally in the defense of the thesis which is closer to their hearts than their perennialism, namely anti-constructivism—the thesis that mystical reports give us warranted beliefs about reality. For if mystical experiences are a species of perception, then they must be granted the same epistemic status as perceptions. Now the Kuhnian revolution has changed our view of what that epistemic status may be. But perceptually-based beliefs continue to be the prototype of rationally warranted beliefs. To place beliefs based on mystical experiences in the same category as perceptually-based beliefs is to put them in good company. In fact, it elevates the mystic's claims to the same epistemic level as the claims of empirical science. I should think that a lot of mystics and their fellow travelers would be content to accept this conclusion. Certainly Alston regards his perceptual theory as a defense of mystically-based beliefs.

Forman has another counterargument against Katz. He argues that mystics in cultures with vastly different conceptual schemes nevertheless report the same experience because their experience is one of pure, contentless awareness—a state of being conscious without being conscious *of* anything. Because the experience has no content, there is nothing for divergent conceptual schemes to differentially conceptualize. As Forman puts it, "it is on the back of . . . content . . . that cultural training is able to enter into and partially form an experience" (1993, 37). Hence, the argument goes, this candidate for a perennial experience does not fall within the scope of Katz's arguments.

Can there be such a thing as contentless awareness? It isn't obviously an incoherent notion. At least for the sake of the argument, I concede that such states occur. However, granting that these experiences exist and are transcultural, it's difficult to see how one might hold that they are *noetic*: how can pure awareness without content at the same time be a state of insight? Forman evidently envisions that this is at least a possibility:

> People *can* temporarily forget their constructive language, turning their backs on their traditionally furnished rooms and walking outside. They can get to some experience of whatever-it-is that is not constructed by language. They can, as it were, walk outside their culture's room.
>
> (1993, 40)

The claim is that states of pure awareness show us how the world looks without the distortions and biases produced by viewing it through the filter of our conceptual scheme. Among perennialists, this is a popular view of the noetic dimension of mystical states. Aldous Huxley, for instance, describes the mystical experience which is sometimes produced by the ingestion of mescalin as the experience of being shown "the outer and the inner world, not as they appear to . . . a human being obsessed with words and notions, but as they are apprehended, directly and unconditionally, by Mind at Large" (1968, 57). There is a palpable lacuna in the move from contentless awareness to its supposed epistemic significance. Forman tells us that "in pure consciousness, one has ceased perceiving, labeling, and thinking" (1993, 37). But if one has ceased perceiving, one can't very well be seeing how the world looks, with or without concepts. Evidently, the experience of "whatever-it-is that is not constructed by language" is not the experience of contentless awareness. It's an experience wherein there *is* visual content, but that content is not affected by our conceptual scheme. Forman's confounding of these two types of experiences is close to the surface in the following passage:

> To be entirely without thought or perception in the pure consciousness event *is*, it seems, to precisely lay aside the conceptual system which might enter as a mediating or shaping factor of thoughts and perceptions. We *can* forget our set for awhile. Moving outside the constructed room the mystic may

> encounter something that is not at all shaped by the expectations of that culture.
>
> (1993, 39–40)

In other words, if you want to be having perceptions that are unmediated by concepts, you should be in a state where you're entirely without perceptions!

What about this notion of experiences that have content, but that are unaffected by our conceptual apparatus? Well, it's a respectable (albeit controversial) position among philosophers of perception that there is a nonconceptual component to perceptual experiences (e.g., Peacocke, 1989, 1992; Crane, 1992). Interestingly, but parenthetically in the present context, Crane's characterization of nonconceptual content is incompatible with the conjunction of (1) Huxley's and Forman's identification of the ineffable mystical insight with learning how the conceptually unmediated world looks and (2) my idea that the ineffable mystical insight can have effable consequences. Together, (1) and (2) entail that the ineffable knowledge of how the conceptually unmediated world looks has effable consequences. But according to Crane, to have a concept just *is* to be disposed to make certain inferences. Thus a nonconceptual state is one from which no inferences can be drawn. This dilemma doesn't make problems for my view, (2), since, as will be seen immediately below, I reject (1).

The main argument for the existence of nonconceptual experiences is that

> an experience can have a finer-grained content than can be formulated by using concepts possessed by the experiencer. If you are looking at a range of mountains, it may be correct to say that you see some as rounded, some as jagged. But the content of our visual experience in respect of the shape of the mountains is far more specific than that description indicates. The description involving the concepts *round* and *jagged* would cover many different fine-grained contents which your experience could have, contents which are discriminably different from one another.
>
> (Peacocke, 1992, 111)

There are two differences between Peacocke's hypothesis and a view like Huxley's or Forman's. The first difference is that Peacocke

maintains that there is a nonconceptual *component* of perceptual experience, whereas Huxley and Forman envision the possibility of a *totally* nonconceptual state of perception. I don't think that this difference is significant: if our perceptual states have nonconceptual parts, then it seems compelling to suppose that the idea of a totally nonconceptual perception is at least a coherent possibility. A Huxley–Forman moment would be brought about when we enjoy a Peacockean nonconceptual component without its usual conceptual accompaniments. At the very least, the burden of (dis)proof would lie on those Peacockeans who would deny the coherence of the Huxley–Forman hypothesis—they would have to show us why Peacockean nonconceptual components must always be accompanied by conceptual baggage.

The second difference between Peacocke and Huxley–Forman is that the former regards nonconceptual experiences to be routine and universal accompaniments to every perceptual state, whereas the latter regard them as extraordinary events that occur only to special individuals (mystics) or in special circumstances (the ingestion of mescaline). Were it not for this difference, adherents to the Huxley–Forman hypothesis might have been able to avail themselves of Peacockean arguments to underwrite their notion of nonconceptual perception. As things stand, however, this warrant would have to be purchased at the cost of giving up the account of what sets the mystical experience apart from run-of-the-mill perceptual states.

The Huxley–Forman hypothesis also suffers from another liability. The idea is that nonconceptual experiences cause the experiencer to come to know the ineffable truth about what the world looks like without the filtering effect of her conceptual apparatus. Whether this state of affairs would indeed count as an instance of ineffable knowledge is a complicated question to which I presently have no answer. It may be a non sequitur to suppose that the way the world looks when we're not conceptualizing it is necessarily unconceptualizable. For one thing, there's the possibility that a conceptless experience can be effed by means of demonstratives, as in "The unconceptualized world looks like *this*." For another thing, there's the possibility of a *retrospective* description of the unconceptualized experience. The "fine-grain" argument shows at most that there is a nonconceptual portion of perceptual experience; it doesn't show that this portion can't be conceptualized. But suppose these qualms can be laid to rest. If this is the ineffable

secret that the canonical mystics have learned, it's difficult to understand why they ascribe so much importance to it. The message between the lines in Forman's discussion is that "the experience of whatever-it-is that is not constructed by language" is an experience of *reality as it is in and of itself*. But I don't see how unconceptualized experience can be regarded as privileged in this or any other way. If you take a *realist* perspective on the world, then conceptless experience is still going to be conditioned by the organs of perception. The way the world looks without concepts is no more the way the world *really* looks, in and of itself, than the way the world looks through Hindu or Christian eyes—it's still a way the world looks through *eyes*. The very idea that there is a way the world looks in and of itself is incoherent. Things have looks only for creatures that possess visual systems. If instead you take an *idealist* perspective, you get the same conclusion: conceptless and concept-laden experiences are in the same epistemic and metaphysical boat. Only the argumentative details are different. The supposed difference between conceptless and concept-laden experience is that the former reveals the world as it really is in and of itself. But idealism is precisely the denial that there is a world independently of our thoughts and experiences. For idealists, concept-laden experiences do not stand in need of a corrective. They are what they are, and that, for idealists, is the end of the story. Idealists therefore have no more basis for privileging conceptless experiences than do realists. Either way, the mystic's ineffable insight turns out not to be a big deal. Nevertheless, if conceptless experience does occur, it could produce knowledge of what the unconceptualized world looks like to beings who are endowed with sensory receptors like ours. Moreover, this knowledge *might* be ineffable; at least I've presented no argument to the effect that it wouldn't be.

The last few pages have been devoted to a critique and ultimately a repudiation of Forman's defense against Katz' argument for anti-perennialism. I turn now to my own critical appraisal of Katz's argument. Recall that Katz begins by enunciating the principle that there are no unmediated experiences. A common perennialist retort is that this principle is laid down without proof (e.g., Evans, 1989). On some interpretations, however, the principle is hardly controversial. If Katz means that every experience is dependent on some property of the experiencer, then the assumption that the notion of unmediated experience is "if not self-contradictory, at best

empty" follows directly from the principle that what we experience is causally dependent on our bodies. There are no doubt some mystics who would want to deny this principle. But the anti-perennialist argument can be de-fused on much less radical grounds. The problem with the argument is that the assumption of universal mediation isn't nearly strong enough to underwrite Katz's anti-perennialist conclusion. Mystical experiences might all be mediated by some neurophysiological condition, yet they could still be the same experiences the world over. In fact, before Katz gets to the comparison between the Hindu and the Christian mystic, the assumption of the argument has been made substantially stronger. It's that there are no experiences which are unmediated *by the experiencer's conceptual scheme*. This principle is no longer an obvious truth. In fact, it was noted above that some contemporary philosophers like Peacocke are arguing that experiences *do* have nonconceptual content. To be sure, these arguments aren't universally accepted. But it's pretty clear that nonconceptual experiences are at least logically possible. There's no contradiction in supposing that there are sentient beings whose experience consists of a mechanical mirror-like reflection of the scene before their sensory organs. The experience of these beings would still be mediated in the broad, initial sense of the term. For one thing, their experience would be causally dependent on the reflective properties of their internal mirror. But their conceptual scheme (if any) need not play a role. We may not be such mirror creatures ourselves. But now that it is clear that there is no *logical* necessity for concepts to shape our experiences, it becomes an empirical question whether we have any concept-free experiences.

Moreover, even the assumption that all experiences are mediated by our conceptual system is not strong enough to refute perennialism. For suppose we grant this assumption. It could still be the case that our conceptual apparatus is in part innate and universal (à la Jerry Fodor and Noam Chomsky) rather than socially determined. If this is so, mystics the world over could be enjoying the same concept-mediated experiences. So Katz needs to suppose (1) that all experiences are shaped by our conceptual apparatus and (2) that our conceptual apparatus is shaped entirely by our culture. But that's still not enough: different cultures may have produced overlapping conceptual schemes—schemes that contain some of the same concepts—and it might be just those overlapping concepts that go into the shaping of mystical experiences. To get to his

conclusion from his premises, Katz would have to add the assumption (3) that there are no conceptual commonalities at all across cultures. This requirement might be met by an appeal to *conceptual holism*, according to which concepts are individuated by their place in the total system of concepts—so that if there are any conceptual differences at all between two conceptual schemes, *all* their concepts will be different. But that still wouldn't be enough. Even if it's granted that every experience is shaped by our conceptual apparatus, that this apparatus is totally determined by culture, and that no two cultures overlap conceptually, it still doesn't follow that *every feature* of every experience is shaped by our conceptual apparatus. In fact, Peacocke's "fine-grain" argument discussed above renders this hypothesis quite implausible. But then it may be the case that what makes an experience a mystical experience is its possession of this concept-free feature. In sum, a priori refutations of perennialism are not to be expected. If we want to settle the perennialism–constructivism dispute, there's no avoiding the need to ascertain what the mystics actually say about their experiences.

