The Knowability Paradox

JONATHAN L. KVANVIG

THE KNOWABILITY PARADOX

This page intentionally left blank

The Knowability Paradox

by
JONATHAN L. KVANVIG

CLARENDON PRESS • OXFORD

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

First published 2006

British Library Cataloguing in Publication Data

Data available

Library of Congress Cataloging in Publication Data

Kvanvig, Jonathan L.
The Knowability Paradox / by Jonathan L. Kvanvig.
p. cm.
Includes bibliographical references (p.) and index.
1. Knowledge, Theory of. I. Title.
BD161.K86 2006 121–dc22 2005030409

Typeset by Newgen Imaging Systems (P) Ltd., Chennai, India
Printed in Great Britain
on acid-free paper by
Biddles Ltd., King's Lynn, Norfolk

ISBN 0–19–928259–5 (Hbk.) 978–0–19–928259–3

1 3 5 7 9 10 8 6 4 2

Table of Contents

This page intentionally left blank

Introduction

The knowability paradox is a paradox deriving from a proof that if all truths are knowable, then all truths are known, first published in 1963 by Frederic Fitch.[1] The proof and the paradox have, since their first publication, existed in relative obscurity. Even works from the past decade or so explicitly aimed at defending realist theories of truth, theories seemingly buttressed by the threat the paradox raises against anti-realist theories of truth, are unaware of the paradox.[2] I recall two conversations in the early 1990s, one with Al Plantinga and the other with Bill Alston about the paradox, neither of whom had even heard of the paradox. Not yet familiar with the obscurity of the paradox, I was stunned to find philosophers whose work clearly touched on the paradox unaware of its existence.[3]

Part of the explanation of the obscurity of the paradox is the insignificant role given to it by Fitch in his original publication. Fitch's theorem, Theorem 4, states, "For each agent who is not omniscient, there is a true proposition which that agent cannot know,"[4] and Fitch notes the origin of the theorem in footnote 5: "This theorem is essentially due to an anonymous referee of an earlier paper, in 1945, that I did not publish."[5] The proof of the theorem takes only one short paragraph, and is followed by two other theorems along the same lines that are proved with equal brevity, after which Fitch turns away from this issue to issues connected with action. The problem raised by Fitch's proofs seems not to have impressed itself very much on Fitch himself, so even

[1] Frederic Fitch, "A Logical Analysis of Some Value Concepts", *Journal of Symbolic Logic*, 28.2 (June 1963), pp. 135–142.

[2] See, e.g., William Alston's *A Realist Conception of Truth* (Ithaca, 1995).

[3] Alston's work related to the paradox is his book *A Realist Conception of Truth*. Plantinga's Presidential Address to the APA "How to Be an Anti-Realist", *Proceedings of the American Philosophical Association* 56, pp. 47–70.

[4] Fitch, "A Logical Analysis of Some Value Concepts", p. 138.

[5] Ibid., p. 138, footnote 5.

careful readers of his work might pass over it to focus on what would be perceived as the more central issues.

The relative obscurity of the paradox has begun to dissipate, which is as it should be. For the paradox has deep significance for our conception of truth and knowledge. In its usual incarnation, the paradox is presented as a threat to semantic anti-realist views that endorse the idea that all truths are knowable. I argue here that this threat is an implication of a more fundamental paradoxicality, one arising from a lost logical distinction between actuality and possibility in a given domain. Philosophers have become familiar over the past forty years or so with such a lost distinction in modal contexts: when dealing with necessary truths, there is no distinction between actual and possible truth. That is, there is no logical distinction, as long as the dominant view is correct according to which S5 is the correct modal logic, to be drawn between

$$\Box p \text{ (it is necessary that p)}$$

and

$$\Diamond p \text{ (it is possible that it is necessary that p).}$$

Even so, contexts in which there is no logical distinction between actual and possible truth are the exception, and place a burden of proof on those who claim to have found such a context. For example, no such proposal regarding empirical truth has any hope of success, for there surely is a logical distinction between the claim that it is raining and the claim that it might be raining.

The aspect of the knowability paradox that is most troubling is that the paradox threatens the logical distinction between actual and possible knowledge in the domain of truth. That is, if we consider the class of truths, the proofs that constitute the paradox imply that there is here no distinction between what is known and what might be known. This result is seriously disturbing, for it is no more plausible to assume that there is no such distinction between known truths and knowable truths than between empirical truths and empirical possibilities. Such a result tells us that there is something seriously wrong with our conceptions of truth, knowledge, or possibility, or with our understanding of logical inference.

In short, my thesis is that there are two problems created by Fitch's proof. One problem is a perceived threat to anti-realism and the other problem is the paradox created by the proof. The standard assumption is that these two problems are two faces of the same coin. That assumption,

I will argue, is false. The threat to anti-realism is one problem and the paradox another and, though I will discuss both problems here, my goal is a solution to the paradox. The result, if I'm successful, will be the somewhat surprising result of finding a solution to the paradox that leaves the threat to anti-realism unanswered. The two problems arising from Fitch's proof are certainly related but, as I will argue here, a defense against the threat to anti-realism is no solution to the paradox, and a proper solution to the paradox need not disarm the threat to anti-realism.

Chapter 1 is devoted to a careful analysis of the logical structure of the paradox, but a brief description may prove useful here. The central feature of the proof involves a contradiction derived from assuming that it is known that there is such an unknown truth. If it is known that there is a particular truth that is unknown, then that truth is known; and if it is known that some particular unknown truth is unknown, then the original truth (the unknown one) is unknown. So, it would seem, it can't be known that there is an unknown truth, from which it follows that if all truths are knowable, then it can't be true that there is an unknown truth (since that can't be known by the above argument). The result is that if all truths are knowable, then all truths are known.

This brief synopsis of the fundamental argument in the paradox will, in all likelihood, be viewed with suspicion by anyone unacquainted with the paradox, and the task of the first chapter is to investigate this argument to reveal fully its underlying structure. As we will see, the proof involves no sophistical argumentation at all.

After examining the logical details of the paradox, Chapter 2 examines one line of motivation for concern about the paradox, for the paradox threatens, most obviously, anti-realist views of truth that endorse the claim that all truths are knowable. I argue there that, though there is no compelling argument for holding that all truths are knowable, this claim has much more going for it than one might initially imagine. I argue that there is a variety of positions with which this feature of semantic anti-realism fits quite naturally, and that a rejection of it puts serious tension into a broad range of overall philosophical outlooks, including, most tendentiously, theism and physicalism.

Chapters 3 through 5 investigate the approaches that have been taken in the last thirty years or so to the paradox. Chapter 3 considers approaches to the paradox that wish to save anti-realism from the paradox by denying that the knowability assumption is a commitment of anti-realism. Such approaches maintain instead that the claim that all truths are knowable must be restricted in some way in order to express

an anti-realist commitment. I argue against all examples of such an approach, and argue further that even if there were a successful restriction strategy, the paradox would remain untouched. For the fundamental paradoxicality we must address is not about whether all truths are knowable. It is, instead, about a lost logical distinction between possible knowledge of all truth and actual knowledge of all truth. The result is that restriction strategies are all red herrings when it comes to the fundamental perplexity engendered by the knowability paradox.

Chapter 4 examines the idea that the logical principles governing the knowledge operator are the root cause of the paradox. As we will see in Chapter 1, there are two such principles. The first is that knowledge implies truth, and the second is that knowledge distributes over conjunction, so that knowledge of a conjunction constitutes knowledge of the conjuncts. I argue that the paradox cannot be avoided by questioning these principles.

Chapter 5 examines the proposal that the paradox derives from our commitment to classical logic. The motivation for this maneuver is the seminal work of Michael Dummett in the philosophy of logic[6] and the way in which his work has supported intuitionistic and other alternatives to classical logic. I argue that in spite of some initial promise at being able to solve the paradox, the attempt to get rid of the problem by a change in logic fails.

In light of this last point, I pursue in Chapter 6 a strategy for solving the knowability paradox in terms of the general category of the fallacies involved in substituting into intensional contexts. It is well known that such substitutions are not always valid: from the fact that Clark Kent is Superman and that Lois adores Superman, one can't infer she adores Clark;[7] and from the fact that 9 is the number of planets, we can't infer that the number of planets is necessarily greater than 7 simply because 9 is necessarily greater than 7. As we will see when we examine the logical details of the paradox, it involves substitutions into intensional contexts as well, and that fact should alert us to the possibility that the substitution is illicit.

I argue that the paradox is another example of failure of substitutivity in intensional contexts by proposing a neo-Russellian treatment of

[6] Most of his seminal work is collected in *Truth and Other Enigmas* (Cambridge, Mass.: 1978).

[7] At least, one can't allow the inference without additional explanation as to why when asked whom she adores and whom she doesn't, Lois places the name 'Superman' on the first list and 'Clark Kent' on the second list. Direct reference theories tend to validate the inference in the text, but they can do so only by shouldering this further explanatory burden.

quantification. Philosophers of language are familiar with Russellian treatments of names and neo-Russellian treatments of natural kinds on the basis of arguments from Kripke and Putnam, among others.[8] On the basis of such arguments, such terms do not refer or pick out features of the world through the mediation of some Fregean sense or other property with which we are directly acquainted. The name 'Aristotle', for example, does not refer to a famous philosopher in virtue of our grasping some definite description that singles him out from all the other famous Greek philosophers. Again, our talking about and referring to birch trees is in no way dependent on some accurate conception of what distinguishes birch trees from other kinds of trees. Such terms have a way of reaching out directly into the world without such reach being mediated by conceptual machinery in the head.

Chapter 6 argues for a similar conception of quantification. The ordinary, Fregean view treats a quantifier as expressing a second-order property, with the domain of quantification entering the picture at a later semantical stage, the stage at which an evaluation of the truth-value of the proposition expressed is calculated. In this way, the Fregean view is akin to a descriptional theory of names, in which a sentence with a name in it expresses a proposition containing a property that could have been expressed by a definite description instead of a name, and things themselves, as opposed to properties, enter the picture at a later semantic stage, the stage at which an evaluation of the truth-value of the proposition expressed is calculated. A neo-Russellian view eschews such mediation, just as does a Russellian treatment of names, holding that the connection between the quantifier and the domain of quantification is unmediated. The task of Chapter 6 is to explain this neo-Russellian view and its relationship to the knowability paradox. In particular, I will argue that the neo-Russellian view turns the proofs underlying the knowability paradox into a particular case of failure of substitutivity in intensional contexts.

My approach to the paradox differs importantly from extant attempts to avoid it. First, I argue that my solution is the only plausible one among those presently available in the literature; in fact, we will find that it is not misleading to say that it is the only solution to the paradox itself (as opposed to the threat to anti-realism created by Fitch's proof).

[8] Saul Kripke, *Naming and Necessity* (Cambridge, Mass., 1980); Hilary Putnam, "Meaning and Reference", *Journal of Philosophy* 70 (1973), pp. 699–711, and "The Meaning of 'Meaning' ", *Minnesota Studies in the Philosophy of Science VII: Language, Mind, and Knowledge*, Keith Gunderson, ed. (Minneapolis, 1975).

Second, though I argue that my solution can provide some comfort to anti-realists who wish to maintain that all truths are knowable, I make no pretensions to salvaging anti-realism from the threat created by Fitch's proof. I sometimes will remark in discussion that the points I am making are congenial to anti-realism, but such remarks should not be construed as a defense of the view or as a commitment to the idea that a proper disarming of the paradox answers in any way to the motivations for being an anti-realist in the first place. By the end of the discussion, it will become clear that the dissolution of the paradox argued for here leaves anti-realism still in jeopardy. Finally, my approach to the paradox denies the working assumption of prior discussions of the paradox, which have treated it as a problematic implication of global anti-realism, the view that all truths are knowable. Viewed from this perspective, the problematic implication is that there is a lost logical distinction between unknown truth and unknowable truth, a lost distinction that constitutes the heart of the paradox. I, too, view the heart of the paradox in terms of a lost logical distinction between actuality and possibility, but a proper understanding of the lost distinction frees the paradox itself from any anti-realist assumptions. The knowability paradox itself is one problem created by Fitch's proof, and the threat to anti-realist conceptions of truth quite another.

My work here has benefited immensely from discussion and comments by a host of colleagues and friends. Among them are: Chris Menzel, Robert Johnson, Andrew Melnyk, Tim Williamson, John Hawthorne, Wayne Riggs, Michael Hand, Joe Salerno, Debby Hutchins, Stig Rasmussen, Paul Weirich, Peter Markie, and Matt McGrath. I wish the errors that remain were their fault.

1

The Paradox

The knowability paradox derives from the work of F. B. Fitch in 1963, in particular from one of several theorems that Fitch proved in service of examining the logic of certain value concepts.[1] This theorem was fairly well ignored in the philosophical community until rediscovered by H. D. Hart and Colin McGinn in 1976,[2] who used the theorem to show that the crucial claim of verificationism that all truths are verifiable is false. J. L. Mackie responded to this argument, claiming that verificationism could be formulated in ways that escaped the paradox.[3] Mackie argued that verificationism need not maintain an anti-realist conception of truth according to which adequate verification entails truth. Instead, verificationism could maintain only that all truths are in principle confirmable, where confirmation for p does not entail the truth of p.

More recently, Dorothy Edgington has offered a partial solution to the paradox, one which she claims yields a defense of a version of verificationism against the paradox.[4] Timothy Williamson argued against Edgington's solution,[5] and has investigated the prospects for anti-realism by rejecting classical logic in favor of intuitionistic logic in the face of the paradox.[6] Since the early 1990s interest in the paradox has grown dramatically.[7]

[1] F. B. Fitch, "A Logical Analysis of Some Value Concepts", *Journal of Symbolic Logic* 28 (1963), pp. 135–142.

[2] W. D. Hart and Colin McGinn, "Knowledge and Necessity", *Journal of Philosophical Logic* 5 (1976), pp. 205–208. See also David Bell and W. D. Hart, "The Epistemology of Abstract Objects", *Proceedings of the Aristotelian Society Supplementary Volume* (1979), pp. 154–165.

[3] J. L. Mackie, "Truth and Knowability", *Analysis* 40 (1980), pp. 90–92.

[4] Dorothy Edgington, "The Paradox of Knowability", *Mind* 94 (1985), pp. 557–568.

[5] Timothy Williamson, "On the Paradox of Knowability", *Mind* 96 (1987), pp. 256–261.

[6] Timothy Williamson, "Knowability and Constructivism", *Philosophical Quarterly* 38 (1988), pp. 422–432; "On Intuitionistic Modal Epistemic Logic", *Journal of Philosophical Logic* 21 (1992), pp. 63–89; and "Verificationism and Non-Distributive Knowledge", *Australasian Journal of Philosophy* 71.1 (March 1993), pp. 78–86.

[7] For a discussion of the literature and a fairly complete bibliography, see Berit Brogaard and Joe Salerno, "Fitch's Paradox of Knowability", *The Stanford Encyclopedia of*

The paradox is an important but largely ignored threat to verificationism and to anti-realist views of truth. I will argue in the next chapter that it is also a threat to a significant range of other positions as well, thereby showing that it is among the most far-reaching of paradoxes in terms of the variety of philosophical positions it threatens. Before doing so, however, it is important to become familiar with the details of the proofs that underlie the paradox and to see more clearly exactly which elements of these proofs give rise to paradox.