What do the mystics say? Not surprisingly, most of the debate between perennialists and antiperennialists consists of a battle of quotations: the perennialists juxtapose passages from the world's mystical literature that sound as though they're saying more or less the same thing; and the anti-perennialists juxtapose passages that sound incompatible. The task of deciding which is the better supported hypothesis is by no means an easy one. Moreover, both sides have available, and make much use of, Duhemian strategies for attributing apparently disconfirming evidence to the failure of auxiliary hypotheses. On the anti-perennialist side, Katz notes that even identical reports don't necessarily betoken identical experiences:

> language is . . . contextual and words "mean" only in contexts. The same words—beautiful, sublime, ultimate reality, ineffable, paradoxical, joyful, transcending all empirical content, etc.—can apply and have been applied to more than one object. Their mere presence alone does not guarantee anything: neither the nature of the experience nor the nature of the referent nor the comparability of various claims is assured by this seemingly common verbal presence alone.
>
> (1978, 47)

Conversely, on the perennialist side, Stace (1972) makes the point that very different reports could be different *ex post facto* interpretations of identical experiences. Another hypothesis-saving move available to perennialists is to say that patently different mystical reports are reports of different facets of one and the same experience. It's significant that neither of these alternative explanations for apparently disconfirming data conflicts with Katz's theoretical assumptions—reports may differ for these reasons even if it's conceded that all experiences are shaped by our concepts. Similarly, perennialists have no basis for repudiating the alternative, non-perennialistic explanation of identical reports proffered by Katz. Still another source of interpretative difficulty is the fact that many of these contrasting passages are written in different languages, so that the issue inherits all the practical and philosophical problems involved in translation. Given the same corpus of mystical accounts, reasonable and unbiased investigators may very well disagree about how well the consensus hypothesis is supported.

The reference to translation suggests an interesting strategy for settling the dispute about perennialism. There are similarities between the question whether two mystical experiences obtained in different cultural milieus are noetically equivalent and the question whether two sentences in different languages have the same truth-conditions. The most straightforward way to settle the second question is to solicit the testimony of bilinguals, for they can tell us whether the two sentences say the same thing by inspection. If we could find "bicultural" mystics—mystics who had obtained experiences that were accredited by, say, the Catholic Church, as well as other experiences that were accredited by a Zen master—they might be able to tell us by inspection whether the two types of experiences were relevantly the same. This procedure might be objected to on the ground that to obtain the bona fide Catholic mystical experience, one has to believe in the Catholic theology, and one cannot believe in both Catholicism and Buddhism at the same time. However, I see no reason to believe that the effect on experience of a conceptual scheme or theory requires an attitude of exclusive belief in the truth of that scheme or theory. I've discussed this issue elsewhere in relation to the theory-ladenness of perceptual experience:

> it isn't at all obvious that assembling a perceptual system requires us to believe in the theory that it embodies. I don't have

> to believe that what I see is really a duck when I see the duck-rabbit as a duck; I don't even have to believe that ducks *exist*. By the same token, I bet that I could learn to see the phenomena relating to combustion phlogistically without believing the phlogiston theory. I just need to have mastered the conceptual machinery of the theory—I don't have to be committed to it. Similarly, if belief in elementary particle theory were a precondition for seeing particle tracks in cloud chambers, then you wouldn't be able to see them if you were an *instrumentalist*, no matter how steeped you were in the theory. And if a physicist were persuaded by van Fraassian arguments to *become* an instrumentalist, his perceptions would automatically revert to the layman's. All this strikes me as utterly implausible.
>
> (Kukla, 1998, 123)

So this avenue for the resolution of the perennialism issue is open. Whether there are or ever have been bicultural mystics is another matter.

Finally there's an anti-perennialistic argument that convinces me that perennialism is false. The argument is due to William James, and it makes no use of any constructivist assumptions about the concept-ladenness of experience and the like. Suppose that perennialists are able to persuade us that there have been many identical mystical experiences in every culture and every era. The evidential value of these data would be utterly defeated if there were *also* experiences in all cultures of ineffable insights with *incompatible* effable consequences. But, as James notes, there certainly are some such conflicting reports:

> The classic religious mysticism . . . is only a privileged case. It is an *extract*, kept true to type by the election of the fittest specimens and their preservation in "schools." It is carved out from a much larger mass; and if we take the larger mass as seriously as religious mysticism has historically taken itself, we find that the supposed unanimity largely disappears . . . religious mysticism is only one half of mysticism. The other half has no accumulated traditions except those which the text-books on insanity supply. Open any one of these, and you will find abundant cases in which "mystical ideas" are cited as characteristic symptoms of enfeebled and delusional states of mind. In delusional insanity, paranoia, as they sometimes call

> it, we may have a *diabolical* mysticism, a sort of religious mysticism turned upside-down. The same sense of ineffable importance in the smallest events, the same texts and words coming with new meanings, the same voices and visions and leadings and missions, the same controlling by extraneous powers; only this time the emotion is pessimistic: instead of consolations we have desolations; the meanings are dreadful; and the powers are enemies to life.
>
> (1902/1985, 424–426)

In brief, the accusation is that the corpus of experiential reports which both perennialists and constructivists have agreed to examine is not a random sample of reports that satisfy the definition of "mystical." It has been pre-selected in a manner that favors the perennialist thesis. When the sample is broadened, all semblance of consensus vanishes. Thus, while consensus is a nomologically possible source of evidence for accepting the mystics' claims, the actual state of the evidence is not supportive. In fact, the evidence positively supports anti-perennialism. This doesn't mean that the canonical mystics don't have a valid ineffable insight. Their valid insight may be buried in a sea of lesser mystics' delusions. Be that as it may, there's no support for the validity of any mystical insight from consensual evidence.

Now to Goldman's fifth: a novice may justifiably accept the testimony of an expert if she has independent access to information about the expert's past *track record* of cognitive successes in the relevant field. We may not ourselves be able to follow an astronomer's calculation of the next solar eclipse, but if we see that her calculation proves to be correct, it is reasonable for us to believe that the theory informing the calculations is also correct. Demonstrable technological expertise is also evidence that the expert has expert knowledge—at least if the expert claims that the technological feats depend on the knowledge. To be sure, the expert's claim in this regard may be wrong. It's commonly granted in our society that the shamans of non-Western cultures have a certain technological expertise in the field of introspective psychology or psychotherapy—they are able to effect some psychological cures; but it's just as commonly denied that the shamans have the right theoretical understanding of how these effective procedures work. We could entertain the same qualms about a straight prediction whose derivation we can't follow—for all we nonastronomers

know, it may be that the astronomer's calculations are merely a device for focusing her mind in such a way that the right intuition comes through. Nevertheless, the hypothesis that the expert's predictions come from her possession of a true theory is a good explanation for why the prediction is borne out. And it's the same with technological feats. Like correct prediction, technological prowess is defeasible evidence of knowledge.

Do mystics have demonstrable capacities (including the capacity for making novel predictions) which (1) they attribute to their ineffable knowledge, and (2) which can't be accounted for by any other accepted lore (we won't be impressed by a mystic who claims that her ineffable knowledge enables her to predict eclipses)? Numbers of mystics don't matter here—all we need is one good mystic with an extraordinary track record. Occult powers would do the trick: if we could be persuaded that a mystic is adept at telepathy and levitation, and if the mystic attributed these powers to her ineffable insight, her epistemic relation to the non-mystic would be approximately the same as that of the astronomer, with her precognition of eclipses, to the nonastronomical layperson. It would be reasonable for us to give some credence to the ineffable parts of her claim. There are legions of vulgar mystic claimants who do maintain that their mystical insights have endowed them with occult powers. But the canonical mystics of both East and West eschew such claims. In fact the canonical literature contains many admonitions to the novice not to confuse the mystic quest with the quest for occult powers. Still, it's worth noting that occult powers provide us with *possible* evidence to credit the ineffable claim. The claim is therefore not beyond the reach of confirmation.

Another possibility is expertise in the field of the human spirit. What often precipitates an interest in the mystical hypothesis is an encounter with a mystic who displays a remarkable degree of equanimity, unflagging benevolence, and an enormous capacity for pleasure. Let me put some data on the table. Here is the account of Matthieu Ricard, a biologist turned Buddhist monk, in his attempt to explain—and justify—his conversion to his scientistic and skeptical father:

> I stayed [at a Tibetan Buddhist lama's house], simply in his presence, for the next three weeks. It left a deep and unforgettable impression on me. He was a man of seventy, radiating goodness and compassion. . . . I received few words of teaching,

> almost none. . . . It was his personality, his being, that made such an impression on me; the depth, strength, serenity, and love that emanated from him. . . . That special quality of his kept coming back to my mind all the time. I became aware that I'd found a reality that could inspire my whole life and give it direction and meaning, even if I still couldn't say exactly how. . . . I only started studying [Buddhism] quite a bit later.
>
> (Revel and Ricard, 1999, 10–11)

And here is R. M. Bucke's description of Walt Whitman, who is counted among the non-denominational mystics by James:

> Perhaps, indeed, no man who ever lived liked so many things and disliked so few as Walt Whitman. All natural objects seemed to have a charm for him. All sights and sounds seemed to please him. He appeared to like (and I believe he did like) all the men, women, and children he saw (though I never knew him to say that he liked anyone), but each who knew him felt that he liked him or her, and that he liked others also. I never knew him to argue or dispute, and he never spoke about money. . . . When I first knew [him], I used to think that he watched himself, and would not allow his tongue to give expression to fretfulness, antipathy, complaint, and remonstrance. It did not occur to me as possible that these mental states could be absent in him. After long observation, however, I satisfied myself that such absence or unconsciousness was entirely real.
>
> (quoted in James, 1902/1985, 84)

One more bit of lore belonging to the same class. During the Vietnam war, a succession of Vietnamese Theravada Buddhist monks immolated themselves to protest the war. Most of us have seen the film. They sat cross-legged on the ground and quietly, without any appearance of anxiety or excitement or tension, set themselves on fire. They continued to sit quietly and without any appearance of tension as their bodies were consumed in flames, until their charred remains slowly toppled over in a heap. Questions of political advisability aside, I'm quite certain that nobody I have ever met could have behaved in this manner.

These are undoubtedly remarkable individuals. The question is: might their behavior also be evidence that they know something we don't? Let me start by asking the same question in another arena,

where we're pretty confident that the answer is yes. I observe a team of nuclear physicists causing an enormous detonation—they call it a "chain reaction." I'm certain that I couldn't duplicate their feat. Does that mean they know something I don't? Well, there are four possibilities.