PROOF-THEORETIC DETAILS OF THE PARADOX

The theorem proved by Fitch on the way to investigating the logic of certain value concepts and from which the paradox arises is:

$$\vdash \sim \alpha(p \& \sim \alpha p),$$

where 'p' is some sentence in a formal language and 'α' is an operator of that language meeting certain restrictions. It is sufficient for meeting these restrictions that α is at least as strong as a truth operator, and that it distributes over conjunction. If we let α be the truth operator itself, then the theorem implies the unremarkably obvious idea that the following conjunction is provably untrue: p and it is not true that p. If, however, we let K be the value for α in the above theorem, where 'K' is interpreted as "it is known by someone at sometime that", we have the material for paradox. On this interpretation, the theorem records that the following conjunction is provably unknown: p and it is not known that p. That such a theorem is provable generates an inconsistency between the claim that all truths are knowable and the claim some truths are not known. According to the theorem just noted, no instance of this second claim, that some truths are not known, can be known. For to know such an instance would be to know that some claim is true and unknown, which is equivalent to knowing that a claim is true while also knowing that its truth is unknown. Since at least one instance of this assumption must be true for the existential claim to be true, it follows that there is at least one truth that is unknowable, contrary to the first claim above that all truths are knowable. In this brief account of the conflict between the

Philosophy (Winter 2002 Edition), Edward N. Zalta (ed.) <http://plato.stanford.edu/archives/win2002/entries/fitch-paradox/>.

knowability of all truth and the existence of unknown truths resides paradox, for it is jarring to our ordinary conceptions to think that the modal claim about knowability could be undermined by a humble admission that we are not omniscient. Lack of omniscience obviously implies the existence of unknown truths, but unknown truths can be known, one would think, even if they are not in fact known. So it would be surprising to learn that endorsing our epistemic limitations requires abandoning the idea that unknown truths might nonetheless be known.

A common reaction to the paradox is to think that any attempt to formulate it precisely will see its demise. I will first describe Fitch's own work and the historical development of treatment of the paradox. I will follow with the explicit derivations of the paradox that I prefer to use in the remainder of this work, with the final intent of showing that the paradox can be derived using only one extra-logical assumption. That assumption is that the operator in question is strictly stronger than truth, as the "it is known that" operator is. The rest of the paradox is nothing more than boring details of first-order theory supplemented by modal operators understood in a perfectly unexceptional way. The point to drive home in the discussion of the technical details in this chapter and their philosophical implications in the next chapter is that the paradox not only has consequences of considerable scope (as laid out in the next chapter), but is also powerful because the paradox has such scope while being derivable from such prosaic assumptions.

The paradox can be used to show that an interesting version of verificationism entails a quite silly version of the view. Silly verificationism is the view that all truths are known (or verified). The proof relies on Fitch's theorem:

$$\vdash \sim \alpha(p \& \sim \alpha p),$$

which can be proven as follows. Assume not, that is, assume $\alpha(p\&\sim\alpha p)$. Distribute the operator, yielding $\alpha p\&\alpha\sim\alpha p$. Since the operator in question is assumed to be truth-implying, the last result implies $\alpha p\&\sim\alpha p$, giving us a proof by reductio of the theorem.

Once we have proved this theorem, we simply note that the K operator, understood as "it is known at some time by some one that", satisfies the distribution rule and the truth-implying rule needed for this reductio proof, and we get $\sim K(p\&\sim Kp)$ as an instance of the theorem. We then apply two unremarkable modal claims, which can be formulated using the usual modal operators $\Diamond$ (which is read "it is possible

that") and $\Box$ (which is read "it is necessary that"). The first is the Rule of Necessitation:

(RN) If p is provable, then it is necessary, i.e., $\vdash p$ implies $\Box p$.

The second is a set of rules establishing the interdefinability of the modal concepts of necessity and possibility, rules which mirror the interdefinability of the universal and existential quantifiers in first-order logic:

$$\Box p \dashv\vdash \sim\Diamond\sim p \text{ (Dual)}$$

$$\Box\sim p \dashv\vdash \sim\Diamond p \text{ (Dual)}$$

$$\sim\Box p \dashv\vdash \Diamond\sim p \text{ (Dual)}$$

$$\sim\Box\sim p \dashv\vdash \Diamond p \text{ (Dual)}.^{8}$$

Since $\sim$K(p&$\sim$Kp) is provable as shown above, we can derive from it $\Box\sim$K(p&$\sim$Kp) by (RN), and by (Dual) we can infer $\sim\Diamond$K(p&$\sim$Kp). If some truths are unknown, there must be a specific unknown truth of the form p&$\sim$Kp. Moreover, if all truths are knowable, p&$\sim$Kp must be knowable, i.e., $\Diamond$K(p&$\sim$Kp), contradicting what was derived above. So the assumption that some truths are unknown yields a contradiction in the presence of the verificationist claim that all truths are knowable or verifiable, and a denial of that assumption is the distinctive claim of silly verificationism: $\forall p(p \rightarrow \text{K}p)$.

Why is this formulation of verificationism silly? Timothy Williamson gives the following reason:

> . . . [I]f p is the proposition that the number of tennis balls in NN's garden on 7th November, 1991 is even, and q the proposition that it is odd, both propositions have been conceived by a human, but either p is a truth never known by any human or q is; one or the other is a counterexample . . . [9]

Since silly verification is subject to such easy counterexamples, any form of verificationism that entails it must be abandoned.

We saw above a proof of the crucial theorem, employing the schematic operator α, but I want to belabor the details of the proof a bit

[8] The latter three can be derived from the first if the assumed logic is classical. I present all four claims independently because the assumption of classical logic is one of the disputes in the literature.

[9] Timothy Williamson, "Verificationism and Non-Distributive Knowledge", *Australasian Journal of Philosophy* 71.1 (1993), p. 79.

since the predominant response one encounters upon by those seeing the paradox for the first time is that some logical mistake is being made in the derivation of a contradiction. The usual presentations of the paradox employ schematic formalizations of the two assumptions that all truths are knowable and some truths are not known:

$$p \rightarrow \Diamond \mathrm{K}p \tag{1}$$

and

$$p \,\&\, \sim \mathrm{K}p. \tag{2}$$

The proof from these assumptions depends on two extra-logical rules. The first is that the K operator is at least as strong as truth, i.e., Knowledge Implies Truth:

$$\mathrm{K}p \vdash p \text{ (KIT)}. \tag{3}$$

The second is that the K operator distributes over conjunction, i.e.,

$$\mathrm{K}(p \,\&\, q) \vdash \mathrm{K}p \,\&\, \mathrm{K}q, \text{ (K-Dist)}. \tag{4}$$

The proof proceeds by substituting (2) in (1) as the value for p. Given this substitution and (2) above, by modus ponens we get:

$$\Diamond \mathrm{K}(p \,\&\, \sim \mathrm{K}p). \tag{5}$$

Suppose we then assume

$$\mathrm{K}(p \,\&\, \sim \mathrm{K}p). \tag{6}$$

Distributing the K operator over the conjunction in (6) yields

$$\mathrm{K}p \,\&\, \mathrm{K} \sim \mathrm{K}p, \tag{7}$$

and an application to the second conjunct of (7) of the truth implying property of the K operator represented in (3) leaves us with

$$\mathrm{K}p \,\&\, \sim \mathrm{K}p. \tag{8}$$

So by reductio, we obtain

$$\sim \mathrm{K}(p \,\&\, \sim \mathrm{K}p), \tag{9}$$

which, by the Rule of Necessitation

$$(\vdash p) \vdash \Box p \text{ (RN)} \tag{10}$$

gives us

$$\Box \sim K(p \ \& \sim Kp). \tag{11}$$

One version of the interdefinability of the modal operators,

$$\Box \sim p \dashv\vdash \sim \Diamond p \ \text{(Dual)} \tag{12}$$

allows us to derive from (11)

$$\sim \Diamond K(p \ \& \sim Kp). \tag{13}$$

So, by reductio, one of the two original assumptions must be abandoned, and the most vulnerable one would seem to be the first claim, the claim that all truths are knowable.

To some it will seem that this derivation glosses over too much complexity to be capable of assuring them that the paradox does not rely on some logical sleight of hand. To take but one example, the K operator has buried in it two different existential quantifiers. One might suspect, therefore, that a more careful representation of the initial assumptions might reveal some fallacy in an attempt to derive a contradiction from them.

Such is not the case. We can represent the argument using either first- or second-order quantifiers, depending on one's preferences. If we wish to use only first-order quantifiers, we proceed by adding a possibility operator '$\Diamond$', a truth predicate 'T', and a three-place relation 'K' (where 'KxTyt' is read "x knows that y is true at time t") to standard first order theory. The argument also uses two other rules of proof, the "Knowledge Implies Truth" rule (KIT), according to which one to infer p from the claim that someone knows p at some time, and a principle about the distributivity of knowledge. The distribution rule allows one to apply conjunction-elimination within knowledge contexts (K-Dist), so that if you know p&q, you know p and you know q. With this apparatus, the argument runs as follows. By assumption, we are given

$$\forall p(Tp \rightarrow \Diamond \exists x \exists t KxTpt) \ \text{(All truths are knowable)}, \tag{1}$$

and

$$\exists p(Tp \ \& \sim \exists y \exists s KyTps) \ \text{(Some truths are unknown)}, \tag{2}$$

An instance of (2) is

$$Tq \ \& \sim \exists y \exists s KyTqs, \tag{3}$$

which we can substitute into (1) as the value for p, yielding

$$(Tq \,\&\, \sim\exists y\exists sKyTqs) \rightarrow \Diamond\exists x\exists tKx(Tq \,\&\, \sim\exists y\exists sKyTqs)t. \quad (4)$$

(3) and (4) compose a modus ponens argument, the conclusion of which is

$$\Diamond\exists x\exists tKx(Tq \,\&\, \sim\exists y\exists sKyTqs)t. \quad (5)$$

Assume

$$\exists x\exists tKx(Tq \,\&\, \sim\exists y\exists sKyTqs)t. \quad (6)$$

and distribute the K predicate, applying KIT to the second conjunct of the result. We thus obtain

$$\exists x\exists tKxTqt \,\&\, \sim\exists y\exists sKyTqs), \quad (7)$$

from which we can derive

$$(\exists x\exists tKxTqt \,\&\, \sim\exists x\exists tKxTqt), \quad (8)$$

by appropriate quantifier introduction and elimination rules. By reductio, we obtain the denial of (6),

$$\sim\exists x\exists tKx(Tq \,\&\, \sim\exists y\exists sKyTqs)t, \quad (9)$$

from which, by the Rule of Necessitation, we get

$$\Box\sim\exists x\exists tKx(Tq \,\&\, \sim\exists y\exists sKyTqs)t. \quad (10)$$

Using one of the (Dual) rules describing the interdefinability of the modal operators, we obtain the denial of (5),

$$\sim\Diamond\exists x\exists tKx(Tq \,\&\, \sim\exists y\exists sKyTqs)t. \quad (11)$$

The conclusion is, then, to deny once again the most vulnerable assumption, which seems to be the one claiming that all truths are knowable.

Alternatively, we might wish to avoid using a truth predicate, and we can do so by employing second-order quantifiers ranging over zero-place predicates of our language. Doing so allows us to formulate the two assumptions as

$$\forall p(p \rightarrow \Diamond Kp)$$

and

$$\exists p(p \ \& \sim Kp),$$

ignoring the unneeded complexities introduced by explicitly representing the quantifiers implicit in the K operator. By now, it should be obvious how to generate the paradox from these assumptions; they are introduced only to provide alternative ways to generate the paradox for those squeamish in any way about the earlier derivations.

The conclusion which cannot be avoided here is that the logic of the paradox is not in any simple way problematic. Attention to the details of these various derivations suggests that if there is a problem with these proofs, it is most likely to be found in the two extra-logical rules employed in it: the distribution rule regarding the K operator, or the requirement that the K operator is at least as strong as a truth operator rather than some hidden flaw in the application of the devices of first-order theory. The issue of the special rules for the K-operator will be pursued in Chapter 3.

Before pursuing that line of inquiry, however, it is important to consider a prior issue concerning the question of whether the derivations above get us to the heart of the matter. As presented, I have been maintaining that the paradox results from an operator stronger than truth and in whose context &-Elimination is allowed. There is in the literature a challenge to this conception of the paradox, built in the thought that there are analogues of the knowability paradox employing operators weaker than truth that can seem as puzzling as the knowability paradox itself, suggesting that there is a deeper and more general paradoxicality than is found solely by attending to the knowledge operator and its close cousins.

ANALOGUES OF THE PARADOX

Those who search for a more general paradoxicality here do so by focusing on alternative interpretations of the crucial operator in the knowability paradox. Let α be a variable for any operator whatsoever; the two general assumptions of which those in the knowability paradox are but an instance are:

$$p \rightarrow \Diamond \alpha p \qquad (1)$$

and

$$p \,\&\, \sim\alpha p. \tag{2}$$

Greater generality would be achieved, then, by finding values for α that yield the same paradoxicality seen when the value is the K operator as above. The first requirement for achieving such a result is to be able to derive an instance of Fitch's theorem, $\sim\alpha(p\&\sim\alpha p)$, after which we can inquire whether the instances of (1) and (2) above are sufficiently plausible to generate paradoxicality on the basis of the Fitch result. Two such values may seem promising: mental states such as "it is thought by some person at some time that" and epistemic conditions such as "there is adequate evidence for some person at some time that".[10] I will argue that such attempts to find some more general paradoxicality fail.

Mental State Operators

Consider the first proposal, "it is thought that". The most that can be generated from this operator is the claim that

> It is thought by S that p & it is thought by S that it is not thought by S that p. (3)

The further inference to S's both thinking p and not thinking p is blocked because thinking does not imply truth.

So the question reduces to whether (3) is paradoxical, and if so, whether it is of the same kind as that involved in the knowability paradox. Some may try to find such a relationship indirectly, by first claiming that (3) is paradoxical in the same way as Moore's paradox, and then attempting to link Moore's paradox to the knowability paradox. This latter identification would be a mistake, however. Moore's paradox is one of *assertion*: to assert "*p*, but I don't believe it", is paradoxical in some way. Moreover, (3) is not paradoxical, since it is not paradoxical for *p* to be true and not be thought to be true, nor is it paradoxical to find oneself in a situation where one mistakenly thinks that one does not think that *p* is true.[11] The contents of consciousness are simply not transparent in the way required for such a claim to be paradoxical.

[10] See Dorothy Edgington, "The Paradox of Knowability," p. 558.

[11] A new link can be forged between the knowability paradox and Moore's paradox by endorsing the idea that knowledge is the norm of assertion, since it could then be argued that the Moorean sentence is not assertible because not knowable (thanks to Joe Salerno

There are, of course, philosophical viewpoints that ascribe such transparency to consciousness, but any appeal to such viewpoints here undercuts the attempt to find some general paradoxicality beyond that involved in the knowability paradox in virtue of the fact that knowledge implies truth. For an appeal to a transparency thesis about the contents of consciousness merely reintroduces the truth implication I have claimed is central to the paradox. That is, if the contents of consciousness are transparent to us, then thinking that one has a certain thought implies that one has that thought and thinking that one does not have a particular thought implies that one does not. In such a case, the operator "it is thought that" generates the Fitch result, (i.e., an instance of Fitch's theorem $\sim\alpha(p\&\sim\alpha p)$) but not at some level of generality wider than what we have already noted. For if TTp (it is thought that it is thought that p) implies Tp and T$\sim$Tp implies $\sim$Tp, then from (3) above, we can infer a direct contradiction, as we do in the knowability paradox. The conclusion then is that the operator "it is thought that" yields a paradox in the context of our assumptions only when special philosophical theses are accepted which turn that interpretation into a precise analogue of the knowability paradox. This operator gives us no reason whatsoever for thinking that there is some more general paradoxicality to be found beyond that involved in operators that are truth-implying.

There is another way to put this point. In the knowability paradox, the general truth-implying nature of knowledge is employed, i.e., for any proposition whatsoever, knowing it implies its truth. In the above case, such general factivity is not present; the operator "it is thought that" is not truth-implying. Yet, if a transparency thesis is appealed to, such an appeal yields a limited factivity to the "it is thought that" operator that is the only kind of factivity needed to generate a contradiction in the original knowability paradox. In that original paradox, it really doesn't matter that all knowledge implies truth; all that matters is that knowledge *about whether or not one has knowledge* implies truth. Just so, all that matters for the "it is thought that" operator is whether what we think about our own thoughts is truth-implying; this type of limited factivity is sufficient to generate a special case of the knowability paradox. For that reason, appeal to a transparency thesis simply undercuts

for reminding me of this point). For an endorsement and defense of the claim that knowledge is the norm of assertion, see Timothy Williamson, *Knowledge and its Limits* (Oxford, 2000). For arguments against the idea that knowledge is the norm of assertion, see my *The Value of Knowledge and the Pursuit of Understanding*, (Cambridge, 2003), chapter 1.

the attempt to find a more general paradoxicality underlying that found in the knowability paradox beyond that of truth-implying operators.

Moore's paradox differs from the knowability paradox in another way. Moore's paradox implies no necessary falsehood as does the knowability paradox. Self-deception or confusion of the sort required to affirm Moore's paradoxical claim is possible. This fact is confirmed by the usual treatments of Moore's paradox which argue that the paradox is not a semantic paradox, but involves some kind of pragmatic or epistemic defect of speech or thought.[12] Perhaps such a belief would be epistemically self-defeating, or the remark indefensible. Each of these claims, however, is compatible with consistently thinking precisely the conjunction in question.

So the operator "it is thought by someone at some time that" provides no counterexample to the claim that the key to the knowability paradox is the factive character of the K operator. There are, nonetheless, other possibilities involving mental states that require discussion. Consider the operator "it is doubted that". No Fitch result can be obtained using this operator, but suppose we place this operator in the context of an infallible and omniscient being. Doing so yields a new operator "it is doubted by some infallible and omniscient being at some time that". For such an operator, the Fitch result can be obtained, i.e., it is provable that

$$\sim D_I\ (p\&\sim D_I p),$$

where "D_I" is the operator in question. But the reason this Fitch result is provable is that no infallible and omniscient being can experience doubt. (If the reader wishes to question whether certain beliefs are incompatible with doubt, I will change the example to refer to the mental state of believing with less than full confidence, which is obviously incompatible with the certainty an infallible being experiences regarding whatever such a being believes.)

This operator provides a counterexample to the claim that a Fitch result depends on factivity at some level, but it does not threaten the view I've articulated that the paradoxicality involved in obtaining Fitch's result depends on factivity. For the air of paradox to arise, we need also to establish that the premises involved in the derivation of a Fitch result have some plausibility. In the case at hand, it would need to be plausible to suppose that some truths are doubted by an infallible and omniscient being, and it would need to be plausible to suppose that any truth can be

[12] See, e.g., Jaakko Hintikka, *Knowledge and Belief* (Ithaca, 1962).

doubted by an infallible and omniscient being. Neither claim is plausible; in fact, both claims are necessarily false and obviously so. So it should come as no surprise whatsoever to find out that they imply a contradiction.

The same kind of response can be given to other attempts at trivial counterexamples. Suppose M is a mental state and the related operator "it is M'ed by some being at some time that". There is no reason to suppose that this schema generates counterexamples to our original thesis, but trivial counterexamples can be easily generated. Consider the mental state of meta-M: the state involved in taking mental attitude M toward a content which includes mental attitude M. Combining these ideas, we can generate the following operator: "it is M'ed by some being for whom meta-M-ing is impossible whether it is M'ed at some time that". Call this operator $M_{\sim M}$. By the definition of $M_{\sim M}$, it is provable that

$$\sim M_{\sim M}(\sim M_{\sim M} p),$$

and, if we stipulate the $M_{\sim M}$ has the distributive property (if the reader wishes, we can build that property into the operator itself), provable that

$$\sim M(p_{\sim M} \ \& \sim M_{\sim M} p).$$

So $M_{\sim M}$ is an operator that is not factive in any way, and yet for which a Fitch result obtains. Just as before, however, the original assumptions involved in the derivation are obviously and necessarily false. As such, the derivability of a Fitch result fails to provide a counterexample to the claim that the concept of factivity is central to a proper understanding of the knowability paradox.

So it is true to say that Fitch results can be obtained without involving any type of factivity, but only by building into the operator special cognitive abilities or limitations that themselves account for the Fitch result. In the cases examined, the versions of the original assumptions used to derive the Fitch result (instances of (1) and (2) at the beginning of this section) are necessarily false, and nothing paradoxical results from demonstrating that a logically impossible claim implies the denial of another logically impossible claim. More generally, to find a counterexample to my claim that factivity is central to the knowability paradox, one will have to obtain a Fitch result from plausible instances of (1) and (2), the original assumptions underlying the knowability paradox. Our discussion suggests that mental state operators will not be able to meet these conditions, either because they do not generate a Fitch result,

or because the Fitch result is not generated from plausible original assumptions.

Epistemic Condition Operators

Much of what we have just noted about the operator "it is thought that" is true of the other kind of example above, where the operators are epistemic ones such as "there is adequate evidence for S that". Similar to what we found in the above case, the two assumptions imply

> there is adequate evidence for S that p & there is adequate evidence for S that there is not adequate evidence for S that p. (4)

At first glance, this claim is a bit more perplexing than that found in (3) above. One reason for our perplexity might be, however, that the operator in question partakes of limited factivity of the sort seen earlier. That is, it might be the case that evidence about our evidential state is truth-implying; that if we have adequate evidence that we have (or lack) adequate evidence for p, that meta-evidence cannot be misleading, i.e., that evidence that there is evidence for *p* implies that there is evidence for *p*. Internalists about epistemic status are most inclined toward such a view, for one quite common version of the view requires of a theory of justification that one be able to tell by reflection alone what epistemic state one is in. One very straightforward way to generate such a result is to claim that epistemic status partakes of the type of limited factivity under discussion here.[13]

As in the prior case, however, any such appeal to the limited factivity of epistemic operators weaker than knowledge undercuts the attempt to find some more general paradoxicality beyond that requiring factivity of some sort. So, if there is to be any hope of finding some deeper issue here than that found in the knowability paradox itself, we will have to assume that the epistemic operators in question do not partake of

[13] It is worth noting that various anti-realist commitments can founder on Fitch-like results even though they are not strict instances of the paradox. For example, if an anti-realist holds that any truth can be warranted while true, we can derive that all truths are warranted by substituting into this general formula the claim that p is true but unwarranted. Such problems are analogues of the paradox but are not instances of it, though we should expect that any solution to the paradox will have something to say regarding these further difficulties for anti-realism.

limited factivity (i.e., deny that evidence that there is evidence for *p* implies that there is evidence for *p*).

Yet, if we suppose that no such factivity obtains, the air of paradox about (4) dissipates rapidly. For (4) is nothing but a denial of the limited factivity claim required to deduce from (4) the characteristic contradiction of the knowability paradox. If the type of limited factivity needed to generate a precise analogue of the knowability paradox fails, then it is possible to have adequate evidence that one lacks adequate evidence for p, and yet have adequate evidence for p—just what (4) claims.

So a defender of the view that the knowability paradox is but a special case of a deeper paradoxicality displayed by (4) is in the awkward predicament of insisting that (4) might be true, but is paradoxical nonetheless. Consider Dorothy Edgington's attempt to do so:

If we restate the argument in terms of evidence, we are able to derive

It is logically possible that someone should at some time have evidence that both *p* and no one at any time has any evidence that *p*.

It is clear that no possible state of information could support the hypothesis:

p and no one at any time has any evidence that *p*.

The conclusion is as paradoxical as before.[14]

To the contrary. If no possible state of information could support such a hypothesis, one would think that it would also be true that one could not simultaneously have evidence for *p* and evidence that no one has evidence for *p*. If so, however, precisely the limited factivity needed to generate a version of the knowability paradox is present, contrary to Edgington's intent to show that the truth-implying character of knowledge is not really needed for paradoxicality.

There is, of course, a further possibility to consider, and that is that the operator is not limitedly factive, but that there could be no state of information to support this claim. After all, requiring otherwise smacks of verificationism of an obviously inadequate sort. Even if that point were true, however, any paradoxicality involved in such a situation would not be similar to that involved in the knowability paradox. The claim would simply be one that might be true, but could never be discovered to be true. That would constitute a refutation of strong verificationism as well as the claim that all truths are knowable, but it

[14] Edgington, "The Paradox of Knowability", p. 558.

would do so on the basis of some fact not derived from the central claim of strong verificationism as in the knowability paradox itself. Furthermore, and this is the more telling point, the primary paradoxicality of the knowability paradox is not that it purports to refute strong verificationism or anti-realism, but that it implies a lost logical distinction between actual and possible universal knowledge. I will argue for this construal in detail later, but it is important to see in the present context how this understanding of the paradox makes the present issue different from that which is central to the knowability paradox.

It is also worth nothing that it is not plausible to think that (4) cannot be supported by any state of information. We can argue for this point by showing how to construct confirmation for the conjunction in question. First, suppose there is evidence that p is true. Suppose also, however, that there is evidence that there is no evidence for p; that is, there is (second-order) evidence against the existence of such (first-order) evidence for p. Both kinds of evidence are simultaneously possible, since having evidence for p does not guarantee the truth of p. If both of these sources of information are present, one need only put them together to form evidence for the conjunctive claim that p is true and no one has confirmation for p. After all, that is how we usually obtain evidence for a conjunction: take evidence for each conjunct and combine the two bodies into one, which then provides evidence for the conjunction.

One might think that any decent account of the way in which bodies of evidence interact will require that the two bodies of information possessed in such a case function so as to defeat the confirming power of at least one of the two bodies of evidence. So either there will be no (all-things-considered) confirmation for p, or there will be no (all-things-considered) confirmation for the claim that there is no confirmation for p.

Such a requirement may be true, but if it is true, then limited factivity obtains for the operator in question. Notice that the claim is that the separate bodies of information must interact in such a way that the confirming power of one of the two bodies of information is defeated by the other. If so, however, then it is not possible to have adequate evidence for a claim and also to have adequate evidence that one does not have adequate evidence for that claim, and this impossibility is simply a reformulation of the point that the operator in question partakes of limited factivity.

It is worth noting, however, that there is some ground for denying that bodies of information must interact in this way. Justified inconsistent beliefs are possible, and the reasons for this possibility lead to a

rejection of this interaction assumption about bodies of evidence. Consider, for example, the preface paradox. Billy Bob writes a book on the craft of calf roping. The book is a complex and thorough treatment of the subject, well beyond what anyone who knows only the author's name would expect. After completing the book, Billy Bob writes a reflective preface to the book. In the preface, he admits that the subject is a very complex one and that, though he has done his very best to get all the details correct, he is sure that he has missed some subtleties and thus that errors remain. He therefore asks for input to help correct such errors in case there is a second edition of the book.

The paradox arises if we suppose that Billy Bob is justified in believing both what he originally wrote in the book as well as the preface statements regarding the existence of errors in the book. If we also hold that justification is transmitted from premises to conclusions through competent deductions, then we can derive the paradoxical result that justified contradictory beliefs are possible (as long as we refuse to insult Billy Bob by questioning his capacity for competent deductions).[15]

One might approach this paradox by insisting that the inconsistent beliefs held by Billy Bob cannot all be justified, but that is a mistake. Billy Bob's predicament is simply a particular instance of the general human condition engendered by awareness of our own fallibility. We have overwhelmingly good reasons for thinking that some of our present beliefs are false, if we have a reasonable appreciation of our fallibility; but being justified in thinking that some of our present beliefs are false does not prevent us from being justified in holding those present beliefs. The alternative to allowing for the possibility of justified inconsistent beliefs is to insist that one not form an opinion on whether some of our present beliefs are false in spite of the overwhelmingly good evidence that some of them are. Such an insistence makes explaining oneself quite awkward:

THERAPIST	Are some of your present beliefs false?
SELF	I hold no opinion on the matter.
THERAPIST	Well, is it likely that some of your beliefs are false?
SELF	Yes, of course.
THERAPIST	Can we assume that each of your beliefs is highly likely to be true?
SELF	Yes, I'm an epistemologist! I have evaluated each belief and retained only those that are justified.
THERAPIST	Are some of your justified beliefs more likely to be true than others?
SELF	Yes.

[15] For a defense of a closure principle for knowledge along these lines, see John Hawthorne, *Knowledge and Lotteries* (Oxford, 2004).

THERAPIST OK, take one of your justified beliefs that is least likely to be true.
SELF OK, I'm thinking of one.
THERAPIST Which is more likely to be true—that belief or the claim that some of your beliefs are false?
SELF Do I have to be honest?
THERAPIST We won't get anywhere if you aren't.
SELF OK, in all honesty, I have to admit that the justified belief I'm thinking of is less likely to be true than the claim that some of my present beliefs are false.
THERAPIST Yet, you refuse to believe that some of your present beliefs are false?
SELF Yes, I have to refuse.
THERAPIST Why?
SELF Because if I don't refuse, I'll have inconsistent beliefs and they'll all be justified and that's paradoxical.
THERAPIST (*aside*) This is going to take a lot longer than I had thought!

The point here is that our fallibility is so obvious to each of us that we'd have to adopt a new cognitive goal in order to explain why we don't believe that some of our present beliefs are false. Our ordinary cognitive goals include items like trying to get to the truth and avoid error, trying to make sense of the variety of experiences ourselves and others have, and the like. Avoiding inconsistent beliefs is simply not one of our usual array of cognitive goals.

One might object that if one's beliefs are inconsistent, one has foregone any possibility of satisfying the goal of getting to the truth and avoiding error; so perhaps the goal of getting to the truth and avoiding error implies by itself that one can't have justified inconsistent beliefs, insofar as justification is to be understood in terms of the getting to the truth and avoiding error. The first point is correct: if one holds inconsistent beliefs, one is guaranteed to have at least one false belief. Even so, to find a prohibition from inconsistency on this basis is to give too much weight to the avoidance of error and too little weight to finding the truth. It has long been noted by epistemologists that some weighting is needed between these two aspects of the goal in question, for one can be guaranteed to find the truth if one simply believes everything and one can be guaranteed to avoid error if one refuses to believe anything. A balanced cognitive life will take some risks at both ends. Doing so involves adopting belief forming policies that balance the risks in an acceptable way, and then abiding by those policies. The policies, then, establish what quality of evidence is needed in order to justify a belief, and it shows a foolish obsession with avoiding error to include in one's policies a clause that prohibits following the evidence

where it leads when that guarantees that one will have at least one false belief. A foolish obsession in the other direction—an obsession that one will miss out on an important truth—would counsel adding a belief anytime it is consistent with what one presently believes. Better policies are found by adopting ones that take into account risks on both sides, refusing to add special qualifier clauses to the theory of adequate evidence embodied in those policies to prevent some perceived horror of having a false belief or of failing to have a particular true one.

So, an adequate approach to the preface paradox requires seeing it as a case of justified inconsistent beliefs, where the inconsistency can even be patently obvious to the person in question. The problem posed by the paradox is the problem of explaining how such justified inconsistent beliefs are possible without having to embrace the idea that one can be justified in believing an explicit contradiction.[16]

One thing to note about this problem is that for it to arise, there must be two independent bodies of information that fail to interact in the way needed to sustain Edgington's claim that it is paradoxical for one to believe with justification the conjunction that p is true but unjustified. If such bodies of information always interacted in the prescribed way, this diagnosis of the problem involved in the preface paradox would guarantee that no solution to the paradox could be found, for it would be equally impossible for confirmation to exist for both the claim that the claims in the body of one's book are all true and the preface claim that some statements in the book are mistaken.