1 One possibility is that the physicists possess theoretical knowledge such as $e=mc^2$, from which they derive practical knowledge such as that if you pull out these rods, you get a chain reaction, which dictates a skilful action (pulling out the rods) which produce the effect (the chain reaction). By "practical knowledge," I mean knowledge of conditionals that specify that a verifiable effect will follow if a particular action is performed. By "theoretical knowledge," I mean broad principles (not necessarily involving nonobservational terms) that *explain* and enable one to *predict* the truths of the practical principles.
2 Another possibility is that the physicists' knowledge is only practical. Instead of deriving the practical principles from theoretical principles, they've stumbled on a successful recipe for producing enormous explosions by a process of trial and error.
3 Alternatively, they may not even be guided by practical principles, any more than we rely on principles when we want to walk across the room. The skilful action may follow on the heel of the intention without the intervention of any principles. When the physicists want an explosion, they make one happen —but they can't tell us how they do it.
4 Finally, the production of explosions may not even be a voluntary act. It may instead be akin to a shudder—or better, to a case of spectacular and uncontrollable flatulence.

The physicists themselves know that the first scenario is the one that obtains. But how is a novice to know? Let's look at the possibilities in reverse order. What sort of evidence would impel us to reject (4) and posit that the production of explosions is at least a voluntary act? This one's easy: we reject (4) if we observe a relationship between the physicists' clear intentions and the production or nonproduction of explosions. The nuclear blasts aren't an involuntary form of flatulence if they can be held back when it's wise to do so.

Now what about going from (3) to (2)? Granting that the physicists can produce explosions at will, what might persuade us that this capacity is based on the knowledge of practical principles? At first glance, the answer seems obvious: if the physicists base their skilful action on their knowledge of practical principles, they ought to be able to tell us how to make the explosions ourselves. They may not be able to convey their esoteric theoretical explanation for why the practical principles work, but they should be able to tell us what the practical principles are. Certainly their provision of recipes that we can follow for explosions would be sufficient to reject the third scenario and move to the second. But is it necessary? Having granted that there can be esoteric theoretical knowledge that's inaccessible to the novice, don't we also have to grant that there can be esoteric *practical* knowledge as well? May it not be the case that the physicists can produce explosions at will, and that they follow a practical recipe in doing so, but that the recipe is incomprehensible to the novice? On the other hand, is it conceivable that there can be effective practical knowledge that doesn't have at least some exoteric practical consequences that are transparent to the nonspecialist? Doesn't even the most esoteric lore ultimately entail principles such as that if you turn the knob on the left in a clockwise direction, a red light will flash? I've placed the foregoing disquisition entirely in the interrogative mode because I have no answer to offer. But at least it can be said that there are possible data that would impel us to move from (3) to (2), namely the provision of practical instructions that can be followed by novices. Do the physicists pass this test? Could they come up with a cookbook description of how to make an atomic bomb that could be followed by someone who is utterly ignorant of the theory? I think it's pretty clear that they could, though the cookbook is likely to be very long.

Finally, the move from (2) to (1). What would convince a novice that the practical knowledge of how to make big explosions is derived from a theoretical understanding? One source of evidence might be the possession of many different but related practical principles. The reasoning once again relies on inference to the best (or at least to a good) explanation: one way of accounting for the possession of many different but related practical principles is to say that they are all consequences of the same theoretical knowledge. This would be very weak evidence, however, since the diverse practical principles could also have been obtained by extensive engineering investigations of the field of interest. But what if the

experts can produce *new* practical knowledge *without* engaging in any new engineering investigations? If they can correctly answer new questions of the form "What will happen if I do this?" simply by thinking about them, we would have good reasons for supposing that they have a theoretical understanding of the practical effects. In sum, the usual sources of evidence for accepting a proposed theory are also evidence for the possession of a theory that's been verified.

Indeed, when new practical knowledge is produced without new engineering investigations, it might appear that its being derived from theoretical principle is the *only* available explanation. The claim that this is so is akin to the "miracle" argument for scientific realism (Putnam, 1975). According to this argument, the predictive success of scientific theories can only be explained by the truth of their principles, which in turn entails the existence of the theoretical entities posited by those principles. The claim is that the predictive success of the theories would be a miracle if the theories weren't true. The new miracle argument is that the predictive success of *scientists* can only be explained by their *possession* of a true theory. In fact, the new miracle argument is stronger than the old one. The old argument that the predictive success of theory *T* can only be explained by the truth of *T* has a traditional antirealist counter: the predictive success of *T* can also be explained by the weaker hypothesis that *T* is merely empirically adequate rather than true —i.e., that the observable world behaves *as if T* were true (van Fraassen, 1980). If the empirical adequacy of *T* is accepted as an adequate explanation of *T*'s predictive success, the old miracle argument is effectively refuted: there is an explanation that doesn't require the existence of the theoretical entities posited by *T*. But the new miracle argument remains unscathed. In the new miracle argument, the scientist's possession of an empirically adequate theory isn't an *alternative* explanation to the hypothesis that her predictive successes are due to the knowledge of broad theoretical principles—it's another instance of the same hypothesis. The fact that the alternative explanation doesn't require the existence of theoretical entities is irrelevant to the new argument. The knowledge that the world behaves as though *T* were true is itself knowledge of a broad theoretical principle in the relevant sense of "theoretical principle." For all that, the new miracle argument still doesn't work, for there are other explanations for the predictive success of scientists besides their possessing either a true theory or an

empirically adequate theory. One of them is that new practical insights tend to come to those who have already a lot of practical knowledge in the field. So the possession of a true theory isn't the only way to explain the physicists' predictive success—but it's a good way.

Now what about the mystics? My analysis of the mystical case falls directly out of the foregoing discussion on the basis of the following principle: from the viewpoint of the novice, *it makes no epistemic difference whether the expert's theory is ineffable or merely incomprehensible to him*. If the mystic could manifest occult powers at will, and give practical recipes that can be followed by novices for producing occult effects, and come up with new, never-before tested recipes for occult effects, then the case for the novice's believing that the mystic has ineffable knowledge would be about the same as the current case for novices' believing that physicists have knowledge which is incomprehensible to them. Of course, such evidence of occult powers doesn't exist. But the conclusion already makes the anti-Feiglian point that the possession of practical skills may, under the right circumstances, rationally persuade us that the possessor has ineffable knowledge.

What about Walt Whitman's and the monks' extraordinary traits? The same four possible explanations are available here as were available in the case of the physicists' propensity to produce large explosions. The most reductive way to understand the self-immolating monks' fearlessness and imperviousness to pain is simply to say that that's what these people are *like*. They may be more admirable than whiners and cowards, but they're not different from whiners and cowards in their relation to knowledge. It may be true that an extraordinary religious experience caused them to have this character, and this may be reason enough to value and seek out the experience. But in this case, the experience would be valued for its *causal* as opposed to its probative or noetic properties.

Alternatively, the monks' fearlessness may be the result of a skilful performance—a manifestation of technological expertise in the internal manipulation of one's own mental states. We'd all like to be fearless and impervious to pain, and the monks know how to do it. How can one tell the difference between a passively endured (or enjoyed) character trait and a skilful performance? By its relation to the actor's intentions. There's a problem here for the skilful performance theory of the monks' fearlessness: since the monks' fearlessness, as well as Walt Whitman's pleasure and benevolence,

are always *on*, we novices have no opportunity to observe their correlation with the comings and goings of intentional states that would be evidence of a skilful performance. It's as though a saint were always levitated two inches above the ground—we'd suspect that the levitation might be something that happened *to* her rather than something that she did. Nevertheless, it has to be conceded that the monks' fearlessness, as well as the saint's perpetual state of levitation, could be the result of something that they did. The inability to turn these states off is only one explanation for their always being on. Another is that the actor chooses to leave them always on. This would admittedly be implausible in the case of the levitating saint—why would anyone always want to be two inches above the ground? But it makes perfectly good sense in the case of the fearless monks: if you know how to turn fearlessness on, why would you ever turn it off? In sum, the evidence doesn't favor the hypothesis that the monks' fearlessness is a skilful performance; but it doesn't rule it out either.

The third possibility is that the monks' fearlessness isn't only a skilful performance, like walking or riding a bicycle, but that it's produced by the application of practical principles. If Whitman and the monks could have provided us novices with recipes for becoming fearless, the issue would have been settled in an unambiguous manner. Now the scriptures of Buddhist and other mystical traditions do offer recipes involving spiritual exercises such as meditation, devotion, and austerities. But these are not the kinds of recipes that can be applied by a novice who wishes to test the would-be mystical expert's bona fides. In fact, they are recipes for obtaining mystical experiences and therefore becoming an expert oneself. It's not easy to say whether mystics are able to supply practical instructions that can effectively be applied by novices qua novices. Perhaps the only way to find out is to make Matthieu Ricard's decision to put oneself under their tutelage. The state of the evidence may not compel such a move; but neither does it mark it as flagrantly irrational.

The fourth and final possibility is that the monks' fearlessness stems from the application of practical knowledge, and that this knowledge is in turn derived from the monks' ineffable insight. Having run out of evidence in support of the third scenario, it's not to be expected that some will turn up for the fourth. But at least we can get a grip on what would be required. In order to demonstrate that her practical lore stems from a broader insight, the mystic

would have to show us that she can derive new consequences from her insight. For example, in addition to the prototypical mystic capabilities for fearlessness, universal pleasure, and unflagging benevolence, the mystic insight may also enable one to produce effects of lesser value which are usually not exploited. Perhaps a good mystic knows how to strike terror in the hearts of the novitiate, or how to be preternaturally funny. Such demonstrations, while lacking the intrinsic value of fearlessness, might nevertheless possess instrumental value in helping to establish the mystic's claim that she knows something we don't.

In conclusion, the evidence that the mystics' fearlessness is a consequence of their ineffable insight is much weaker than the evidence that the physicists' ability to make large explosions is a consequence of their theoretical knowledge. Still, it's a hypothesis that's in the running. Unusual as they are, the mystics' mental characteristics are not so extraordinary as to defy a more pedestrian explanation than that they are privy to an ineffable secret. But it isn't unreasonable to lend some degree of initial credence to their own view of the matter—perhaps sufficient credence to justify Ricard's decision to learn the secret oneself.

There is, by my reckoning, a sixth, non-Goldmanian source of evidence available to novices in their assessment of experts. Conceivably, Goldman might have intended to include it under the "track record" rubric. But it's sufficiently different from the examples that he discusses under that heading that a separate treatment seems called for. I call it evidence of the putative expert's *epistemic modesty*. The previous fifth source of evidence is evidence of past cognitive successes—say, correct predictions. The more numerous, the more diverse, and the more surprising the correct predictions, the more justified we are in believing the next one. But now consider an individual who does *not* have a record of diverse and surprising confirmed predictions to her credit—not because she's made a lot of mistakes, but because she's kept her claims to a bare minimum. What she's had to say has by and large been correct, but it's also been pretty much restricted to what everybody already knows. So she gets no credit for a past track record of cognitive successes. However, her track record provides us with evidence that her utterances must pass an exceedingly high standard of certitude before they pass her lips. Suppose now that this person, who has never before made any but the most pedestrian of factual claims, reports that she saw a UFO. There would, it seems to me, be some

reason to believe her. The grounds are inductive, but it's a different induction from the one in case five. In case five, the inductive premise is that most of the expert's tested beliefs have turned out to be true, and the conclusion is that her untested beliefs are also true. In case six, it's also the case that most of the person's tested beliefs have turned out to be true. But in this case, the beliefs that have been tested are for the most part beliefs that we all already have ourselves. Thus, on these grounds, we have no more reason to accept her untested beliefs than anyone else's. But there is another induction to be made. In case six, the inductive premise is that most of the person's beliefs have a high degree of justification. Other people share the same beliefs, but they additionally have more speculative beliefs. This person is known to have no truck with speculative beliefs. The inductive conclusion is that her future beliefs will also have a high degree of justification.