There is an important difference between Edgington's example and the preface paradox, for one approach to the preface paradox prohibits the use of conjunction introduction within the context of a confirmation operator,[17] whereas that rule is required to obtain evidence for the conjunctive claim that p is true but unevidenced. The difference is fully appropriate, for the reason for not using conjunction introduction in the case of the preface paradox is to avoid violating a condition of adequacy on a theory of justification, according to which a theory is adequate only if it does not allow both p and $\sim p$ to be justified at the same time. So conjunction introduction is restricted in order to honor the platitude that contradictory justified beliefs cannot occur. This reason does not carry over to the present case, however, because the conjunction *p is true*

[16] See, for example, Richard Foley, *The Theory of Epistemic Rationality* (Cambridge, Mass., 1986), especially chapter 6.

[17] See, e.g., Henry Kyburg, "Conjunctivitis", in Marshall Swain, ed., *Induction, Acceptance, and Rational Belief* (Dordrecht, 1970), pp. 55–82.

but unconfirmed is clearly not a contradiction. So the application of conjunction introduction to yield that conclusion violates no obvious conditions of adequacy on a theory of justification, and hence should not be viewed as a problematic application of that rule.

So the problem raised by the operator "it is confirmed that" lacks the depth found in the knowability paradox. The problem might be only a special case of the paradox (if "confirmed" partakes of limited factivity), but when it is not, it relies on additional epistemological assumptions to make the problem appear. Moreover, these assumptions about the interaction of bodies of evidence are false, as is shown by the above discussion of the preface paradox. In contrast, these special epistemological assumptions do not underlie the knowability paradox, and even if these assumptions were all true, the paradoxicality involved would still differ from that involved in the knowability paradox. The paradoxicality in Edgington's example, if that example were successful, is just the paradoxicality involved in thinking that anti-realism could be refuted so easily. The paradoxicality engendered by Fitch's proof, however, is different. The proof may threaten anti-realism, but the heart of the paradox is located elsewhere, located in the lost logical distinction between actual and possible universal knowledge.

Moreover, the operator "it is confirmed that" can be thought of as an arbitrarily chosen example of a non-factive epistemic operator. The results we have found regarding it are appropriately generalizable to other non-factive epistemic operators, such as "it is justified that", "it is evidenced that", "there is adequate evidence that", "there is all-things-considered confirmation that", and the like. We may therefore conclude that it does not appear plausible in the least to suppose that consideration of other non-factive epistemic operators will generate paradoxical results of the sort found in the knowability paradox.

TENNANT'S ARGUMENT

The most thorough attempt in the literature to find a non-factive operator that generates knowability-like paradoxical results is developed by Neil Tennant.[18] Tennant uses the example of the operator "wondering whether" together with Moore's paradox to argue that the

[18] Neil Tennant, *The Taming of the True* (Oxford, 1997), chapter 8.

knowability paradox is but an instance of a more general phenomenon. Tennant advances the discussion by laying out in explicit and rigorous fashion the logical principles used in finding a non-factive operator that nonetheless creates a Fitch result. We can begin the discussion of Tennant's argument by noting the principles his argument uses:

> Rule of Rational Commitment (RC): If a (suitably idealized) rational thinker T has attitude A toward $p1 \ldots pn$, and $pn + 1$ follows from $p1 \ldots pn$, then T has A toward $pn + 1$.
>
> Rule of Credibility (C): If the claim that T believes p is consistent, then that claim is consistent with the truth of p.
>
> Rule of Self-Intimation (SI): If T takes A toward p, then T believes that T takes A toward p; and if T fails to take A toward p, then T believes that T fails to take A toward p.
>
> Rule of Assertion/Belief (AB): If T's belief that p is inconsistent, then so is his assertion of p.
>
> Rule of Wonderment/Belief (WB): T's wondering whether p is inconsistent with T's believing p.

Two Rules relating Wonderment, Belief, and Negation that Tennant employs are:

> (WB$\sim$) Believing$\sim p$ is inconsistent with wondering whether $p\&q$.
>
> (W$\sim$B) If T wonders whether $p \,\&\, q$ and fails to believe p, then T wonders whether p.

Tennant first uses these principles to present a resolution of Moore's paradox, and then uses them to show that, where 'W' means "Someone at some time wonders whether", $W(p\&\sim Wp)$ has the same kind of inconsistency found in the knowability paradox regarding the claim $K(p\&\sim Kp)$.

Tennant wishes to resolve Moore's paradox by showing that the assertion of *p and I don't believe it* is inconsistent, i.e., that, using the principles above, we can demonstrate the truth of $\sim A(p\&\sim Bp)$. The argument therefore begins by assuming

$$A(p\&\sim Bp). \qquad (1)$$

Tennant further assumes that "sincere declaratives betoken belief",[19] i.e., that from (1) we can derive

$$B(p\&\sim Bp). \tag{2}$$

We can then derive, by (RC)

$$Bp. \tag{3}$$

We also note that

$$\sim Bp \tag{4}$$

follows if we assume that the content of the belief at line (2) is true. So from the assumptions of the belief and its content, we can derive a contradiction ((3) and (4)). By (C), it follows that $B(p\&\sim Bp)$ itself is inconsistent, i.e.,

$$\sim B(p\&\sim Bp). \tag{5}$$

Since the original assumption at line (1) implies (2), we can conclude

$$\sim A(p\&\sim Bp), \tag{6}$$

i.e., that Moorean assertions are inconsistent for rational thinkers.

Similarly, Tennant tries to show that $W(p\&\sim Wp)$ is inconsistent. Assume

$$W(p\&\sim Wp). \tag{1}$$

By (WB) we get

$$\sim(Wp\ \&\ Bp). \tag{2}$$

Next, show by subproof $\sim Bp$, beginning by assuming

$$Bp. \tag{3}$$

(3) yields, by (WB),

$$\sim Wp, \tag{4}$$

from which by (SI) we can derive

$$B\sim Wp. \tag{5}$$

[19] Tennant, *The Taming of the True*, p. 250.

By (RC), we can derive from $B{\sim}Wp$ and Bp

$$B(p\&{\sim}Wp). \tag{6}$$

We can then derive, by (WB),

$${\sim}W(p\&{\sim}Wp) \tag{7}$$

Since (7) contradicts (1), we infer

$${\sim}Bp, \tag{8}$$

thereby completing the subproof.

If we then combine (8) with the original assumption (1), we get

$${\sim}Bp\,\&\,W(p\&{\sim}Wp). \tag{9}$$

From (9) we can derive

$$Wp \tag{10}$$

by (W∼B), and then

$$BWp \tag{11}$$

by (SI). But by (WB∼), BWp is inconsistent with $W(p\,\&\,{\sim}Wp)$, thereby showing that

$${\sim}W(p\&{\sim}Wp). \tag{12}$$

Tennant thus claims to show a strong connection between Moore's paradox, the knowability paradox, and other operators that are not truth-implying, such as the operator concerning wondering whether a claim is true. If successful, Tennant's proposal would show that the central paradoxicality involved in the knowability paradox is not distinctively about an operator stronger than truth, contrary to what I have been claiming.

Both proofs fail, however.[20] In the first proof, Tennant assumes that from

$$A(p\&{\sim}Bp) \tag{1}$$

we can derive

$$B(p\&{\sim}Bp), \tag{2}$$

[20] My discussion here mirrors results achieved independently by David DeVidi and Tim Kenyon in their excellent paper on Tennant's operator. See "Analogues of Knowability", *Australasian Journal of Philosophy* 81.4 (2003), pp. 481–495.

on grounds, as noted above, that sincere assertion expresses belief. Tennant is not entitled to this justification, at least not as a logical principle. First, it is neither analytic nor a priori, for there are too many interesting philosophical positions that deny it for it to be so. Stephen Stich, for example, doubts that there are any beliefs,[21] and when he sincerely asserts his views, we do not find our interpretive situation reduced to absurdity. Pyrrhonian skeptics counsel against the possession of beliefs, holding that mere acquiescence to the appearances will leave our practical ability to cope with the world intact and improved. Among such practical abilities is the ability to make sincere assertion, so Pyrrhonians join with Stich and others in a commitment against Tennant's principle. And William James suggests to those unable to choose to believe in the existence of God the wisdom of adopting theistic practices and hoping for the best, some of which surely involving sincere assertion of claims not (yet) believed. Most telling, however, is the common experience reported by therapists, of having to work to correct the self-understanding of their patients who sincerely report certain self-assessments but who obviously do not believe what they are reporting. In fact, many reflective people come to view their former selves as being in precisely that predicament, the predicament of sincerely mouthing certain claims but clearly not actually believing them. If the Moorean claim is paradoxical, its paradoxicality will not be revealed by trying to find analytic or a priori logical connections between assertion and belief, and it is doubtful that there is anything more than a strong probabilistic connection between the two. A proof of absurdity, however, requires more than a contingent, defeasible evidential relationship between (1) and (2).

The second proof relies on purported logical principles about the connection between wondering and believing that are not compelling. (WB), for example, claims that wondering and believing are inconsistent when directed at the same content—that is, for idealized rational agents, it is impossible both to believe that, for example, God exists and also wonder whether God exists. To the extent that doubt and wondering are linked, this result is implausible, since belief and doubt are not incompatible. To doubt that *p* does not imply, of course, that one wonders whether *p*, but it does give one reason to so wonder; and sometimes we do what we have reason to do. Furthermore, it is not clear that if one wonders, one has thereby abandoned one's initial belief.

[21] Stephen Stich, *From Folk Psychology to Cognitive Science* (Cambridge, Mass. 1983).

When skeptical theists say, "I'm a theist and have always been, but I sometimes wonder whether there really is a God," they do not contradict themselves. Furthermore, holding such views need not imply any lack of idealized rationality, for the same reasons that belief is not incompatible with doubt for idealized rational agents.

Tennant believes he can grant this point and still make his case. He says,

> In the case of analytical disagreement here, let us simply define the concept schmondering whether so that if one either believes or disbelieves that ϕ, then one cannot (logically) schmonder whether φ. In all other respects the newly defined concept will be like the . . . concept of wondering whether.[22]

I will return to this ploy later, but for now it is worth noting that such a definition fails to guarantee that there is any possible mental state corresponding to it. For example, suppose someone had said, "Take the mental state of hoping for something, and now consider the related concept which is just like hoping but without any non-cognitive features whatsoever." I don't know whether there could be any such mental state, even if the concept so described implies no contradiction. Thus, this reply is only as strong as some as yet forthcoming defense of the possibility of such a mental state.

Principle (SI), the principle according to which mental states are all self-intimating, is also problematic for similar reasons. It is simply not a logical truth at all; in fact, it is false. Tennant remains undeterred, however, appealing to the idealization involved in his concept of a rational thinker, holding that it is "at least logically consistent . . . to postulate a rational thinker . . . whose attitudes are self-intimating".[23]

Such a rejoinder is troubling, however. First, it is not obvious that such a postulation is consistent. What is true is that there is no obvious inconsistency in the idea, but it would be nice to have a defense of the consistency of the postulate rather than relying on any form of argument from ignorance.

Second, idealizations are useful or not relative to some purpose or other. Scientists idealize the phenomena in order to get a model which is useful for predictive and explanatory purposes; they don't idealize in order to reveal more accurately the truth of the matter. Bayesians sometimes idealize to the point of logical omniscience in order to reveal more clearly the nature of empirical confirmation and rational belief revision in empirical matters (though I make no endorsement of these

[22] Tennant, *The Taming of the True*, p. 252, n. 7. [23] Ibid., p. 248.

claims here). What justifies the idealization here that allows claiming that (SI) is a logical principle governing the attitudes of the idealized rational thinkers? I don't know the answer to that question, but I do know what I want to hear, but can't. I want to be told that the idealization allows us to see more accurately what the phenomenon of wondering whether is really like. It is clear, however, that such an answer is utterly implausible.

Tennant appeals to idealization to rescue (RC) as well. This principle claims that rational belief is closed under logical entailment. It doesn't take a deep grasp of the literature of the last thirty years or so to know that any kind of closure principle in epistemology will be controversial.[24] And it is worth noting the incredible amount of idealization of cognitive abilities required for the truth of Tennant's closure principle. In particular, he must assume that rational thinkers grasp the content of everything implied by what they believe. Methinks they have more than a three-pound brain. Equally troubling, such an assumption is at odds with the fact that the logical consequence relation is not clearly decidable, a fact of which Tennant is well aware.[25] This issue perhaps need not concern realists when they idealize, but it is an important issue for anti-realists such as Tennant.

There is a deeper point to make here about how radical an idealization Tennant is making. Unless we idealize the cognizer in question past the point of fallibility, the preface paradox will continue to show that (RC) is false even for idealized rational agents. For if you are fallible, it can be reasonable for you to believe that some of your beliefs are false even though it is a logical consequence of your beliefs that non of your beliefs are false, in which case (RC) is violated.

The final point I want to make about Tennant's extra-logical rules concerns the principle of credibility, according to which if the proposition *T believes p* is consistent, then *T believes p* is consistent with the truth of *p*. Suppose, however, that *p* itself is inconsistent; then it is inconsistent for every proposition whatsoever. Even so, the claim that some person believes an inconsistency is not itself inconsistent. If it were, it would be impossible for anyone to believe an inconsistency, but such is not impossible.

Two deeper problems arise for Tennant's argument, however, when we note the implications of combining Tennant's defenses of (WB) and

[24] For a penetrating discussion of this issue, and a thorough guide to the literature, see John Hawthorne, *Knowledge and Lotteries* (Oxford, 2003).

[25] Tennant, *The Taming of the True*, chapters 6–8.

(SI). The first problem is that these defenses will allow certain forms of inference that are known to be defective. To see why, take the operator "it is insinuated that" (I) and ask whether it is coherent to assume $I(p \& \sim Ip)$. To do so, one would have to find a way to insinuate, say, that Billy Bob is an idiot in a way that also insinuated that there was no such insinuation. That is political insult at its finest. I know of people who are skilled in the art.

Now let's "pull a Tennant" on the art. We insist that insinuating and believing are inconsistent (the analogue of (WB)), and if you disagree, we define a new concept, *schinsinuating*, which lacks precisely the logical features of insinuating that led to your objection. We also assert that insinuating is closed under logical implication (the analogue of (RC)) and that it is self-intimating (SI). You disagree. But we say it is consistent to postulate a suitably idealized cognizer for whom both of these claims are true. We then produce the analogue of Tennant's argument above to show that the operator "I" generates a Fitch result. Is this an interesting result? No, and nothing can be gained by way of insight into the paradox or requirements on a solution to it by such connivance.

So the first deeper problem, beyond the difficulties in defending the principles Tennant's argument employs, is that we can mimic that argument with other operators to derive conclusions that are known to be false. The second deeper problem adverts to the discussion above in which technical and trivial attempts at counterexamples to my claim that factivity of the operators in question is the key to the knowability paradox, attempts such as that involved in the operator "it is doubted by an infallible and omniscient being that". Such operators show that it is possible to generate Fitch results without the operator in question being factive, but as we saw, such results generate no air of paradox since the assumptions that some truths are doubted by such a being and that any truth can be doubted by such a being are necessarily false. Once the necessary falsehood of these assumptions is noted, obtaining a Fitch result fails to generate a paradox. This point is relevant to Tennant's example, since the combination of (WB) and (SI) entails that Tennant's idealized cognizers are incapable of meta-wonderment, the kind of wondering involved when we wonder about our own mental states or lack thereof. The argument that meta-wonderment is impossible is simple: (SI) implies that the idealized cognizers always have a correct belief about what mental states they are in, and (WB) states that wondering and believing are incompatible; hence, meta-wondering is impossible for these idealized cognizers.

This fact shows that even if all the other problems already discussed were to disappear, Tennant's argument concerning the operator "wondering whether" suffers from the same problem that plagued the attempted counterexample relying on the operator "it is doubted by some infallible and omniscient being that". A Fitch result might be obtainable, but only at the expense of the plausibility of the original assumptions, especially the analogue of the assumption that all truths are knowable. The analogue of this assumption in Tennant's argument is logically equivalent to the claim that any truth can be wondered about by a being incapable of meta-wondering at the same time about one's wondering, and that claim is quite obviously logically impossible (since some of the truths will be truths about what such a being is wondering about). As a result, it will not surprise us to learn that this claim is inconsistent with other claims, since a logically impossible claim is inconsistent with everything.

Tennant's example initially looks more promising than it really is, because it plays on the ordinary operator "wondering whether", and it is certainly plausible to suppose both that any truth can be wondered about and that some truths will never be wondered about. Appeal to this operator in the discussion is misleading, however, and careful attention to Tennant's maneuvers show it. The operator "wondering whether" is soon replaced by a new operator "schmondering whether", and this operator is soon supplemented by additional clauses specifying the kind of idealized cognizer whose schmondering is in view. In the end, these additional clauses render Tennant's operator such that it logically entails the operator "it is wondered by a cognizer incapable of wondering about any of its own mental attitudes at some time that" (where one operator O logically entails operator O' if and only if for all p, Op implies O'p). This operator renders logically impossible the analogue of the claim that all truths are knowable. That claim is logically equivalent to the claim that all truths can be wondered about by a being incapable of meta-wondering, even those truths that could only be wondered about by engaging in meta-wondering. Nothing paradoxical results from finding out that some further proposition is inconsistent with a logically impossible one. In sum, Tennant's discussion reveals a non-factive operator for which a Fitch result can be obtained, but regarding which nothing paradoxical exists since the derivation of the Fitch result relies on an analogue of the claim that all truths are knowable that is logically false.

Common to all of the above attempts to find something analogous to the knowability paradox is an operator weaker than truth. Most

important to note is that in such cases, the air of paradox surrounding such operators can be explained by noting the philosophical possibility that each such operator partakes of limited factivity. If such possibilities are realized, if the operators in question are iterationally redundant, both operators generate precisely the paradox at the heart of the knowability paradox. The lesson to be learned is that the heart of the knowability paradox involves operators that are iterationally redundant.

As we have seen, not all operators that yield a Fitch result also generate a paradox. We have seen several examples above of such, but the simplest examples are truth and actuality operators. Fitch results can be obtained for both of these operators, but clearly the assumptions used to generate these results are implausible. In particular, the second assumption would be either that there is some truth that is not true, or not actually true, and both claims are obviously false. The Fitch result would show only that these obviously false claims need to be rejected, and there is nothing paradoxical in the least in drawing such a conclusion.

CONCLUSION

I have argued for a certain understanding of the key element in obtaining the paradoxical results involved in the knowability paradox, and I have argued as well that there is no simple and obvious logical mistake in the derivation of the knowability result. Still, a paradox has deep significance only if it arises from plausible premises. Those in question in Fitch's proof are the claim of epistemic modesty, that some truths will never be known, and the knowability principle that all truths are knowable. Though the first claim is quite plausible, the second claim does not have the same intuitive pull as the first. In spite of this lack on the part of the knowability claim, there are substantive grounds in its favor, grounds that some hold show that all truths are knowable but which show at the very least that it is plausible to maintain that all truths are knowable. To this issue we turn in the next chapter.

2

What's Paradoxical?

The paradox that is the focus of this work begins from a commitment to a claim about the nature of truth, a commitment to the effect that all truths are knowable. Upon seeing the derivation of Fitch's result, a common reaction is to shrug off its significance, thinking that there was little reason to begin with to think that all truths are knowable. Such nonchalance is hardly justified, I believe, and I want to pursue several lines of thought to motivate greater concern about the paradox.

A first way to show that the results of the last chapter are truly paradoxical is to argue for the first assumption of the proof, namely, that all truths are knowable. The first part of this chapter will be devoted to that claim, for the variety of positions in philosophy that are committed to the claim, or for which the claim is a natural accompaniment, is much broader than is usually appreciated.

In the end, however, I do not think that the paradoxicality that exists here is a result of the attraction of the knowability assumption. The chapter will close with an explanation of the heart of the paradox, which has little to do with whether one is inclined to believe that all truths are knowable.

ARE ALL TRUTHS KNOWABLE?

There is strong motivation from a number of different directions for thinking that all truths are knowable. In particular, there are two general ways that have led philosophers either to presuppose or endorse something like the claim that all truths are knowable, one metaphysical and one semantic. I begin with the semantic route.

Interest in the nature of truth and meaning have long been topics of philosophical inquiry tracing at least to Socrates' perplexity about the possibility of false belief in the *Theaetetus*, and perhaps earlier to the

Parmenidean conclusions about the connection between the nature of reality and the possibility of what can be (truly) said and what cannot. In our own post-Kantian era, however, a certain approach to truth and meaning has come to play an increasingly important role in discussions concerning these issues, an approach to truth and meaning has come to be known as anti-realism. Since there is a variety of positions called by this name, the term itself will do little to convey what view is in question, but we can begin by first distinguishing between metaphysical realism and semantic realism. The first view signals an ontological commitment, so to be a metaphysical realist about numbers is to think that numbers actually exist. Semantical realism is, however, a view about the nature of truth and meaning, though there are different claims regarding how realists and anti-realists differ. Some anti-realists view the issues between realism and its denial in terms of questions concerning the connection between meaning and use and a related question concerning the validity of bivalence for selected regions of conceptual space, while others think of the questions in terms of whether truth or meaning are tied in some way or other to what we are capable of knowing or verifying. The first group takes its clue from the work of Michael Dummett, whereas the second view has a long and rich philosophical tradition among those who have thought that alternative, realist conceptions of truth lead to skeptical consequences.

I will begin with a discussion of the anti-skeptical motivation and then turn to the "meaning as use" motivation. After treating these motivations, I will turn to other philosophical grounds for being attracted to the knowability claim, where I will argue that both physicalism and theism contain elements that incline such positions to endorse that claim.

Anti-Skeptical Motivations

A dominant concern of philosophy has always been the issue of skepticism, whether of a practical sort of the kind displayed by the Pyrrhonians or of the more theoretical sort characteristic of the modern period of the history of philosophy. There are various attempts to try to avoid skepticism, one of which is most important in the present context. Many philosophers since Descartes have thought that the problem of skepticism is rooted in an incorrect picture of the nature of reality and truth. The heart of this approach involves holding that ordinary understanding of the distinctions between mind and world, subject and object, internal and external, appearance and reality is the root of all skeptical evil, that the source of skepticism is not found in the discipline

of epistemology but rather in a misguided ontology or in a mistaken conception of truth. An early proponent of this view is George Berkeley:

> [W]e have been led into very dangerous errors, by supposing a two-fold existence of the objects of sense—the one *intelligible* or in the mind; the other *real* and without the mind, wherey unthinking things are thought to have a natural subsistence of their own, distinct from being perceived by spirits. This, which, if I mistake not, hath been shown to be the most groundless and absurd notion, is the very root of Scepticism . . . All this sceptical cant follows from our supposing a difference between *things* and *ideas*, and that the former had a subsistence without the mind or unperceived.[1]

Berkeley's position is that the root of skepticism is to be found in certain ontological commitments, and it is important to notice the natural affinity of these commitments with a denial of the claim that all truths are knowledge, and to see the affinity of denying these commitments with endorsing the claim that all truths are knowable. If there is a difference between things and ideas such that things have a subsistence without the mind, then it would seem to be possible for things to exist without any mental beings existing at all. Non-mental things could be pretty much just as they actually are without minds to know about them. Or, to put the point semantically, what is presently true of non-mental things could continue to be true of them even if there were no minds to know these truths. So, it would seem, the ontology of things and ideas against which Berkeley rails is one that would refuse to endorse any necessary connection between truth and knowability, for on this view truth does not require that there be knowers at all.

Much the same position about the root of skepticism is endorsed by John Dewey:

> Those who have followed the previous discussions will not be surprised to hear that, from the standpoint of experimental knowing, all the rivalries and connected problems grow from a single root. They spring from the assumption that the true and valid object of knowledge is that which has being prior to and independent of the operations of knowing.[2]

John Greco holds that this line of thinking pervades the branch of philosophy standardly referred to as Continental Philosophy:

> While this diagnosis of skepticism can be found in analytic philosophy, it is almost ubiquitous in the Continental tradition. In this regard it is helpful to understand

[1] George Berkeley, *A Treatise Concerning the Principles of Human Knowledge* (Chicago, 1986), pp. 107–108.

[2] John Dewey, *The Quest for Certainty* (New York, 1929), p. 196.

that much of Continental philosophy is essentially Kantian, beginning from the premise of Kant's Copernican revolution. Roughly, the line of thinking is that Hume's skepticism presupposes a distinction between mind and mind-independent reality. To avoid skepticism we must give up empirical realism in favor of transcendental idealism. In other words, we must give up the idea of a knowable mind-independent reality. On the basis of roughly this line of thought, many Continental philosophers now take it for granted that the object of knowledge is created (or at least shaped) by the mind that knows it, and Continental philosophy is largely devoted to exploring the implications of this idea.[3]

Greco is certainly correct to cite Kant in this regard, for Kant's Copernican revolution in philosophy has a rejection of such realist conceptions of truth as one of its cornerstones; for Kant, truth is something intersubjective, and hence phenomenal rather than noumenal. Moreover, however ubiquitous the view is or is not in Continental circles, there is a vast array of proponents of this view of the source of skepticism beyond those cited to this point. Richard Rorty, for example, not only endorses the view that the source of skepticism is bad ontology, but refers to this explanation as "the usual story" about skepticism, citing Etienne Gilson and John Herman Randall as proponents of this explanation.[4]

In our context, the importance of this story about skepticism is how it requires more intimacy between mind and world and hence between knowers and truth. If the world itself is shaped by our thinking, then the possibility of truth completely beyond our ken may come to be seen to be so remote as to disappear entirely. If truth becomes so in virtue of the activity of mentation itself, it becomes thoroughly mysterious how our minds could succeed in creating something completely beyond our grasp. Such thinking about skepticism thus results in endorsements of *epistemic* conceptions of truth, conceptions which constrain the range of truth by our cognitive capacities. Examples of such endorsements are multitude. Above, I noted Kant's endorsement of an intersubjective, phenomenal account of truth, and in more recent times, such epistemic conceptions of truth can be found both among classical American philosophers and their progeny, as well as among the Logical Positivists and their heirs. Charles Sanders Pierce, for example, understands truth in terms of what the scientific community would agree on in the long run; he says, "The opinion which is fated to be ultimately agreed to by all

[3] John Greco, *Putting Skeptics in Their Place* (New York, 2000), p. 78.

[4] Richard Rorty, *Philosophy and the Mirror of Nature* (Princeton, 1979), p. 49, n. 19, and pp. 51–52, n. 21.

who investigate is what we mean by truth",[5] and "Truth is that concordance of an abstract statement with the ideal limit towards which endless investigation would tend to bring scientific belief..."[6] William James, in certain moods, understands truth in terms of what is born out by experience. He says, "[I]deas ... become true just in so far as they help us to get into satisfactory relationship with other parts of our experience",[7] and "Those thoughts are true which guide us to *beneficial interaction* with sensible particulars as they occur, whether they copy these in advance or not."[8] John Dewey seems to hold that truth is to be understood in terms of warrant or verification: "Men have slowly grown accustomed in all specific beliefs to identifying the true with the verified. But they still hesitate to recognize the implication of this identification and to derive the definition of truth from it."[9] And Hilary Putnam in one of his myriad of commitments holds that truth is what would be verified under ideal epistemic conditions.[10] What is common across all these characterizations is the idea that truth simply cannot forever and always outstrip our capacity to uncover it. Such a realist conception of truth is rooted in the bad ontology which is the source of skepticism, and if such skepticism is to be avoided, a minimal first step is to endorse a theory of truth according to which truth is simply incapable of exceeding the range of possible knowledge.

Of course, it is a long way from the claim that all truths are at least in principle knowable to a denial of skepticism, to a denial of the claim that we never know anything. For even if all truths are in principle knowable, the falsity of skepticism does not follow but only the mere possibility of such. There are modally strong versions of skepticism whose falsity is undermined by the in principle knowability of truth; for example, a version that claims that knowledge is simply impossible is incompatible with anti-realism. Weaker versions of skepticism remain unscathed, however. Even so, this position on truth does allow a certain foothold against skepticism, for even if the knowability of all truth fails to refute skepticism, it does undermine attempts to argue for skepticism by

[5] Charles Sanders Peirce, *Collected Papers of Charles Sanders Peirce*, 8 vols, edited by Charles Hartshorne and Paul Weiss (vols. 1–6) and A. W. Burks (vols. 7–8) (Cambridge, Mass., 1931–1958), vol. 5, p. 407.

[6] Peirce, *Collected Papers*, vol. 5, p. 565.

[7] William James, *Pragmatism*, originally published in 1907 (Cambridge, Mass., 1975), p. 34.

[8] William James, *The Meaning of Truth* (Cambridge, Mass., 1909), p. 51.

[9] John Dewey, *Reconstruction in Philosophy* (Boston, 1957), p. 159.

[10] Hilary Putnam, *Reason, Truth, and History* (New York, 1981).

employing premises that imply the impossibility of knowledge (and such arguments are common, even the most common, in the history of epistemology). So jettisoning the bad ontology does have some benefits, since theoretical skeptical arguments generally do not imply that we lack knowledge in some merely contingent fashion, but that knowledge is impossible. For example, the regress argument in the hands of skeptics is intended to show that knowledge is impossible, as are the arguments from the fallibility of all our beliefs, the circularity of all warrant, the existence of alternative explanations of our experience, and our inability to answer the skeptic in a non-question-begging way. In one fell swoop, all these skeptical arguments can be eliminated if we jettison any ontology incompatible with the knowability of truth. So endorsing a restructured ontology with a commitment to the claim that all truths are knowable would seem to bear some significant philosophical fruit.

The "Meaning as Use" Motivation

Another motivation for thinking that all truths are knowable hinges on connections between meaning and truth, and the Wittgensteinian idea that ties meaning together with (competent) use. Suppose our untutored selves begin thinking about meaning and truth in such a way that the meaning of a claim involves its truth conditions, and that these conditions may forever exceed our capacity to discover them. In such a case, the truth conditions in question are objective, mind-independent aspects of the world, and our understanding of a language involves our knowing which mind-independent aspects constitute the meaning of a sentence, word, or, more generally, symbol.

Against such a theory, the anti-realist challenges our untutored selves to cite a feature of our use of the language that would confirm this picture. Presumably, our knowledge of the meaning of the language can be displayed in our use of the language, so if our knowledge of meaning is as above, we should be able to cite some way in which our use of the language would confirm that picture. The challenge for the untutored view is that the feature in question must count in favor of the truth-conditional view over against its primary rival, the verification-conditional view. On this latter view, meaning is understood in terms of the conditions under which the sentence in question would be verified. The relationship between meaning and use on this theory of meaning is best seen by generalizing from what might be said about simple connectives in propositional logic. To know the meaning of 'and', for

example, is simply to actually apply and be disposed to apply the appropriate introduction and elimination rules in one's use of that connective. For example, when one assents to "Today is Friday and it is raining", one who knows the meaning of 'and' will also be disposed to assent to "Today is Friday." Moreover, when one assents to "Today is Friday", and also assents to "It is raining", one who is competent in the language will also be disposed to assent to "Today is Friday and it is raining." Similar remarks can then be made about terms in general: to grasp the meaning of the terms is to apply and be disposed to apply it in the conditions which would verify the claim in question.