Mystical claimants to occult power, for example, would score no credibility points by this criterion. I don't know how this additional source of evidence affects other mystics' claims, for the mystical literature has not been read with this potential source of evidence in mind. As was the case with the fifth source of evidence, all we need to find here is one good mystic. One person with an established track record of extreme epistemic modesty may be enough to rationally persuade us that there are UFOs, or that she's had an ineffable insight into the nature of the universe.

Finally, we may have grounds for accepting the mystic's claim even if we have no evidence at all bearing specifically on the mystic's credibility. It was noted in the subsection "Accepting the metaclaim: sufficient conditions," p. 85, that several writers have argued that we have a presumptive epistemic right to accept the testimony of others so long as we are not in possession of reasons for rejecting them. Thus, even if it were to be the case that none of the Goldmanian-plus-one sources of evidence for novices were to yield up any grounds for believing the mystic, the story would not yet be over. It might still be possible to argue that the mystic's report should be believed on the general ground that any reports should be believed unless there are reasons to reject them.

Are there reasons to reject the mystic's claim? One type of defeater is the presence of "distorting interests and biases"—Goldman's third source of evidence. Here we have, I think, the most potent reason for withholding assent to the canonical mystics' claims: the claims are just too good to be true. More precisely, the

claims are so good that we must be wary of the possibility that the mystics have succumbed to wishful thinking. This may be enough of a reason to withhold default acceptance and look into the specific merits of the claim. The greatest specific merit, to my mind, is not to be found in the problematic consensus of the perennialists. It lies in the extraordinary mental and emotional skills of some selected mystics. Whether this is reason enough to overcome our suspicion of wishful thinking is a matter on which reasonable non-mystics might disagree. The best advice is to keep an open mind.

4 Five types of ineffability

The idea of ineffability is open to a variety of interpretations. For one thing, expressive incapacity may come in various degrees. For example, a state of affairs may be ineffable in one particular language, or in any humanly accessible language, or in any possible language. Such degrees of incapacity are discussed in the subsection "Four or five grades of ineffability" in Chapter 1, p. 23, and will be discussed again below. But the main purpose of Chapter 4 is to distinguish five *qualitatively* different senses in which it may be claimed that something cannot be said.

A secondary purpose is to apply the new taxonomy of ineffabilities to the phenomenon of religious mysticism. The canonical mystics characteristically maintain that they have experienced an ineffable insight. The question arises: which of the five types of ineffability is being claimed? I don't plan to deal at depth with this historical question. My discussion is propaedeutic to the historical question: it's a conceptual investigation into the consequences of the several interpretations of the mystics' ineffability claim which are generated by the new taxonomy.

Unrepresentability

A state of affairs may in the first instance be ineffable in a class of languages because those languages lack the expressive resources to *represent* the state. "Snow is white" is ineffable in this sense in any language that lacks color words and conventional means for introducing them. Let's call this type of representational ineffability by the name of *unrepresentability*. Unrepresentability is undoubtedly the prototype of ineffability. It's also the type that's discussed in Chapters 1, 2, and 3 of this book. Examples of unrepresentability

claims can be found in philosophical commentaries on religious mysticism. According to Paul Henle, for instance, "religious mystics may be taken as asserting that their insights are ineffable with regard to all known symbolisms" (1949, 419). I will note below that it's at least arguable that the canonical mystics' protestations of the inexpressibility of their insights may refer to other, non-representational varieties of ineffability.

Unrepresentability comes in degrees. A fact or state of affairs may be unrepresentable in the language of discourse (*weak* unrepresentability), in all humanly accessible languages (*human* unrepresentability), in all languages accessible by any nomologically possible beings (*nomological* unrepresentability), or in all logically possible languages (*logical* unrepresentability). (This is a somewhat simplified version of the taxonomy presented in the section on p. 23.) All these grades of unrepresentability have figured in the literature of ineffability. A famous example of a weak unrepresentability claim is Benjamin Lee Whorf's (1964) thesis that some sentences of Hopi have no English translation. If Whorf is right, then the states of affairs represented by these Hopi sentences are weakly unrepresentable in English. Consider once again Henle's view that mystical insights are purported to be unrepresentable in "all known symbolisms." If Henle is right, then mystics claim that their insights are weakly unrepresentable in every human language that has ever been devised. This still falls short of *human* unrepresentability, for it leaves open the possibility that the insights can be represented in some hitherto unknown but humanly accessible symbolism.

An example of a full-blooded human unrepresentability claim is Colin McGinn's (1989) thesis that the human mind is incapable of conceiving of the theory that correctly explains the mind–brain relation. If McGinn is right, then no human being can ever master a language in which this theory can be represented—the theory is humanly unrepresentable. McGinn also speculates that the correct mind–brain theory may be inaccessible to any nomologically possible being (1989, 361), which provides us with an example of a nomological unrepresentability thesis. Finally, Galen Pletcher has opined that mystics regard their insights as logically ineffable: "Is the mystical experience coherently describable in *some* conceptual system? . . . mystics and mystic interpreters consistently answer this question in the negative" (Pletcher 1973, 207).

The mystic's expressive impasse is often attributed to the experiential deficits of non-mystics. For example, it's likened to the

inability of a sighted person to convey phenomenal facts about visual experience to a blind person: the sighted cannot tell a blind person what it means to say that an object looks red. By virtue of their experiential deficiencies, the blind are unable to master some of the expressive resources of English. The fact that certain objects look red under certain conditions is unrepresentable in the truncated English of the blind. By the same token, we non-mystics may suffer from experiential deficiencies relative to mystics that make it impossible for mystics to inform us of the content of their insight. This view of the matter is exemplified by William James' discussion:

> [The quality of a mystical state] must be directly experienced; it cannot be imparted or transferred to others. In this peculiarity mystical states are more like states of feeling than like states of the intellect. No one can make clear to another who has never had a certain feeling, in what the quality or worth of it consists. One must have musical ears to know the value of a symphony; one must have been in love one's self to understand a lover's state of mind. Lacking the heart or ear, we cannot interpret the musician or the lover justly, and are even likely to consider him to be weak-minded or absurd. The mystic finds that most of us accord to his experiences an equally incompetent treatment.
>
> (1902/1985, 380)

If the Jamesian account is correct, then, *contra* Pletcher, mystical ineffability is merely a species of weak unrepresentability. It's true that the fact that an object looks red is weakly unrepresentable in the truncated English of the blind; but the same fact is straightforwardly representable by the sentence "This object looks red" in the standard English of the sighted. If we sighted were brought up speaking the truncated English of the blind, there would be nothing to stop us from extending our native language by the introduction of terms referring to visual experiences. By the same token, there is nothing in the blind–sighted model of mysticism to stop mystics from introducing the concepts needed to represent their insights. This extension of standard English may not be accessible to us non-mystics, just as the extension of the blind's truncated English to standard English is not accessible to the blind. But in both cases, the more expressive language is still a *humanly* accessible language. Mystics, after all, are still human beings. The only difference between

the two cases is that there are more sighted people than blind, but fewer mystics than non-mystics—and even this difference of degree could change (there could be an epidemic of congenital blindness, or a massive mystical enlightenment of the bulk of the population). The bottom line is that the hyper-English in which mystical insights can be represented is a humanly accessible language. Therefore mystical insights are not *humanly* unrepresentable—they are merely *weakly* unrepresentable in English, as well as in the other languages that are fully accessible to non-mystics. If mystics do *not* believe that their insights can be represented in a jargon that's inaccessible to non-mystics—if, for example, Pletcher is right in claiming that mystics believe their insights to be unrepresentable in *any* conceptual system—then the experiential deficit model doesn't suit the mystics' claims.

Let's look more closely at the concept of weak unrepresentability. It may be objected that natural languages don't have fixed boundaries, so that the question of whether a particular state of affairs can be represented in (say) English has no determinate answer. Alfred Tarski was of this opinion:

> [Natural] language is not something finished, closed, or bounded by clear limits. It is not laid down what words can be added to this language and thus in a certain sense already belong to it potentially. We are not able to specify structurally those expressions of the language which we call sentences, still less can we distinguish among them the true ones.
>
> (1956, 164)

My first reaction to this difficulty was to distinguish between two languages: Narrow English and Broad English. A state of affairs is unrepresentable in Narrow English if it can't be represented by means of the words and syntactical devices that have already been explicitly introduced. Strictly speaking, Narrow English has to be relativized to a date—Narrow English in 2002 is a very different language from Narrow English in 1002. A state of affairs is unrepresentable in Broad English if it can't be represented by means of any words or syntactical devices that *can be* introduced into Narrow English. In effect, Broad English is the union of all the possible Narrow Englishes that can be attained from the starting point of the current (or any other) Narrow English. This distinction, I believed, enabled me to maintain that the question whether a

particular state of affairs can be represented in English admits of a determinate answer despite the open-endedness of natural languages. It's just that the answer may be different for a particular brand of Narrow English and for Broad English.

What does it mean, however, to "introduce" a new expressive device into (Narrow) English? What methods of introduction are permissible in order for us to continue to call the new language an extension of English—an explicit realization of what "already belong to it potentially"—rather than a different language? We can't insist that all new words be introducible by a verbal definition; if we did, then Broad English could have no more expressive power than the Narrow English with which we began. At the very least, we have to allow the introduction of new terms by means of demonstratives, as in "by 'fuchsia', I mean *this* color." Now the range of new concepts that can be introduced by ostension depends not only on the conventional properties of the language being extended, but also on the psychological characteristics of the speakers. For example, the very same utterances, accompanied by the very same physical pointings, could indicate the concept of green to one type of mind and grue to another type of mind. But since its effect depends on the psychology of the speakers, the use of demonstratives slides seamlessly into the use of any and all non-conventional methods of conveying new concepts: metaphor, analogy, drama, reward and punishment, etc. In sum, the methods available for broadening English are coextensive with the methods available for teaching a new language like Chinese. This means that Chinese has to be viewed as a component of Broad English: you can get there from here. In fact, you can get from any humanly accessible language to any other humanly accessible language. That is to say, Broad English is the union of all humanly accessible languages. The same can be said of Broad Chinese or Broad Urdu. It follows that weak unrepresentability in Broad English is identical to *human* unrepresentability. The moral is that it isn't interesting to talk about what can or cannot be represented in English "potentially." There is no middle ground between unrepresentability in natural languages narrowly conceived and unrepresentability in any humanly accessible language.