The story for verificationist theories of meaning will have to become more complicated than the above in order to explain the possibility of slips of the tongue and other misuses of language, not to mention problems of misperception and other factual errors, but we need not enter into the complications of the view for present purposes. Here, we need only note that the challenge to truth-conditional theories of meaning is contrastive in character. It is a challenge to provide evidence for the truth-conditional view that would not also be evidence for the verification-conditional view, for the complications just noted for the verification-conditional theory will be complications for the truth-conditional theory as well. That said, the contrastive nature of the challenge from verificationists is to produce evidence that meanings can be verification-transcendent.

The idea, then, is this. Our knowledge of the meanings of our sentences is normally not explicit knowledge, but rather implicit, revealed by our linguistic behavior. A theory of meaning is then a model of this implicit knowledge, and we judge the adequacy of the theory by how well the model predicts linguistic behavior. If the model predicts behavior that we never actually display, even when presented with an appropriate stimulus, then the model is inadequate. Moreover, if the model contains elements that do no explanatory work, and there is another model lacking those elements, then the simpler model is preferable.[11]

After setting up this criterion of adequacy, we then suppose that we have two theories of meaning to consider, one a truth-conditional one and the other a verification-conditional one. The truth-conditional theory posits implicit knowledge that the verification-conditional theory does not, so the truth-conditional theorist needs to find cases in which these extra elements do some explanatory work. The most plausible cases for such a theorist to focus on are cases in which we have undecidable

[11] For an argument along these lines, see Dag Prawitz, "Meaning and Proofs: On the Conflict Between Classical and Intuitionistic Logic", *Theoria* 43 (1977), pp. 2–40.

sentences—sentences for which no decision procedure exists (or could exist) for determining whether the sentence is true or false. After all, for decidable sentences, there will be no distinction between the truth conditions and the verification conditions, for the verification theorist can construct a decision procedure for a decidable sentence in terms of checking to see if the conditions cited by the truth-conditional theorist obtain. So the debate focuses quickly on undecidables, and in this crucial test case, the question is whether any linguistic behavior would manifest implicit knowledge of the truth conditions of such sentences, as opposed to their verification conditions.

Here is where the problem lies for truth-conditional semantics, for, it is argued, in order to manifest knowledge of a statement's truth conditions, one would have to find oneself in a position where one's linguistic behavior would manifest knowledge of the statement's truth-value.[12] If so, however, truth-conditional theorists will be at a loss to find a case in which the implicit knowledge of truth conditions posited by such a theory does any explanatory work whatsoever.

If we then reject truth-conditional semantics in favor of verification-conditional semantics, we thereby tie the concepts of meaning and truth more closely with those of verification and knowledge. If the meaning of a sentence is its verification conditions (and if we assume that all truths are expressible by meaningful sentences), then one would expect it to follow that all truths are verifiable, simply by checking each of the conditions of verification. Furthermore, the argument from verification to knowledge should be relatively simple, for it would be an awkward theory that implied that it is possible for a claim to be verified and yet unknown. So, anything verifiable should be knowable, and hence if meaning is to be understood in terms of verification, anything true should be knowable as well.

These two strands within the history of philosophy—the anti-skeptical strand and the verificationistic strand—provide the most obvious motivations for accepting the claim that all truths are knowable. These positions are controversial, however, and it is far from the truth to maintain that everyone should find here a motivation for endorsing the claim that all truths are knowable, even though these motivations are present in a very large percentage of philosophers. What is more

[12] Richard Kirkham gives this reconstruction of Michael Dummett's argument in *Theories of Truth* (Cambridge, Mass., 1995), 255–258, taken from Dummett's "The Philosophical Basis of Intuitionistic Logic", in *Truth and Other Enigmas* (Cambridge, Mass., 1978), pp. 215–247.

surprising is that there are other motivations as well for endorsing the claim that all truths are knowable. In particular, there are lines of thought from both physicalist and theistic assumptions in support of the claim that all truths are knowable.

Methodological Physicalism

The view I have in mind is alternately referred to as materialism, or naturalism, or physicalism. This viewpoint is very hard to characterize, but central to it is the denial that there are supernatural beings, such as God, Cartesian minds, or Leibnizian monads. When pressed to give a positive characterization of the view in question, defenders of the view can go in one of two directions. One way to go is to give a substantive characterization of the view. One such way adopts a Cartesian viewpoint about matter, claiming that matter is essentially extended, and that matter is all there is. Such a substantive characterization of materialism is problematic for several reasons. First, it is incompatible with the existence of empty space between material particles. Second, it is incompatible with contemporary physics. Third, it classifies philosophers in a somewhat embarrassing way. Hartry Field, for example, must be an immaterialist on this construal because he thinks space-time points are actual, concrete individuals; and he comes out as an immaterialist in spite of his explicit avowals of a desire to maintain the materialist viewpoint.[13] Perhaps the most embarrassing of these difficulties is the second, and it is an irremediable problem for substantive characterizations of the materialist perspective. For any substantive characterization of that perspective will be open to refutation by the future of science, and even if we adopt a more contemporary expression of the perspective in terms of the kinds of things contemporary physics posits, we still have fairly strong inductive evidence that such a view will also be refuted by some future physical theory.

There is no inconsistency, of course, in holding a view for which one has inductive grounds to reject. It may be that there is countervailing evidence that defeats these inductive grounds (such as the very grounds for accepting the theory on the basis of the empirical evidence for it). It may also be the case that the view in question is better than any known alternative, and one may hold a view on which it is rational to hold a view when it is the best view among the known competitors. Still, there is a bit

[13] Hartry Field, *Science Without Numbers* (Oxford, 1980).

of embarrassment to be addressed, for when the pessimistic induction reaches fruition through development of a new scientific theory that conflicts with one's substantive characterization of materialism, those who hold the materialist perspective do not conclude that materialism has been refuted. Instead, they typically maintain that the earlier expression of that perspective was mistaken about the real content of the view. The earlier view is an outmoded expression of the materialist perspective, one only retained by those not up to speed with advances in science.

This pattern of revision in one's understanding of materialism or physicalism strongly suggests that the viewpoint is not a substantive one at all but rather a methodological one. The methodological viewpoint holds that it is the sciences that define what kinds of things exist, and when science advances in a way that undermines our present conception of what there is, we should change our conception to accord with these advances. When asked about their view, such physicalists often sound as if they hold a substantive view, for they will often offer the present scientific conception of things. Once the pattern of reformulation is noticed, however, the best interpretation of the data is that they are methodological physicalists rather than substantive ones. The key datum here is whether such a physicalist views the advance of science as amounting to a refutation of their view. The presence of the pattern above is evidence that such physicalists are not disposed to treat scientific advances as capable of refuting their view, thereby confirming that their view is methodological rather than substantive. On such a methodological construal, present science may tell us what is at present rational to believe (if anything is rational to believe) about what is real, but it is not definitive of what there is.

Consider J. J. C. Smart's account of the commitments of materialism:

> By 'materialism' I mean the theory that there is nothing in the world over and above those entities which are postulated by physics (or, of course, those entities which will be postulated by future and more adequate physical theories).[14]

This proposal suffers from a number of inelegancies of expression. Does Smart mean to affirm the existence of everything that any physical theory posits from now till eternity, as his use of disjunction implies? What are we to make of the qualifier "more adequate"? If all there is, is what physics says there is, then it is hard to find room for that which

[14] J. J. C. Smart, "Materialism," in *Essays Metaphysical and Moral* (Oxford, 1987).

physics must explain, unless the physical theory in question first tells us what the data are and then explains them.

The intended idea, however, is not that hard to decipher. The idea is that nothing falls outside the domain of inquiry by physics, that everything that exists is (in principle, at least) discoverable by physics, that if we were able to construct the best possible physical theory, one subject to no anomalies whatsoever, that theory would be complete in the sense that nothing exists that the theory did not posit and nothing was true except what falls within the purview of the theory. Such a view does not tell us in substantive terms what is real, but it does tell us the method by which we can discover what is real, and for that reason, the best interpretation of the above quote takes Smart to be endorsing methodological physicalism.

It is important here to distinguish two quite different claims that might be intended by such an insistence that we trust physics, or science more generally, in this way. One claim is epistemological and the other metaphysical. The epistemological claim insists that the best way to think about ontology at any given time is in terms of the best science of the day. Such a claim is not a metaphysical claim, but rather an epistemological claim. It informs us about the most justified position to take on a metaphysical matter, but it is not itself a metaphysical claim. The metaphysical claim will have to identify what is real with what is discovered by the scientific method, at least at some ideal limit of inquiry. Both positions can maintain the view that the best science of any day is partial, incomplete, and defeasible by further evidence. To adopt a metaphysical position on the nature of what is real, however, some version of the metaphysical position will have to be adopted; merely endorsing the epistemological claim is not yet to have adopted any metaphysical position. Hence, to the extent that Smart is endorsing a metaphysical position, the best position to attribute to him is the view that what is real is defined in terms of what science is capable of discovering.

Such a viewpoint, it appears, maintains an anti-realist view of truth, and it strongly suggests that truth cannot outrun our best scientific theories. If this assessment is correct, the best scientific theories would have to be complete, yielding as theorems every truth whatsoever. Such theories might be only within the domain of physics, or for the less reductionistic, they might be within the domain of any of the sciences. But what is central appears to be that truth can never extend beyond what our best scientific theories can reach.

One might grant that methodological physicalism cannot tolerate truths about entities over which the best physical theory does not

quantify, and yet question whether the best theory needs to be complete. The idea behind this worry is that there might be truths not implied by the best physical theory that are nonetheless truths about objects within the scope of that theory. Such a possibility is easy to dismiss, however, for an adequate version of physicalism ought to require not only that the objects that exist fall within the purview of science, but also that all the properties and relations do so as well. For example, no version of physicalism should be comfortable allowing a dualism of mental and physical properties, allowing that the mental realm need not be explicable in terms of the physical realm.

One might also question the idea of there being a "best" physical theory. Instead, a more rational view is that there might be an unending sequence of theories, each an improvement on a prior theory. Methodological physicalism would then be committed to the idea that nothing falls outside the explanatory scope of this infinite sequence of theories. Though I have appealed to the idea of a "best" theory, there is no reason to suppose that there is one, and nothing in the above discussion requires that there be a best theory in order for methodological physicalism to have a stake in determining whether all truths are knowable. Instead of requiring that all truths are captured in terms of one single, grand unified theory of everything, the methodological physicalist can demand only that every truth is suitably encoded by some physical theory in a succession of theories each of which is an improvement on the prior theories in the sequence.

One might also resist the idea that physicalism is committed to the knowability assumption by leaving some place for skepticism. That is, a physicalist might endorse a methodological construal of the view and yet insist that nothing, or not much, can be known. This option is open to physicalists, but a dilemma still faces the view. We should want to know whether the theories at the limit of inquiry are assumed to be explanatorily complete or not. If they are not explanatorily complete, that will give us some reason to doubt that the metaphysical perspective in question is itself well motivated. For example, suppose the theories in question posit laws, but the laws themselves are unexplained and inexplicable. If so, then it might be argued that there is a property dualism involved in this metaphysical perspective that is incompatible with physicalism. Presumably, what would make the laws inexplicable would be something about their modal status, and so sentences or propositions stating nomically necessary laws would be different in kind from the rest. It would appear, in such a case, that the theories can explain everything

except nomic necessity itself, and thus that there is a property that falls outside the explanatory scope of physical theory. Yet, if nomic necessity itself need not be explicable by physical theory in order for physicalism to be defensible, why should that be acceptable when it is clearly not acceptable to maintain that mentality is not explicable?

Such questions might push the methodological physicalist to grant the claim that explanatory completeness is included in the idea of the limit of scientific theorizing, but still resist the conclusion that all truths are thereby knowable. That position is tenable, but it is not one that can escape the form of argument at the heart of the knowability paradox. As we have seen, the crucial elements necessary to Fitch's proof are concepts (operators) that imply truth (more carefully, concepts that are iterationally redundant) and distribute over conjunction. One such concept is the concept of explanation: if p is explained, p is true; and if the conjunction $p\&q$ is explained, both conjunct are individually explained as well. Hence, if one grants that all truths are explicable, one can mimic Fitch's proof to show that all truths are in fact explained. The derived claim is, however, implausible. More plausible it is to think that some truths will go forever unexplained, even if, had we been interested in explaining them, we could have.

Hence, there is reason to view physicalism as committed to the central assumption of the knowability paradox that all truths are knowable, or at least to analogues of it that are equally problematic. Substantive versions of physicalism can avoid that commitment, but they avoid this commitment at the expense of likelihood of truth if the pessimistic induction is to be believed. Whether substantive physicalists suffer embarrassment in the face of the pessimistic induction is, however, not the point here, for I intend only to point out the attraction of the methodological view as well as its affinity for the position that endorses the knowability of truth.

Theism

A final position that has grounds for claiming that all truths are knowable is classical theism, derived from religions in the Abrahamic tradition. The obvious basis concerns the doctrine of omniscience, according to which God knows all truths,[15] for if all truth is known, it is obvious that all truth is knowable. More can be said, however. Religions

[15] Some hold that omniscience requires only something weaker, namely, knowing all that can be known. For a defense of the standard view, see Jonathan L. Kvanvig, *The Possibility of an All-Knowing God* (London, 1986).

in this tradition hold that human beings are created in the image of God, one aspect of which is the capacity for knowledge. Once this aspect of the image is introduced, a question will arise as to why human beings are singled out as the unique bearers of the image in question, for human beings are not the only created beings who have the capacity for knowledge. In order to defend the uniqueness claim, theologians will be under some pressure to adopt the knowability of all truths for humans. Lower animals, it may be claimed, are always limited in what they can know, but once we reach the plane of human rationality, the limits are removed. Though it may not be possible for a mere human being to be omniscient, the removal of limitations suggests that no truth is intrinsically beyond the epistemic powers of the human mind, and it is this feature that grounds the cognitive image of God in human beings.

The important point to note is that theologians will be under pressure to find something more than a gradual enhancement of cognitive abilities when humans are considered in comparison with other animals capable of knowledge. Without something discontinuous in the relationship between human and non-human mentation, theists will find it difficult to sustain the view that human cognition is an expression of the image of God but non-human cognition is not. A quite natural way to sustain a claim of discontinuity here is to hold that all truths are knowable for human beings whereas only some are for nonhumans.

Even more pressure exists for Christians to adopt the knowability thesis. Traditional Christianity holds both that God is essentially omniscient and that Jesus was both divine and human. If the universal knowability claim were false for ordinary human beings, it would be necessarily false (since it is a modal claim), but then an essential property of humanity would conflict with an essential property of divinity. An individual's being divine would imply that all truths are knowable (and known) by that individual, and an individual's being human would imply that not all truths are knowable by that individual. How then could one individual be both divine and human?

ASSESSING THE GROUNDS FOR THE KNOWABILITY ASSUMPTION

What is the strength of these considerations on behalf of the knowability assumption? One way to address this question is to consider what positions fit naturally with a denial of the assumption. One such

position takes an evolutionary perspective on cognition. From a cognitive ethological perspective, one can find evolutionary development from species with limited cognitive abilities up through more advanced cognitive abilities. For each such species, however, it is fairly obvious that there are truths exceeding their cognitive capacities. Dogs, for example, may know who their master is or that they are about to be fed, but they cannot understand Fermat's last theorem. They are simply not far enough along the evolutionary progression for that. Once we get to human beings, however, there is no reason to say anything more than that they can grasp many truths not able to be grasped by animals with lesser cognitive abilities. Nothing about the discipline of cognitive ethology could give us reason to think that once human beings have arrived on the scene, all limitations on mental capacities are superseded. It is, after all, only a three-pound brain.

If one is inclined toward such a perspective that fits naturally with denying that all truths are knowable, one may be faced with the cognitive dissonance created by holding other positions that provide grounds for believing that all truths are knowable. For example, the combination of a commitment to evolutionary theory and methodological physicalism is a quite common one, but these two positions pull in opposite directions on the question of the knowability assumption.

If we look more directly at the grounds cited above for endorsing the knowability assumption, probably the least convincing is the theistic argument. First, it is weak because it relies on a special aspect of one kind of theistic view, the aspect appealing to the image of God in human beings. Such a humanistic version of theism is the most common form in western culture, but it is surely not the only possible version of theism. So the tie between theism and knowability is somewhat loose simply because only for the most popular version of the view can we find any such attachment. Moreover, even for humanistic theism, the argument for endorsing the knowability claim is far from compelling. One might also hold that, on the cognitive side, the image of God is displayed by a capacity to grasp certain kinds of truths that only those in whom the *imago Dei* resides can know. Perhaps the kinds in question are moral truths, or perhaps spiritual truths. But there is no compelling reason to think that the more general capacity to know any truth is required as an expression of this image. Finally, though the attraction of the knowability claim is strongest for traditional Christianity because of the doctrine of the Incarnation, the history of Christian philosophy is filled with sophisticated (and sophistical) attempts to avoid apparent

implications of just the sort noted above. For one example, there is the venerable tradition of appealing to the *qua* aspect to block apparent contradictions. In this tradition, one can say that Jesus *qua* divine was omniscient, but *qua* human was not. In my opinion, this maneuver is best thought of as sophistical rather than sophisticated, but it would be premature to pronounce at this point that no such maneuver could successfully free traditional Christianity from commitment to the knowability claim.

Second, the motivation for the knowability assumption that derives from considerations in the theory of meaning must answer strong rebuttals. These motivations land their defenders on a path well worn by those with Logical Positivist inclinations in the early decades of the twentieth century. As Michael Dummett argues, the manifestation argument cited earlier leads to a theory of meaning on which the key notions are either verification or falsification,[16] mirroring in terminology, at least, a view that one would think has long been discredited. Seen from this perspective (a perspective, of course, which the anti-realist will claim is misleading), the "meaning as use" idea seems to be just a different way of jumping on a train headed for disaster.

Finally, there is some room for physicalism to avoid a commitment to the knowability claim. The most obvious point is that a physicalist may bite the bullet and retain a substantive characterization of the view, insisting that if the future of science shows that view mistaken, then they will admit that physicalism is false. Furthermore, even methodological physicalists may be able to avoid the knowability claim. For the methodological physicalist may appeal to a type/token distinction, maintaining that physicalism is true because every type of event or aspect of reality can be successfully investigated by scientific methods. Such a viewpoint would require that every type of truth is knowable, a claim logically weaker than the claim that all truths are knowable. It must be admitted, however, that this ground for avoiding the knowability claim is a bit strained. In order to avoid the knowability claim, such a physicalist would have to maintain that some truths are unknowable even though they fall within a type that is knowable. It is hard to see what grounds there could be for such a distinction, independent of the derivation of Fitch's result. Furthermore, if Fitch's result is the only motive for the distinction, a better position to hold is that

[16] See especially, "What is a Theory of Meaning II", in G. Evans and J. McDowell (eds.), *Truth and Meaning* (Oxford, 1999).

methodological physicalism has a stake in the paradox in virtue of the fact that the knowability assumption fits better with the view than does its denial. Perhaps a better path for the physicalist to follow is to deny the explicability and knowability of either the initial conditions of the universe or the laws that govern transitions from one state to another. There are some grounds for concern here, as already noted, but perhaps these grounds can be overcome. In any case, a final possibility for the physicalist rests on the apparent fundamental indeterminacy that might still be found in physical theories at the limit of inquiry. If such irreducible indeterminacy remains, it is plausible to maintain that events will happen for which there can be no explanation (even though the likelihood of such events will be explicable). Armed with such a position, a physicalist may deny that all truths are knowable and that all truths are explicable.

The last motivation is the anti-skeptical one, and even though adopting an anti-realist theory of truth blocks certain skeptical arguments, such a position cannot resolve skeptical worries. Skepticism is rebutted only if there are some things that we know, and endorsing the claim that all truths are knowable is simply irrelevant to that issue. The knowability claim is neither necessary nor sufficient for having an adequate response to skepticism, so if one wishes for an adequate response to the problem of skepticism, one had better engage the epistemological assumptions of the problem directly, rather than hoping to find some mistake in other areas of philosophy, such as metaphysics or philosophy of mind, that leads to skeptical problems.[17]

In all, it must be admitted that the case on behalf of the claim that all truths are knowable is not sufficient to undergird the claim of paradoxicality for the proof that originated in Fitch's 1963 paper. For that to be true, the epistemic credentials of the knowability claim would have to be stronger than they in fact are. For those inclined toward the knowability claim, the proof presents a serious problem, but general paradoxicality is not supplied solely on the basis of the plausibility of that claim. I think, however, that the proof signals paradoxicality, even though the paradox is not simply that the knowability of all truth is inconsistent with the modest assumption that some truths are not known.

[17] This point is the thesis of Greco's examination of the variety of positions that try to find non-epistemological mistakes that are responsible for the power of skepticism. See his *Putting Skeptics in Their Place*.

THE SOURCE OF PARADOXICALITY

The first thing to note about the paradox is the surprise it contains. Even if a person were not inclined to think that all truths are knowable, could it really be the downfall of this view that some truths are not known? The first is a modal claim, the second non-modal; the first asserts the mere possibility of that which the second denies as a matter of contingent fact. We would object to such inferences in ordinary cases. The existence of unredeemed chits does not show that some chits are unredeemable, nor does the existence of criminals who never change their ways show that any criminals are incorrigible.

It is the surprising character of this inference that provides the source of the paradoxicality of the problem. We begin with two claims, the first a modal claim and the second non-modal:

$$p \rightarrow \Diamond Kp \tag{1}$$

$$p \rightarrow Kp. \tag{2}$$

These two claims are related logically, for it is an utterly pedestrian proof to show that (2) entails (1), a proof needing only the modal principle that what is actual is possible. This logical relationship is not surprising, but when we combine it with the proof from the last chapter, we end up with a theorem that ought to leave us incredulous:

$$\vdash (p \rightarrow \Diamond Kp) \leftrightarrow (p \rightarrow Kp).$$

That is, we end up having to endorse the claim that all truths are knowable is provably equivalent to the omniscience-like claim that all truths are known.

I maintain that it is crucial to a proper understanding of the paradox of knowability to clarify it in terms of the above theorem and the inter-derivability claim, and I will expend considerable effort later on arguing for this view. A bit of foreshadowing might help here, however, so let me suggest briefly why this particular characterization of the paradox is crucial.

The history of thought about Fitch's proof has focused on the point that if the proof is sound, a particular version of anti-realism, according to which the universal knowability assumption of Fitch's proof is correct, is false. On this picture, it is easy to construe the paradox as involving an unexpected implication from certain anti-realist commitments about truth.

My view, however, is that what is paradoxical here is not tied in any way at all to a commitment to anti-realism. The above characterization of the central perplexing claims that constitute the heart of the paradox yields precisely this result. The logical result in question requires no anti-realist commitments to prove, since it is a theorem; and denying anti-realist commitments does not eliminate whatever paradoxicality is involved in the equivalence formulated.

In short, it is crucial to the defense of my approach to the paradox to see it in terms of the theorem above and the logical equivalence it presents. The paradox is not a local problem for anti-realism, it is a global problem that affects everyone. The reason it is a global problem is that the claim above specifies a context in which the distinction between actuality and possibility utterly disappears. A response of complacency about whether all truths are knowable cannot eliminate the perplexity engendered by this result.

My claim, then, is that the proof first appearing in Fitch's 1963 paper yields a paradox, but it is not a paradox because it threatens the idea that all truths are knowable. Instead, what is paradoxical is that the proof can be used to show that there is no logical distinction whatsoever between universally known truths and universally knowable truths.

It is important as well to note that, although it is appropriate to characterize Fitch's proof as surprising, the characterization given above involving a collapse in a certain context of the distinction between actuality and possibility is more than surprising. The distinction between what is paradoxical and what is merely surprising is not easy to clarify, but the distinction between what is actual and what is possible is so deeply rooted in our conception of things that it goes beyond a merely surprising result when this distinction is threatened. Upon first seeing Fitch's proof, a natural response is to see it as a novel and surprising refutation of verificationism and its cousins. Deeper inspection reveals, however, that it is not simply that, since the proof forms the core element in establishing a logical equivalence that threatens some of our deepest modal convictions.

I am not claiming here that the paradox is unresolvable, that there is no acceptable way to explain how such a collapse can be tolerated. After all, we know of other contexts in which such a collapse occurs: in one quite popular modal system, there is no logical distinction between actual necessity and possible necessity, for example. It is worth noticing, however, that the loss of this distinction is made palatable by a semantic explanation of the loss, in terms of possible worlds and accessibility

relations among them. Perhaps something similar can be done with the knowability, or perhaps there is some other way to relieve our discomfort at the lost distinction resulting from Fitch's proof. The point I want to insist on is the need for some such explanation, a need arising from the fact that the result above is not merely surprising but is, instead, paradoxical.

Hence, I submit, there is a paradox here, and its source is found in an identification that I will speak of using the phrases 'universally knowable truths' and 'universally known truths'. It is important not to misunderstand my language here. By 'universally known truths', I do not mean truths that are known by everyone, but rather that every truth falls within the class of things known, i.e., the universe of truths is included in the set of things known. Given this understanding of my terminology, the source of the paradox is in the logical relationship that is claimed to exist between universally known and universally knowable truths. Appreciation of the paradox is aided by seeing some plausibility in the idea that all truths are knowable, but the paradox is not simply a matter of an untenable conclusion following from a platitude, either ordinary or philosophical. Though Fitch's proof may provide an argument against verificationism and other philosophical positions that maintain some variant of the knowability claim, the paradoxicality of the proof is found elsewhere. The proof undermines the truism that there is a logical distinction between the actual and the possible, examples of which can be found in many areas, including the subject matter of epistemology. There is a distinction between what is confirmed and what can be confirmed, between what is justified and what might be justified, and there is also such a distinction, we should expect, between universally known truth and universally knowable truth.

There is a further point to note here. Let global anti-realism be the view that endorses the claim that all truths are knowable. Global anti-realism implies something stronger than what I have claimed above, for it implies that there is a lost logical distinction between possible known truth and actual known truth, not just a lost distinction between possible universally known truth and actual universally known truth. That is, if global anti-realism is true (and Fitch's proof goes unanswered), then we can demonstrate that any claim is knowable if and only if it is known.

This implication of global anti-realism is related to, but distinct from, the logical distinction cited above. It is one thing to claim that the proposition *all truths are knowable* is logically equivalent to the claim *all truths are known*; it is quite a different thing to claim that, for any value

for *p*, *p* is a knowable truth if and only if *p* is a known truth. The second is an implication of global anti-realism, if Fitch's proof is sound, but I deny that it is the heart of the knowability paradox. The heart of the paradox is about a lost logical distinction between actuality and possibility. Global anti-realism implies one such lost distinction, but it is not the only lost distinction in the conceptual space in question. There is also the lost logical distinction between universally knowable truth and universally known truth, and that distinction is lost to everyone on the basis of Fitch's proof, even to those who find anti-realism unattractive.

CONCLUSION

We have seen a variety of grounds for accepting the claim that all truths are knowable, the assumption that is central to the derivation of Fitch's result. None of these grounds are compelling, but it is nonetheless important to see some of the attraction the claim holds for a variety of positions, for as we saw in the last chapter, derivations of a Fitch result are not paradoxical when they begin from premises that are obviously necessarily false.

In the end, however, the paradox should be felt by everyone, even those who do not accept the knowability claim, for the heart of the paradox is not simply in what is implied by the knowability claim but rather in a lost logical distinction between what is actual and what is possible. Though this characterization of the paradox may seem prosaic at this point, we will find that it has remarkable implications regarding the direction the literature on the paradox has taken. We will find that a great deal of ink has been spilt addressing the paradox in a way that is completely futile for failing to appreciate where paradox lies.

3

Syntactic Restriction Strategies

The central proof involved in the knowability paradox begins with a commitment to

$$\forall p(p \rightarrow \Diamond \mathrm{K}p). \tag{1}$$

For many, (1) is the quintessential implication of semantic anti-realism, and for some, a constitutive claim of the position.

In the face of the paradox, however, some have sought to free anti-realism of a commitment to (1) in order to avoid a refutation of anti-realism by the knowability paradox. Such a move is a bit suspicious, prompted as it is by the paradox itself, rather than by some independent investigation of the position of anti-realism, but I do not wish to press that point here. For it is often the case that deeper understanding of a philosophical position is occasioned by some forceful argument against it so, whatever the cause of the reflection on the nature of anti-realism, the question of its true commitments is always relevant.

In this chapter, I will look at four attempts to argue that anti-realism is not committed to (1). After doing so, we will be in a position to draw some general conclusions about the plausibility of such attempts and their relevance to the paradox. We will see that restrictions on (1) are not relevant to the knowability paradox, though they may allow anti-realism to avoid refutation by Fitch's proof. The key to understanding this point is to see that Fitch's proof is central to the paradox, but that the paradoxicality involved should not be identified with Fitch's proof itself.

It is also worth noting that the extant attempts to save anti-realism by restricting (1) face serious obstacles. As noted already, the most obvious worry is that such proposals will be *ad hoc*, but there are other difficulties that such proposals face as well. I begin, then, by examining the plausibility of restricting (1), leaving until the end of the chapter an explanation of why such restrictions won't help with the paradox.

EDGINGTON'S PROPOSAL

The first proposal to relieve anti-realism of a commitment to (1) is by Dorothy Edgington.[1] She defends verificationism against the challenge presented by the paradox by arguing that, at least for actual truths, truths involving an actuality operator, the paradox does not arise. Thus, she claims, the assumption that all actual truths are knowable and the assumption that some actual truth is not known are consistent. Moreover, she argues, there is no reason to saddle verification with the stronger commitment that all truths are knowable, and hence there is at least one form of anti-realism that escapes refutation by the proof of Fitch's result.

To see how this solution works, we first adopt the following truth condition for the actuality operator @:

$'@p'$ is true at any world w iff $'p'$ is true at @, the actual world.

An instance of the second assumption that some actual truths are not known is:

$$@(p \;\&\; \sim \mathrm{K}p), \tag{2}$$

which together with an appropriate representation of the first assumption yields:

$$\Diamond \mathrm{K}@(p \;\&\; \sim \mathrm{K}p). \tag{3}$$

This claim, in possible worlds semantics, asserts that there is some world i in which it is known that it is true in the actual world that p is true but not known. That claim is fairly obviously consistent, in much the same way that it is consistent for me to know that there is a truth that is unknown to you.