Unabducibility

Here's a highly idealized dynamic model of what's involved in producing speech acts: a candidate sentence *S* is selected and evaluated for its sayability under present circumstances; *S* is spoken if it passes some threshold of sayability; otherwise, *S* is rejected and either another candidate *S′* is selected for evaluation, or the speaker decides not to say anything at this time. I will use this model as a representative for a broad class of models, the differences among which do not figure in the analysis. For one thing, either the selection process or the evaluation process or both processes may be either conscious or unconscious. For another, we may select *batches* of sentences at a time for evaluation and designate the highest above-threshold scorer of the batch for saying, rather than say the first one that exceeds the threshold. More substantially, it may be objected that we don't begin by selecting a sentence for evaluation—we begin by selecting a *proposition*, decide whether it is what we want to express, and only then begin to look for a sentence that expresses the chosen proposition. I want to circumvent the deep and difficult issues concerning the relation between language and thought that come to the fore here. I believe I can do so by allowing that it may be a part of the evaluative process for sentences that they express the proposition that we have previously decided to express.

Let's have a closer look at the initial process of selecting candidates for evaluation. This is the process that Peirce (1901/1957) called *abduction*. Given the infinite number of English sentences that might be submitted to a sayability evaluation, which ones are in fact abduced? One can imagine indefinitely many abductive strategies or mechanisms (one is tempted to call them strategies if they are thought to be conscious, mechanisms if unconscious). For example, one might evaluate sentences in order of their length—first all the one-symbol sentences, then all the two-symbol sentences, and so on, until a sentence is found that surpasses the threshold of sayability. Of course, strategies (or mechanisms) may change from circumstance to circumstance. Strictly speaking, a cognizer needs to be assigned an *abductive function* that maps circumstances which might call for a speech act into the order in which candidate sentences are considered for saying under those circumstances. Most of my discussion, however, is conducted relative to some unspecified circumstance.

Now the strategy of abducing candidate sentences in order of their length has this property: every sentence of the language is

scheduled for eventual evaluation, if only its predecessors fail to be above the threshold (and if the cognizer lives long enough). This is not a universal feature of abductive strategies. For example, consider the alternative strategy of abducing candidate sentences in alphabetical order. No matter how long he, she or it lives, a cognizer who follows this strategy will never evaluate sentences that begin with the letter *b*, even if all their predecessors fail to be above the threshold. The reason, of course, is that there are infinitely many predecessors—sentences that begin with the letter *a*—that need to be evaluated first. If we abduce via the alphabetical strategy, finite information processors like ourselves will never even consider saying "Bisons are bigger than butterflies," because there will always be a candidate of the form "Ants can't count to *n*" whose consideration takes precedence. "Bisons are bigger than butterflies" is *unabducible* relative to the alphabetical strategy

What determines the nature of the abductive function? It could be society, or biology, or (more plausibly) a combination of both. There's no a priori reason to rule out the possibility that sentences are abduced entirely at random—although it's easy to rule out this possibility on empirical grounds: if we considered what to say by randomly selecting candidates from the infinitude of grammatical sentences of English, we would never encounter any sentence that's worth saying. Whatever it is that determines the shape of the abductive function, there's at least the theoretical possibility that *all the abductive functions that are nomologically available to a person or a class of persons render some sentence* S *unabducible in any circumstance*. The abductive function that selects sentences in alphabetical order illustrates this point in dramatic fashion: for a being who is nomologically compelled to evaluate sentences alphabetically, the possibility of saying "Bisons are bigger than butterflies" will never be entertained. More generally, let *L* be the set of all the sentences in the language of discourse, and let *P* be a being that always selects its candidate sentences from the set *L*-{*S*}, where *S* is a sentence of *L*. Then *S* is unabducible by *P*.

Unabducibility is a non-representational type of ineffability. If *S* is unabducible by *P*, then P *can't say it*. *P* isn't suffering from unrepresentability problems, however: *P*'s language contains a perfectly adequate representation of the state of affairs represented by *S*, namely *S*. Nevertheless, it's nomologically certain that *P* will never say *S*. It might be objected that there is no coherent distinction between unabducibility and unrepresentability—that if *P* is

incapable of entertaining *S*, then *P* effectively speaks the language *L*-{*S*}, in which the state of affairs represented by *S* is unrepresentable. (I assume here that *L* doesn't contain any other sentences that have the same meaning as *S*. This presupposition will be discharged in the next two paragraphs.) The reply to this objection is the same as the reply to any verificationist argument. We have two hypotheses that account for *P*'s inability to say *S*: (1) *P* speaks the language *L*-{*S*}, in which *S* is unrepresentable, and (2) *P* speaks the language *L*, but operates with an abductive function that never selects *S*. The fact that these two hypotheses are empirically indistinguishable when taken in isolation doesn't mean that they're indistinguishable *tout court*. It's possible that each of them might be embedded in a broader theory, and that one of these theories receives more empirical support than the other.

Like unrepresentability, unabducibility also comes in degrees. In the case of unabducibility, however, there are two theoretically interesting dimensions of quantitative difference. The first dimension is the extensiveness of the class of *languages* in which a sentence is unabducible; the second is the extensiveness of the class of *persons* for whom a sentence is unabducible. Let's call these the *l-grades* and the *p-grades* of unabducibility, respectively. The l-grades of unabducibility correspond to the grades of unrepresentability discussed in the previous section: a sentence may be l-weakly unabducible in a single language, l-humanly unabducible in all humanly accessible languages, l-nomologically unabducible in all nomologically accessible languages, or l-logically unabducible in all logically possible languages. This formulation is, of course, not exactly right. Since it's *sentences* that are unabducible, the concept of unabducibility is defined only with respect to a given language. No sense attaches (yet) to unabducibility in a *class* of languages. So what do we mean when we say that *S* is l-humanly unabducible? I define the unabducibility in language *L2* of a sentence *S1* of language *L1* as follows:

> *S1* is *unabducible in* L2 if and only if there is no sentence S2 of *L2* which both is abducible and has the same truth-conditions as *S1*.

One adventitious consequence of this definition is that *S1* is unabducible in *L2* if there is *no* sentence *S2* of *L2* which has the same truth-conditions as *S1*—i.e., if the truth-conditions of *S1* are *unrepresentable* in *L2*. But unrepresentability in a class of languages

can't be said to entail unabducibility in that class. The way things have been set up, unabducibility in a class of languages requires that there be at least one language in the class that contains the unabducible sentence.

Another adventitious consequence of the definition of unabducibility is that *S1* is unabducible *in* L1, the language of which *S1* is a sentence, just in case there is no sentence *S1′ of* L1 which is both abducible and has the same truth-conditions as *S1*. Heretofore, when I spoke of the unabducibility of a sentence in its own language, I merely meant that this very sentence could not be abduced. But it's a consequence of the generalized definition of unabducibility in an arbitrary language that the unabducibility of *S* in its own language *L* entails that *all* the sentences of *L* that have the same truth-conditions as *S* cannot be abduced. I have not found an elegant way out of this lexical tangle. But I affirm that the ambiguity of "*S* is unabducible in *L*" does no harm to the subsequent analysis.

To say that *S* is unabducible in *L* (in either sense of the phrase) is not yet to make a definite claim. *S* may be unabducible in *L* by a single person (p-weak unabducibility), or by all of humanity (p-human unabducibility), or by all nomologically possible beings (p-nomological unabducibility), or by all logically possible beings (p-logical unabducibility). The cross-product of the two dimensions of unabducibility—linguistic and personal—yields 16 cases. Here are a few illustrative comments on some of them. The combination of *l-weak* and *p-weak* unabducibility denotes the case where an individual can't abduce *S* in a language *L*. *L-weak* and *p-human* unabducibility is the case where no human being is able abduce *S* in *L*. Presumably, the sentence *S* has properties that put it beyond the reach of any humanly accessible abductive function. This would be the case if humans were nomologically constrained to abduce alphabetically and *S* were a sentence that starts with the letter b. In *l-human* and *p-weak* unabducibility, an individual can't abduce *S* in any humanly accessible language. If the individual were to learn any other language *L2*, he would still be unable to abduce any sentence of *L2* that had the same truth conditions as *S*. In *p-human* and *l-human* unabducibility, no human being can abduce *S* in any humanly accessible language. In this case, no human being will ever come up with a sentence in any language that represents the truth-conditions of *S*.

Another point about these grades of unabducibility: it's clear that there are some logically possible beings who can abduce anything

that there's a sentence for in any logically possible language. That is to say, unabducibility is always a contingent matter. So the three p-logical unabducibility cases are demonstrably empty.

Finally, it's worth noting that some of the 16 cross-combinations collapse into others. For example, consider l-nomological and p-human unabducibility. This is the case where no human being can abduce *S* in any language that's accessible to any nomologically possible being. This condition can be decomposed into the following pair of subconditions: (1) no human being can abduce *S* in any humanly accessible language, and (2) no human being can abduce *S* in any language that's nomologically accessible but not humanly accessible. But the second subcondition is vacuously true, since no human can have access to humanly inaccessible languages. Therefore l-nomological and p-human unabducibility reduces to l-human and p-human unabducibility.

Here is an argument (that I will ultimately repudiate) to the effect that unabducibility is *not* a type of ineffability. Consider the case of p-weak unabducibility—a person P can't abduce *S* in some language or class of languages. In this case, P's inability to say *S* is due to the fact that the option of saying *S* never arises in P's mind. But if that's the *only* obstacle to saying *S*, then it might happen that the option of saying *S* is presented to P by *another* individual, Q, for whom *S* is not unabducible. Q might walk up to P and ask "Would you say that *S*?" So even if it's assumed that *S* is unabducible by P, *S* may nevertheless arrive in P's mind via the external route. (Note that no such palliative is available if P's inability to say something is due to its *unrepresentability*: if *S* isn't a part of P's language, then when Q asks him whether he would say that *S*, P simply wouldn't understand him.) It seems to follow that p-weak unabducibility isn't a form of ineffability. The fact that P can't abduce *S* by himself is certainly an obstacle to his saying *S*. But with a little bit of prompting, P may end up saying *S* anyway. What if all human beings nomologically suffer from the same unabducibilities? Then one human being Q can't provide another human being P with the sentence that P can't abduce. But that role could conceivably be played by extra-terrestrials. This is the main potential benefit of the search for extraterrestrial intelligence: when we finally contact them, they may supply us with hypotheses that we can understand, but that we would never have thought of on our own. More to the point, p-human unabducibility isn't guaranteed to be a type of ineffability either.

Finally, what about p-nomological unabducibility? (I say "finally" because there is no such thing as p-logical unabducibility.) In this case, we can be sure that no being is ever going to supply any other being with the unabducible sentence. But (it may be argued) p-nomological unabducibility is still inadequate to ensure ineffability. It's true that no language-using agent will ever suggest *S* to P. But hypotheses that come in via the external route don't have to be delivered by language-using agents. *S* might accidentally be typed out by one of the proverbial typing monkeys; the sound of the whistling wind or the babbling brook may acoustically be indistinguishable from someone's saying *S*. The effect of such eventualities on P, if he should perceive them, can clearly be the same as if he *had* heard a linguistic agent saying *S*. In fact, P might erroneously believe that he had heard *S* being said. Whether P is right or wrong in this supposition, the event can bring *S* to P's mind, whereupon P might choose to assert it. Thus *S* may end up being said even if it's p-nomologically unabducible. The conclusion is that unabducibility is not a form of ineffability. It makes effing difficult and chancy; but it's still nomologically possible, a fortiori logically possible.

The weakness of the foregoing argument is that it relies on an entirely unprincipled distinction between the internal and the external route to P's mind. Recall that *S* is unabducible if it fails to be in the range of all nomologically possible abductive functions, and that an abductive function is a function that maps circumstances into ordered sets of sentences. Well, hearing someone ask you "Would you say that *S*?" is a circumstance; so is hearing that question in the whistling of the wind. The fact that either of these circumstances would cause a normal English-speaking adult to entertain the possibility of saying *S* just means that the abductive function of normal English-speaking adults assigns a set containing *S* to these circumstances. There is no interesting difference between being impelled to consider saying "Snow is white" by the sight of snow or by a friend's inquiry into our opinion of the color of snow. Sentences that come into our heads via the so-called external route aren't instances of effable unabducibilities—they're perfectly ordinary cases of abduction.

On the other hand, there are at least logically possible beings for whom *S* is unabducible by either the internal or the external route. There is no circumstance that can impel such a being to entertain *S*. Even shouting "Would you say that *S*?" in their face (if they have a

face) or sending them e-mails with the same query will not have the desired effect. Perhaps this comes about because these beings systematically fail to perceive the question. Freudian repression is supposed to work more or less like this. At any rate, unabducibility is at least a conceptual possible variety of ineffability.

Unselectability and unexecutability

According to the idealized model constructed in the previous section, what a person says is the result of a three-stage process: candidate sentences are *abduced*, the abductees are *evaluated*, and the winner is *executed* (i.e., the winning sentence is said). In the previous section it was seen that a sentence might be ineffable by virtue of never being abduced. A sentence might also fail to be expressible by virtue of blockages at either the evaluative or the executive stage. Suppose that sentence *S* is sometimes abduced. *S* would nevertheless be ineffable if it were always evaluated as being below the threshold of assertibility. I will say that a sentence having this property is *unselectable*. We may, under certain circumstances, come to entertain the possibility of saying an unselectable sentence; but we always decide against it in the end. It always seems to be too contentious, or too troublesome, or too trivial a thing to say.

Consider now a sentence that's both abducible and selectable —we entertain the possibility of saying *S* and decide that we *will* say it. These conditions are not yet sufficient to insure that *S* will be said, for we may find that we're unable to execute our intention. When we try to say *S*, we find ourselves tongue-tied, or our head explodes. In such a case, I will say that *S* is *unexecutable*. Unabducibility, unselectability, and unexecutability are all species of *unspeakability*. To say that *S* is unspeakable is to say that there is a nomologically insurmountable obstacle to saying *S*. This obstacle may be located at the abductive, evaluative, or executive stage of speech production.

The formulas "*S* is unselectable in *L*," "*S* is unexecutable in *L*," and "*S* is unspeakable in *L*" suffer in my hands from the same benign ambiguity as "*S* is unabducible in *L*": sometimes they mean only that the sentence *S* itself cannot be selected or executed or said, and sometimes they mean that no sentence having the same truth-conditions as *S* can be selected, etc. In some places in my discussion, the context resolves the ambiguity; in the other places, it's not important that the ambiguity be resolved.

Unselectability, unexecutability, and unspeakability each come in the same 16 varieties as unabducibility. Some selected examples: l-weak and p-weak unspeakability refers to the inability of a single individual to speak a sentence *S* in a single language; l-human and p-human unspeakability is the inability of any human being to speak *S* or any sentence with the same truth conditions as *S* in any humanly accessible language. As in the case of unabducibility, there are some logically possible beings who can say anything that there's a sentence for in any logically possible language. Therefore there can be no case of p-logical unspeakability. The same can be said a fortiori about unselectability and unexecutability. As with unabducibility, some of the 16 cases reduce to others. For example, l-nomological and p-human unspeakability is the same as l-human and p-human unspeakability.

The concept of unspeakability suggests an alternative interpretation of the mystics' ineffability claims. The standard interpretation is that mystical ineffability is due to the *unrepresentability* of the mystic's insight. But might it not be a matter of unspeakability? Perhaps the mystical experience is one which renders the experiencer incapable of stating its content. On this account, there might be a perfectly adequate statement of the mystical insight, but having the insight causes one to be afflicted by a condition wherein one is rendered incapable of making that statement. The same general idea might explain the curious fact that no work has ever been done on the psychology of silliness. Perhaps the contemplation of silliness disrupts one's intellectual processes in a way that results in unspeakability: every time you try to formulate what silliness is, you dissolve in giggles. By the same token, it may be the case that every time the mystic tries to formulate what the mystical insight consists in, she dissolves in rapture. Something very much like this interpretation of mystical ineffability has been suggested by James Austin:

> Some patients who sustain a relatively small stroke in the left thalamus have language problems. They can't retrieve words, nor can they register and retain verbal material. Therefore, it is difficult to exclude the possibility that in some ineffable states, those parts of the thalamus usually involved in language are preempted, or are disarticulated from their usual routines, or bypassed. If such were the case, on one or both sides, then some portion of [mystical] ineffability might imply

> that the corresponding parts of the person's deeper language functions had been out of the loop in a sense, and were then far removed from the rest of the mainstream flow of the experience.
> (Austin, 1998, 515–516)

My guess is that Austin has in mind the unabducibility variety of unspeakability: the mystical experience is thought to cause or consist in a disengagement of the brain functions responsible for finding a sentence that represents the insight. I will have occasion to discuss an unselectability-based interpretation of mystical ineffability in the next section.

Suppose, as per Austin, that it's having the mystical insight that renders one incapable of stating the content of the insight. Then it might still be possible for a *non*-mystic—someone who doesn't have the insight—to state the content of the mystic's insight. Perhaps the non-mystic believes on the basis of indirect evidence that the mystic has twigged onto a cosmic truth, and guesses that this is what the mystical insight comes to. And then mystical ineffability would turn out to be a case of merely p-weak unspeakability—the mystical insight would be ineffable only to the mystics! Alternatively, it may be that non-mystics are incapable of representing the mystical insight, because having a mystical experience is a prerequisite for attaining some of the concepts that are essential to its representation. This case would be a hybrid of weak unrepresentability and p-weak unspeakability: the mystical insight would be unrepresentable in the languages that are accessible to non-mystics, and it would be unspeakable by mystics. Interestingly, the combination of these two weak ineffabilities would result in an across-the-board *human* ineffability: *no* human beings could state the mystical insight, either because they can't master the requisite language or because their mental state renders them incapable of stating it.

Is it conceptually possible that the mystical insight is straightforwardly p-humanly unspeakable?—that no human being, whether mystic or non-mystic, is capable of stating its content, even though that content is representable by sentence *S*? For that to be the case, it wouldn't be enough that the mystical insight itself renders one speechless. We would have to be so constructed that merely *entertaining the hypothesis* represented by *S* renders one speechless. One is reminded of the Monty Python skit in which the military develops jokes that kill. This may be an implausible scenario, but it is a conceptual possibility.

Unreportability

Consider the following variant of Moore's paradox:

(1) Snow is white, but I do not say "Snow is white."

Sentence (1) has this peculiarity—that though it may be true, I can never be in a position to *report* the fact to anyone, where "reporting *S*" is taken to imply that *S* is true. I will say that (1) is *unreportable*: if I try to report *S*, I thereby ensure that it is false; but if I *don't* try to report it, it may very well be true. The occurrence of indexicals in (1) is incidental. The following non-indexical sentence is just as unreportable as (1):

(2) Snow is white, but no one has ever said or will ever say "Snow is white."

There are surely some sentences having the form of (2) that are true. In fact, there are infinitely many of them. For example, there are infinitely many numbers x, y, and z, such that x+y = z and no one has ever said or will ever say that x+y = z.

Unreportability is not the same thing as unrepresentability: the state of affairs represented by (1) is obviously representable—by (1). Unreportability is also not the same thing as unspeakability. There is nothing to stop me from saying (1). It's just that if I *do* say (1), I can be sure that my report will be false. Nevertheless, unreportability is a type of ineffability—it's another way of construing the claim that something "can't be said." In this case, what can't be said is the fact that *S* is true and that I do not say "*S* is true." There are infinitely many facts of that form, but we can be sure that none of them will ever be stated.

Here's another way to describe the differences among the three types of ineffability. Suppose you try to write a universal encyclopedia that contains all the facts. Ineffabilities are facts that can't, for either nomological or logical reasons, be a part of the encyclopedia. Unrepresentable facts can't find their way into the encyclopedia because there's no sentence for them. Unspeakable facts can't get in because they can't be written down. Unreportable facts can't get in because writing them down makes them false.

Here's another *apparent* difference between unspeakabilities and unreportabilities. Unspeakabilities are always contingent. The

attempt to speak the sentence makes our head explode, or puts us to sleep; but there's no *logical* obstacle to speaking any sentence in a language that we've mastered. Only empirical research can reveal whether there are unspeakabilities. In contrast, empirical research is irrelevant to ascertaining the unreportability of (1) or (2). The fact that a sentence is unreportable is a logical or conventional fact that can be established on the basis of a priori reasoning. This imputation is based on a mistake. It's true that unspeakability is always contingent. It's also true that the unreportability of (1) and (2) can be established on a priori grounds. But there can also be contingent unreportabilities. The act of saying *S* is a spatiotemporal event that has a place in the causal nexus of the universe. Speech acts have even been known to cause people to kill themselves. Since saying *S* is a bona fide causal event, there's no a priori reason why its effect can't be that the truth-conditions of *S* do *not* obtain. For example, telling an athlete that he's sure to win the competition may cause him to become complacent and therefore to lose. By parity of reasoning, saying the *negation* of *S* may cause the truth-conditions of *S* to *be* satisfied: if the athlete is told that he will lose, his wounded pride spurs him on to victory. If both these situations were simultaneously to obtain, we would be wrong whether we asserted *S* or its negation. The only way to avoid error would be to keep one's mouth shut. That is to say, both *S* and its negation would be unreportable. But their unreportability would be due to certain contingent causal facts that could only be discovered by empirical research.

Once again, "*S* is unreportable in *L*" may be read either as "One cannot report the sentence *S* of *L*" or as "One cannot report any sentence of *L* that has the same truth-conditions as *S*." Unreportability also comes in the same 16 grades as unspeakability. An example of a p-weak and l-weak unreportability is:

(3) Snow is white, and George Bush doesn't say "Snow is white."

Sentence (3) may be true, but George Bush could never report it in English. However, Colin Powell could report it—therein lies its p-weakness. Also, if George Bush could learn German, he could apprise us of the same fact himself by saying:

(4) Schnee ist weiss, und George Bush sagt nicht "Snow is white."

Therein lies (3)'s l-weakness. Here is a p-human and l-human unreportability:

> (5) Snow is white, and no human being speaks any sentence that has the same truth-conditions as "Snow is white" in any humanly accessible language.

If (5) were true, Colin Powell couldn't report it any more than George Bush could—and learning German wouldn't help either one of them.

A noteworthy difference among the three varieties of ineffability concerns the status of their strongest, "logical" grade. The existence of logical unrepresentabilities—states of affairs that can't be represented in any logically possible language—is a speculative hypothesis at best. At worst, logical unrepresentability may turn out to be an incoherent notion. Logical unspeakability, on the other hand, clearly *is* an incoherent notion: if a sentence *S* is unspeakable by some person or persons, then it's representable by *S*, and then there surely is a logically possible being who can say *S*. On the third hand, it's just as clear that logical unreportabilities *do* exist. Sentence (2) above is an example of a p-logical and l-weak unreportability: there is no logically possible being who could report the truth of (2) in English. Here is a p-logical and l-logical unreportability—a sentence that could be true, but the truth of which cannot be conveyed by any logically possible being in any logically possible language:

> (6) Snow is white and nobody speaks any sentence in any language that has the same truth-conditions as "Snow is white."

Is it possible that the canonical mystics' protestations of ineffability refer to unreportability? A nice feature of this interpretative hypothesis is that it has the potential of underwriting the frequently encountered claim that the mystical insight is ineffable in the highest, "logical," degree. As noted above, this claim is problematic for unrepresentability and demonstrably false for unspeakability. Of course, sentences having the form of (1) or (2) aren't plausible candidates for the mystical insight. For any numbers x, y, and z such that x+y = z, we may have the unspoken insight that x+y = z and that nobody speaks that x+y = z. But this insight is not likely to make

us feel transformed or ecstatic, or make us believe that we've solved the riddle of the universe.

There are other unreportabilities, however, that are not so absurdly unrelated to mystical themes. Consider the virtue of humility, i.e., the virtue of not ascribing virtues to oneself. There are, or at least there can be, humble people. But nobody can truly claim to *be* humble. The humility of the humble is unreportable. Of course, I can *ask* myself whether I'm humble without its being false. It's just that the only correct answer I can give to the question is "no." Consider now the closely related but even more heroic virtue of *self-forgetfulness*—the virtue of never even thinking about yourself, never asking yourself or anyone else whether you have or lack a particular property. The merely humble may wonder whether they are humble, but the self-forgetful don't even ask. Like humility, self-forgetfulness is unreportable. Interestingly, it's also a concept that frequently surfaces in the mystical literature. Here are a few examples taken from Aldous Huxley's mystical anthology:

> Your own self is your own Cain that murders your own Abel. For every action and motion of self has the spirit of Anti-Christ and murders the divine life within you.
>
> (William Law, quoted in Huxley 1944, 178)

> Lucifer, when he stood in his natural nobility, as God had created him, was a pure noble creature. But when he kept to self, when he possessed himself and his natural nobility as a property, he fell and became, instead of an angel, a devil. So it is with man. If he remains in himself and possesses himself of his natural nobility as a property, he falls and becomes, instead of a man, a devil.
>
> (*The Following of Christ*, quoted in Huxley 1944, 180)

> There are few contemplatives, because few souls are perfectly humble.
>
> (*The Imitation of Christ*, quoted in Huxley 1944, 297)

And here is the clincher from Zen master Dogen:

> To study Buddhism is to study the self.
> To study the self is to forget the self.
>
> (Dogen, quoted in Epstein 1995, 20)

Let's explore the hypothesis that the mystical state is the state of self-forgetfulness. This hypothesis accounts at least in a general way for the mystics' protestations of ineffability: when they're asked "What state are you in?" they can't give a true reply. If they're asked "Are you in a mystical state?" the most accurate response they can make is to say "Not any more!" This response points to a new wrinkle in the analysis. Presumably no one is perpetually in a mystical state. The hypothesis under consideration is then that P is in a mystical state at time *t* if and only if P is self-forgetful at *t*. Now P's state is unreportable *at time* t: if P says, at time *t*, that she's currently being self-forgetful, then she isn't. But there's no reason why P couldn't inform others *at another time* t′ of her self-forgetfulness at *t*. So P's self-forgetfulness isn't even p-weakly unreportable: anyone, including P herself, can inform others that P is self-forgetful at *t*. P's self-forgetfulness at *t* is merely unreportable *at certain times*.

We could, if we wished, bring this notion of temporary unreportability into play in the analysis. In fact, we could introduce a temporally indexed notion of any of the varieties of ineffability. An example of temporary unrepresentability is provided by the language which is just like English, except that by linguistic convention, color words have no meaning on Thursdays. In this language, the fact that snow is white is unrepresentable on Thursdays. As for temporary unspeakability: there might be creatures whose heads explode if they try to say *S* on Thursdays. These intermittent ineffabilities are members of an indeterminately large family. There are conceivable languages and speakers relative to which a sentence is effable or ineffable, in any of the five senses of the terms, depending on the geographical location of the speech act, the weather, the gender of the speaker, etc. These ineffabilities generally cross-cut the types that I've formally introduced (weak, human, nomological, and logical). For instance, unspeakability-on-Thursdays neither contains nor is contained by weak unspeakability, e.g., unspeakability by George Bush. However, the temporary unreportability of "P is self-forgetful at time *t*" is weaker than the p-weak unreportability of "P is always self-forgetful." If the self-forgetfulness theory of mysticism is right, then the mystical state isn't even weakly unreportable. To be sure, there is a sentence about self-forgetfulness that *is* p-weakly unreportable, namely "P is always self-forgetful." There's even a sentence about self-forgetfulness that's p-logically and l-logically unreportable: "Everyone is always self-forgetful." But these

sentences suffer from the shortcoming of being patently false, even if no one tries to use them.

Richard Gale (1960) makes essentially the same point in his critique of Walter Stace's (1952) views about mysticism. Like many other writers on mystical themes, Stace holds that the mystical experience involves "the dissolution of the personal ego." If this is not exactly the same thing as self-forgetfulness, at least it *entails* self-forgetfulness. Gale protests that if this is a correct description of the mystical state, then, by virtue of having been correctly described, the mystical state can't be considered ineffable after all. Gale supposes that "Stace's counter to this would be that *within* the mystical experience the mystic cannot conceptualize his experience" (472). But, Gale notes, "this sense of ineffability is trivial because many non-mystical experiences would equally qualify as candidates for the title of ineffable" (472). Here are his examples:

> We would all agree that the experience of wrestling with an alligator is conceptualizable. However, the proposition, "Tarzan is wrestling with an alligator," could not possibly be either formulated, consciously considered, or verified by Tarzan while he is actually having the experience described by the proposition. Similarly, and for slightly different practical reasons, Schnabel is not capable of formulating or verifying the proposition, "Schnabel is concentrating *solely* on interpreting Beethoven's 14th Sonata," while he is actually engaged in performing the sonata; for if he were to attempt to verify this proposition while he was performing the Sonata he would automatically render it false, because that would mean that he could not possibly be concentrating *solely* on interpreting the Sonata.
>
> (472)

In the present context, there are two features worthy of note in Gale's analysis. First, he attributes to Stace a theory of mysticism which equates mystical ineffability with what I have been calling unreportability. The reason that the mystic can't say what her experience is like isn't because there is no adequate description of it—the experience can be accurately described as "the dissolution of the ego." Its ineffability comes rather from the fact that describing it, like Schnabel's informing us that he is concentrating solely on playing the Beethoven sonata, automatically renders the

communication false. Second, the unreportability of the mystical state in this account is of a very feeble variety. It's weaker than p-weak. Just as Schnabel could tell us after the concert that he *had been* concentrating solely on the Beethoven sonata, so could the mystic tell us retrospectively that she had experienced the dissolution of her ego.

Gale offers his disquisition on alligators and Beethoven sonatas as a *criticism* of Stace's view: the sense in which mystical experience is ineffable is "trivial" because many mundane experiences, such as wrestling with an alligator, are ineffable in the same sense. It seems to me that Gale's analysis is rather a *clarification* of Stace's view. The fact that the intermittent ineffability of the mystical experience is shared by mundane experiences can't be regarded as a criticism, for there are mundane cases of every degree of ineffability. Even the strongest variety—permanent ineffability by any logically possible being in any logically possible language—is instantiated by infinitely many mundane facts having the form "x+y = z, but nobody ever says that x+y = z." (Of course, I can't tell you which substitutions for x, y, and z render this sentence-form true—they're ineffable. But I can be sure that there *are* substitutions for x, y, and z that render the sentence-form true.)

If the mystical experience is merely intermittently ineffable, however, why do so many mystics misleadingly claim that it's ineffable *tout court*? Perhaps it's because they don't *want* retrospectively to describe their past mystical states. Their protestations of ineffability may be part of a strategy designed to avoid losing the egoless state, or perhaps to induce it in others. Several commentators have suggested that the mystics' communications are not motivated by the desire to convey factual information (e.g., Gale 1960; Streng 1978). In this respect, mystical discourse is like a comedy routine. Suppose the mystic believes that describing the mystical state is incompatible with *entering into* the mystical state. (Compare: explaining the joke is incompatible with experiencing the humor of it.) This would be the case if the mystical state were to be a state of self-forgetfulness or ego-dissolution. But the point I'm about to make applies to any conception of the mystical state that equates it with a temporary unreportability. The point is that if the mystic wants to foster the experience in others, it will not help her to provide the other with an accurate description. On the contrary, it's obviously self-defeating to tell someone that to attain the goal he should be self-forgetful. Self-forgetfulness must always elude those

who aim for it, at least while they are busy aiming for it. In fact, the best that the mystic can do to foster the requisite mental state in others is to *discourage* their attempts to formulate the state. Perhaps the mystics' protestations of ineffability are part of a strategy to short-circuit the aspirant's self-defeating attempts to formulate the nature of the goal. Moreover, if description is incompatible with the mystical state and if the mystical state is desirable, then the mystic isn't going to want to describe it even to other mystics or to herself. The result would be that the mystic *never* formulates the nature of her mystical state at *t*. But this still wouldn't be a case of (weak or higher) unreportability. It would be rather a decision *not* to report a state of affairs that *could* be imparted to others. The mystic *can* inform us at time *t′* of the nature of her mystical state at time *t*, but chooses not to. On this account, the mystic's protestations of ineffability are a therapeutic mystification.

There may be another way of looking at the mystic's decision not to report. What if this decision is dictated by the psychological laws of human motivation? Perhaps the consequences of abiding in the mystical state are so attractive that all who have enjoyed a taste of it are compelled to try to prolong it by not describing it. If that were to be so, then, as a matter of nomological necessity, no mystic would ever state wherein the mystical experience consists. If the mystic *did* say "I was self-forgetful at time *t*," she would be speaking the truth. But her desire to foster and prolong the state causes her *not* to say it. On this account, mystical ineffability isn't a species of unreportability at all—it's a type of *unspeakability*. More specifically, it's a *p-weak* unspeakability: mystics can never speak the sentence that describes the nature of their mystical state, but nonmystics may figure it out, and they may have no compunction about stating the result of their investigation. Which variety of unspeakability characterizes the mystic's state? Well, if her motivation to prolong the mystical state causes her never even to entertain the sentence that describes her state, then that sentence is unabducible. If it comes to mind, but motivation to foster the mystical state in others causes her always to decide not to say it, then the sentence is unselectable.

I'm not sure whether it makes sense to regard the mystic's purported therapeutic mystification as nomologically compelled. But whether it's regarded as caused or freely chosen, it surely is not an instance of any of the designated varieties of unreportability. This doesn't invalidate the mystification hypothesis. It's possible that

that's what the canonical mystics' protestations of ineffability come to. But then canonical mysticism isn't a matter of (weak or higher) unreportability. In the end, I haven't been able to come up with a plausible unreportabilistic interpretation of canonical mysticism. But the phenomenon of unreportability points to a conceptually possible type of mysticism. It may not be St. Theresa's type, or the Zen masters', but maybe it's mysticism on Mars.

References

Alston, W. P. Ineffability. In J. Donnelly (ed.), *Logical Analysis and Contemporary Theism* (New York: Fordham University Press, 1972).

Alston, W. P. *Perceiving God: The Epistemology of Religious Experience* (Ithaca, NY: Cornell University Press, 1991).

Augustine, St. *City of God* (New York: Dutton, 1947).

Austin, J. H. *Zen and the Brain: Toward an Understanding of Meditation and Consciousness* (Cambridge, MA: MIT Press, 1998).

Bambrough, R. Intuition and the inexpressible. In Katz, S. (ed.), *Mysticism and Philosophical Analysis* (New York: Oxford University Press, 1978).

Bucke, R. M. *Cosmic Consciousness: A Study in the Evolution of the Human Mind* (Philadelphia: Innes, 1901).

Chomsky, N. *Reflections on Language* (New York: Pantheon Books, 1975).

Chomsky, N. Rules and representations. *Behavioral and Brain Sciences*, 3, 1980, 1–15.

Chomsky, N. *Knowledge of Language: Its Nature, Origin, and Use* (New York: Praeger 1986).

Chomsky, N., and Fodor, J. The inductivist fallacy. In M. Piattelli-Palmarini (ed.), *Language and Learning: The Debate Between Jean Piaget and Noam Chomsky* (Cambridge, MA: Harvard University Press, 1980).

Cole, D. I don't think so: Pinker on the mentalese monopoly. *Philosophical Psychology*, 12, 1999, 283–295.

Crane, T. The nonconceptual content of experience. In T. Crane (ed.), *The Contents of Experience: Essays on Perception* (Cambridge: Cambridge University Press, 1992).

Cushing, J. T. Quantum theory and explanatory discourse: Endgame for understanding? *Philosophy of Science*, 58, 1991, 337–358.

Davidson, D. On the very idea of a conceptual scheme. *Proceedings and Addresses of the American Philosophical Association*, 47, 1974, 5–20.

Davidson, D. Psychology as philosophy: Comments and replies. In *Essays on Actions and Events* (Oxford: Clarendon Press, 1980).

Davidson, D. A nice derangement of epitaphs. In E. LePore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson* (Oxford: Blackwell, 1986).

Davidson, D. The structure and content of truth. *Journal of Philosophy*, 87, 1990, 279–328.

Dennett, D. C. *Consciousness Explained* (Boston: Little, Brown and Company, 1991).

Epstein, M. *Thoughts Without a Thinker: Psychotherapy from a Buddhist Perspective* (New York: Basic Books, 1995).

Evans, D. Can philosophers limit what mystics can do? A critique of Steven Katz. *Religious Studies*, 25, 1989, 53–60.

Feigl, H. The scientific outlook: Naturalism and humanism. In H. Feigl and M. Brodbeck (eds.), *Readings in the Philosophy of Science* (New York: Appleton-Century-Crofts, 1953).

Flanagan, O. *The Science of the Mind* (Cambridge, MA: MIT Press, 1991).

Fodor, J. *The Language of Thought* (Cambridge, MA: MIT Press, 1975).

Fodor, J. The present status of the innateness controversy. In *Representations* (Cambridge, MA: MIT Press, 1981).

Fodor, J. *The Modularity of Mind* (Cambridge, MA: MIT Press, 1983).

Forman, R. K. C. Of deserts and doors: Methodology of the study of mysticism. *Sophia*, 32, 1993, 31–44.

Gale, R. M. Mysticism and philosophy. *Journal of Philosophy*, 57, 1960, 471–481.

Gimello, R. Mysticism and meditation. In S. T. Katz (ed.), *Mysticism and Philosophical Analysis* (New York: Oxford, 1978).

Goldman, A. I. Science, publicity, and consciousness. *Philosophy of Science*, 64, 1997, 525–545.

Goldman, A. I. Experts: Which ones should you trust? *Philosophy and Phenomenological Research*, 63, 2001, 85–110.

Goodman, N. *Fact, Fiction and Forecast* (Cambridge, MA: Harvard University Press, 1954).

Hacking, I. The participant irrealist at large in the laboratory. *British Journal for the Philosophy of Science*, 39, 1988, 277–294.

Happold, F. C. *Mysticism: A Study and an Anthology* (London: Penguin Books, 1963).

Hardwig, J. Epistemic dependence. *Journal of Philosophy*, 82, 1985, 335–349.

Hempel, C. *Fundamentals of Concept Formation in Empirical Science* (Chicago: University of Chicago Press, 1952).

Henle, P. Mysticism and semantics. *Philosophy and Phenomenological Research*, 9, 1949, 416–422.

Hick, J. Ineffability. *Religious Studies*, 36, 2000, 35–48.

Huxley, A. *The Perennial Philosophy* (New York: Harper, 1944).
Huxley, A. *The Doors of Perception and Heaven and Hell* (London: Chatto and Windus, 1968).
James, W. *The Varieties of Religious Experience* (London: Penguin Books, 1902/1985).
Kaplan, D. *Themes from Kaplan* (New York: Oxford University Press, 1989).
Katz, S. T. Language, epistemology, and mysticism. In S. T. Katz (ed.), *Mysticism and Philosophical Analysis* (New York: Oxford University Press, 1978).
Kellenberger, J. The ineffabilities of mysticism. *American Philosophical Quarterly*, 16, 1979, 307–315.
Kraut, R. The third dogma. In E. LePore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson* (Oxford: Blackwell, 1986).
Kuhn, T. S. *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press, 1962).
Kukla, A. *Studies in Scientific Realism* (New York: Oxford University Press, 1998).
Kukla, A. *Social Constructivism and the Philosophy of Science* (London and New York: Routledge, 2000).
Kukla, A. *Methods of Theoretical Psychology* (Cambridge, MA: MIT Press, 2001).
Latour, B. and Woolgar, S. *Laboratory Life: The Social Construction of Scientific Facts* (London: Sage, 1979).
Manning, R. N. All facts great and small. *Protosociology*, 11, 1998, 18–40.
Margolis, J. *The Truth About Relativism* (Oxford: Blackwell, 1991).
Maxwell, G. The ontological status of theoretical entities. In H. Feigl and G. Maxwell (eds.), *Scientific Explanation, Space and Time* (Minneapolis: University of Minnesota Press, 1962).
McGinn, C. Can we solve the mind–body problem? *Mind*, 98, 1989, 349–366.
Montague, R. *Formal Philosophy: Selected Papers* (New Haven: Yale University Press, 1974).
Moore, A. W. *The Infinite* (London: Routledge, 1990).
Moore, A. W. Human finitude, ineffability, idealism, contingency. *Nous*, 26, 1992, 427–446.
Nietzsche, F. On truth and lies in a nonmoral sense. In D. Breazeale (ed.), *Philosophy and Truth: Selections from Nietzsche's Notebooks of the Early 1870s* (Atlantic Highlands, NJ: Humanities Press, 1979).
Otto, R. *Mysticism East and West* (New York: Macmillan, 1932).
Peacocke, C. *Transcendental Arguments in the Theory of Content* (Oxford: Oxford University Press, 1989).

Peacocke, C. Scenarios, concepts and perception. In T. Crane (ed.), *The Contents of Experience: Essays on Perception* (Cambridge: Cambridge University Press, 1992).

Peirce, C. S. The logic of abduction. In V. Tomas (Ed.), *Essays in the Philosophy of Science: Charles S. Peirce* (New York: Bobbs-Merrill, 1901/1957).

Penner, H. The mystical illusion. In S. T. Katz (ed.), *Mysticism and Religious Traditions* (New York: Oxford University Press, 1983).

Pletcher, G. K. Mysticism, contradiction, and ineffability. *American Philosophical Quarterly*, 10, 1973, 201–211.

Popper, K. *The Logic of Scientific Discovery* (London: Hutchinson, 1959).

Proudfoot, W. *Religious Experience* (Berkeley: University of California Press, 1986).

Przelecki, M. On the meaning of indexicals. *Studia Logica*, 42, 1983, 285–292.

Putnam, H. *Mathematics, Matter, and Method: Philosophical Papers* (vol. 1) (Cambridge: Cambridge University Press, 1975)

Revel, J., and Ricard, M. *The Monk and the Philosopher: A Father and Son Discuss the Meaning of Life* (New York: Schoken, 1999).

Rucker, R. *Infinity and the Mind: The Science and Philosophy of the Infinite* (Sussex: Harvester, 1982).

Schick, T. W. Rorty and Davidson on alternate conceptual schemes. *Journal of Speculative Philosophy*, 1, 1987, 291–303.

Schmitt, F. F. (ed.). *Socializing Epistemology: The Social Dimension of Knowledge* (Lanham, MD: Rowman and Littlefield, 1994).

Schuon, F. *The Transcendent Unity of Religions* (New York: Harper, 1975).

Smith, H. *Forgotten Truth: The Primordial Tradition* (New York: Harper, 1976).

Stace, W. T. *Time and Eternity* (Princeton, NJ: Princeton University Press, 1952).

Stace, W. T. *The Teachings of the Mystics* (New York: New American Library, 1960).

Stace, W. T. *Mysticism and Philosophy* (New York: Macmillan, 1972).

Streng, F. J. Language and mystical awareness. In S. T. Katz (ed.), *Mysticism and Philosophical Analysis* (New York: Oxford University Press, 1978).

Suzuki, D. T. An Introduction to Zen Buddhism (London: Rider, 1949).

Tarski, A. *Logic, Semantics, Metamathematics: Papers from 1923 to 1938* (Oxford: Clarendon Press, 1956).

Underhill, E. *Mysticism* (New York: Dutton, 1911).

van Fraassen, B. *The Scientific Image* (Oxford: Clarendon Press, 1980).

Watson, J. B. *Behaviorism* (New York: Norton, 1925).

Watts, A. *Myth and Ritual in Christianity* (London: Thames and Hudson, 1954).

Whorf, B. L. *Language, Thought, and Reality: Selected Writings* (Cambridge, MA: MIT Press, 1964).

Index