There are points to complain about in this proposal centering on what to make of the claim that one can know from the vantage point of one world what is true in another world. In particular, it is questionable whether it is possible to entertain any proposition about some singular, individual world not identical to the world one is in. Any way of describing the world will apply equally to a number of different worlds, and any sort of direct reference to other worlds would seem to be

[1] Dorothy Edgington, "The Paradox of Knowability", *Mind* 94.376 (October 1985), pp. 557–568.

impossible in virtue of the lack of any causal links between possible worlds.[2]

For this reason Edgington proposes to talk of possible situations rather than possible worlds, where a situation need not be complete as is the case for possible worlds.[3] This maneuver is not wholly successful, as Timothy Williamson has argued.[4] In general terms, the problem is one of how to get the content of a thought to be *about* a merely possible situation. There are two ways to do so, either by using a referring expression or by a description. The first way is difficult to sustain, given that there could be no causal link between one situation and another merely possible one. So it looks as if the contents of thought about a merely possible situation *s* will need to be fixed by description. If it is fixed by description, then the proposition *In s, p* will need to be necessarily true since *s* needs to be the actual situation (given the demand of finding a way to know from the perspective of another world what is true in the actual world), a situation whose identity conditions include the truth of *p*. Once we see the need for *in s, p* to be a necessary truth, it will be difficult to find a way to give a descriptive specification of *s* without yielding the result that the claim in question is nothing more than a trivial logical truth. Suppose, for example, that *q* is a set of necessary and sufficient conditions specifying *s*. Then so is *p*&*q*, given the necessary truth of *in s, p*. So it would seem that the descriptive specification of *s* could be either *necessarily, if q then p* or *necessarily, if p*&*q, then p*. Edgington needs to find a way to disallow the second, for the knowledge of what is going on in a different situation from one presently occupied is not knowledge of a trivial logical truth.[5]

Moreover, Edgington's view leaves non-actual knowledge of what is actually the case a mystery. She considers what she terms an "unproblematic" case of unactualized possible knowledge, knowledge that it was raining here at midnight (assuming that it was raining here at midnight, but that no one knew it), and she argues as follows:

> To satisfy (1') [$@p \rightarrow \Diamond K@p$], the possible knower that it rained last night has to know that it still would have rained, had neither he nor anyone else noticed. And

[2] For details on these complaints, see Timothy Williamson, "On the Paradox of Knowability", *Mind* 96.382 (April 1987), pp. 256–261.

[3] Dorothy Edgington, "The Paradox of Knowability", p. 564 and p. 566.

[4] Williamson, "On the Paradox of Knowability".

[5] Williamson gives an argument along these lines aimed at the stronger conclusion that the situation is hopeless for Edgington; see "On the Paradox of Knowability", pp. 258–259. I don't think his argument is decisive, though it creates just the problems for Edgington noted in the text.

any possible knower that it rained does, or at least can, know this . . . Our possible observer, if asked 'Would it still have rained, had neither you nor anyone else observed the rain?' would surely answer 'Of course'. If there are truths which fail to satisfy the principle

p would still have been true, had no one known that *p*, that I am in pain, for example, then they satisfy the principle,

If *p*, then someone knows that *p* and, *a fortiori*, they satisfy (1'): $@p \rightarrow \Diamond K@p$.[6]

What is mysterious is how the information cited supports the conclusion drawn, that (1') is satisfied in this case. To see the problem, let @ be the actual situation (in which it rained here last night, but no one knows that), and let *s* be the situation in which our possible knower *K* is supposed to know something which constitutes knowledge that in @, it is an unknown truth that it rained here last night. I think Edgington holds the following:

> *K* knows that it rained last night (i.e., knows that , in *s*, the situation that *K* occupies, it rained last night). (1)

With this information as background, she imagines that *K* has some further knowledge:

> *K* knows that it still would have rained last night even if no one knew that it had rained last night. (2)

She further holds that these two pieces of information constitute knowledge that in @, it is an unknown truth that it rained last night, since her second option above does not apply in this case. For the second option to apply, it would need to be true that if it rained last night, someone knows that it did, and Edgington's example of being in pain shows that she intends this conditional to be a logical, rather than material, conditional. Since the claim that it rained last night provides no logical guarantee that someone knows that it rained last night, she must maintain that (1) and (2) imply knowledge that in @, it is an unknown truth that it rained last night.

The mystery is how we got from knowledge about *s*, which is what the knowledge recorded in (1) and (2) is knowledge about, to knowledge

[6] Edgington, "The Paradox of Knowability", p. 567. Edgington uses 'A' to represent the actuality operator, and I have altered her text to conform to my preference for using '@'.

about @. The only way the answer to this puzzle can go is to hold that the knowledge described in (2) is, somehow, knowledge about @. How could that be? Since the knowledge described in (2) is knowledge of the truth of a counterfactual, perhaps Edgington is thinking as follows:

> @ is a situation very similar to *s*, with the only difference between them being that the knowledge described in (1) obtains in *s*, but not in @; so if we are in *s*, and imagine a situation in which the knowledge described in (1) is absent, the situation we will be imagining is @; and the consequent of the counterfactual knowledge described in (2) will then constitute knowledge that, in @, it rained last night.

This argument makes sense out of the reasoning in the passage above, but it is fallacious, for it assumes that @ is the unique, closest situation to *s* in which no one knows that it rained last night. (Note that it is not enough for Edgington to claim that there is no closer situation to *s* than @, for then there is no entitlement to the conclusion that in thinking about the knowledge described in (1), we are thinking about @.) As is well known, the standard semantics for counterfactuals does not presume that there is a closest world or situation,[7] but instead allows that there might be ties between worlds or situations in terms of closeness. Without the assumption that there is a closest world, I see no way for Edgington's proposal to give an explanation of how the two pieces of knowledge in *s* described above will yield knowledge about what is the case in @.

These problems are serious ones that cannot be handled easily, but there are other issues with Edgington's proposal as well. In our context, there is one way of taking her proposal that makes it irrelevant to the issue of whether various positions on truth and meaning can escape the clutches of the knowability paradox. One might propose Edgington's approach as a *retrenchment* position to take so that some remnant of fully general anti-realism or verificationism can be saved. Such a restricted position escapes the paradox, but does nothing to rescue anti-realism and other threatened positions that make claims about *all* truth, not just truths involving an actuality operator.

The other way to take the proposal is as a *clarification* of the true nature of anti-realism, a nature misunderstood by those who require it to include a commitment to the claim that all truths are knowable. There is evidence that Edgington herself takes her solution in this way. She first

[7] See David Lewis, *Counterfactuals* (Oxford, 1973).

points out that the paradox does not arise when various indexical devices invade the assumptions. For example, suppose one tried to develop the paradox when someone claimed that all truths are knowable, but that not all truths are known *now*, or not known *by us*. In both cases, the indexical character of the qualifiers is ignored: there is no contradiction in your knowing that some claim is true of which I am ignorant, nor is there any contradiction in our knowing things of which ancient peoples were ignorant. Edgington then extends this discussion to the knowability paradox itself. She says,

> I shall argue that an analogous solution can be given to the original paradox, in terms of the modal operator 'actually' . . .
>
> (2) [the second assumption of the paradox, that some truths are unknown] says that there is something which is actually true and actually not known to be true. Let us symbolize an arbitrary instance:
>
> $$@(p \;\&\; \sim \mathrm{K}p).^{8} \qquad (2^{*})$$

Edgington proposes here a fully general solution to the knowability paradox, one which attempts a careful clarification of the commitments of the anti-realist. She does not claim that her solution is a retrenchment aimed at saving some aspect of the view. Instead, she claims that the proper representation of the claim that some truths are not known is

$$\exists p@(p \;\&\; \sim \mathrm{K}p) \qquad (2^{*})$$

rather than

$$\exists p(p \;\&\; \sim \mathrm{K}p); \qquad (2)$$

she says explicitly that the second assumption "says that there is something which is actually true and actually not known to be true".

Here we must demur. Someone might have stated the second assumption, *intending* or *meaning* to assert (2*) rather than (2). But it is utterly obvious that (2) is the proper representation of "some truths are not known" and (2*) is not, but is instead the proper representation of "some claim is actually both truth and unknown".

It is easy to get confused here into thinking that (2) and (2*) say the same thing by failing to distinguish the truth conditions of a sentence from its meaning. Both (2*) and (2) have the same truth conditions, but then again, so do "snow is white" and "snow is white and it is raining or it isn't". Truth conditions alone do not determine meaning (though

[8] Edgington, "The Paradox of Knowability", p. 562.

there may be a way of defining some subclass of truth conditions which does determine meaning). To claim that "some truths are not known" says the same thing as (2*) makes the same mistake as to claim that "snow is white" says the same thing as "snow is white and either it is raining or it isn't". Having the same truth conditions is simply a different thing from saying the same thing or having the same semantic content.

This point connects in an important way with another problem Williamson raises for Edgington's view. This problem arises from the fact that the logical behavior of the actuality operator results in any claim of the form '@p' being either necessarily true or necessarily false (depending only on the truth-value of p). As a result, Edgington's proposal maintains that anti-realism is committed only to the knowability of necessary truths, and it is very hard to imagine that anti-realism is a view that does not require any contingent truths to be knowable.[9]

In the end, Edgington fails to carry through on the idea that a solution to the paradox can be modeled on solutions to similar arguments formulated in terms of *my* knowing all truths and all truths being known *now*. She is correct that the modal analogue of these indexicals is expressible using an actuality operator, but there are deep problems trying to explain how this approach to the paradox can save anti-realism.

TENNANT'S APPROACH

A second restriction strategy is pursued by Neil Tennant. Tennant does not think that Fitch's proof, even if granted, forces one to abandon anti-realism. In order to defend this view, Tennant must qualify claim (1) in some way or other, but not just any qualification will do. Some qualifications are obviously *ad hoc*; for instance, simply excluding from the range of (1) the truths that lead to problems. To address the paradox in a philosophically substantive way, one must go beyond such arbitrary approaches. Realists attempt to do this by observing that truth is 'radically nonepistemic',[10] thereby giving themselves a reason based on their conception of truth for denying (1). Tennant must do something comparable. We should expect him to find some feature of truth, anti-realistically conceived, that disarms the threat to anti-realism by allowing some truths to be unknowable. We will find, however, that Tennant fails to do so.

[9] Williamson, "On the Paradox of Knowability", p. 257.

[10] Hilary Putnam, "Reference and Truth", in *Realism and Reason: Philosophical Papers, Volume 3* New York, pp. 69–86.

Tennant proposes to deny (1), and submits (3) in its stead,

$$p \;\&\; Cp \vdash \Diamond \mathrm{K}p, \quad (3)$$

where '*Cp*' indicates that *p* is *Cartesian*. The notion of a Cartesian proposition is understood in terms of consistency. To be Cartesian, a proposition must be such that the assumption that it is known is consistent. Where K is an operator meaning *it is known by someone at some time that*..., a proposition fails to be Cartesian (is *anti-Cartesian*) precisely when $\mathrm{K}p \vdash \bot$.[12] So, *p* is Cartesian just when $\nvdash \sim \mathrm{K}p$.[13]

Tennant gives three ways that a proposition *p* can be anti-Cartesian. For one thing, *p* might itself be inconsistent. If not, then it may be the sort of proposition Descartes considered, such as *No thinking thing exists*. These propositions are consistent but *existentially inconsistent*[14]—presumably they could be true, but if true they cannot be the object of propositional attitudes. Finally, there are claims like $p\&\sim\mathrm{K}p$; the assumption that *p* is known is inconsistent because of the iteration of knowledge operators required to make the assumption.[15]

There are two problems I want to pursue regarding this proposal. First, there is an argument from Timothy Williamson that the paradox can arise anew even given Tennant's restricted claim. Second, there is the concern that Tennant's proposal is hopelessly *ad hoc*. I will pursue each of these points in turn.

Williamson's Objection

Williamson has argued that Tennant's proposal is subject to a variation of the knowability paradox.[16] His argument uses a sentence of the form:

$$p \;\&\; (\mathrm{K}p \rightarrow En),$$

[11] Tennant's own formulation in *The Taming of the True* is, p. 274, as follows: ϕ; ergo $\Diamond\mathrm{K}\phi$, where ϕ is Cartesian.

[12] The falsity constant, $\bot$, is one that is always assigned the value FALSE, and has as a content one that is never true under any circumstances. Some theorists use it to express the idea of an absurd proposition, or a self-contradictory one.

[13] This characterization assumes the usual negation–introduction rule $\Gamma, p \vdash \bot \Rightarrow \Gamma \vdash \sim p$. Nothing in subsequent discussion hinges on this.

[14] The idea of existential inconsistency is due to Jaakko Hintikka, "*Cogito Ergo Sum*: Inference or Performance?" *Philosophical Review* 72 (1961), pp. 3–32.

[15] Tennant uses a distribution principle of epistemic logic that is not universally accepted, $\mathrm{K}(p\&q) \vdash \mathrm{K}p$, though he does not defend the rule at any length, siding with our results from the last chapter.

[16] Timothy Williamson, "Tennant on Knowable Truth", *Ratio* 13.2 (June 2000), pp. 99–114.

where *p* is some decidable sentence such as "there is a piece of Roman pottery at such-and-such a location", '*E*' is the predicate "is even", and '*n*' is a name that rigidly designates the number of books on a particular shelf.

There are some points to note before examining Williamson's argument. In Williamson's discussion, some stress is placed on the role of the definite description "the number of books on that shelf," leading one to wonder what role the description plays in the objection. To make sure that no confusion results, it is important to note that the description can play one of two roles: it can either be the content of the term 'n' or it can fix the reference of the term 'n', where the content of the term thus becomes the number picked out by the description. In the former case, the sentence has the same content as one where 'n' is replaced with the description in question, but in the second case, the description is only the means by which the term 'n' used has its referent fixed (and thus is not part of the content of the proposition expressed by sentences containing 'n'). In the former case, the proposition will be a contingent one, and in the latter case a necessary one.[17] It is clear that Williamson intends the latter, and it is important to note that the description plays no role in the example when the example is construed this way. In particular, the use of the description in the example does not imply that the proposition is contingent nor does it imply that knowledge of the truth of that proposition is a posteriori. When a description is used to fix the referent of a name, the use of the name takes on a life of its own, independent of the description; the description does not semantically or epistemically infect propositions expressed by the use of the name. So, the proposition expressed is simply the mathematical claim that a certain number is even, even though the name of that number acquires its referential capacity in virtue of an empirical description. So the first point to note is that *En* is either necessarily true or necessarily false in the usual fashion for mathematical claims.

But the notion of consistency Tennant uses in defining a Cartesian proposition should not be identified with possibility, on pain of trivializing his proposal (we shall discuss later whether it is possible for it to avoid such trivialization, but for now we will assume that it can). So our initial assumption should be that there are impossibilities that are consistent, and consistencies that are impossible. If so, however, our

[17] For reasons deriving from Saul Kripke's *Naming and Necessity* (Cambridge, Mass. 1982). Both parties to the dispute agree with this background assumption, so I will not pursue the arguments for it here.

assumption ought to be that the above proposition is Cartesian. For p is decidable, and so is En (we may assume that the shelf is fairly small, so that all one needs to do is count the books), and the conjunction above is thereby decidable simply by applying the decision procedures for p and En. Tennant holds that if p is decidable, then so is Kp,[18] so he cannot object to the Cartesian status of Williamson's claim without undermining his argument.

Given these preliminaries, here is how Williamson's argument proceeds. Williamson's strategy is to show a contentious result (CR):

$$(\mathrm{CR})\ p\ \&\ {\sim}\mathrm{K}p \vdash En,$$

for he thinks that if he can show this, the non-contingency of En will allow him also to be able to prove an equally contentious result:

$$(\mathrm{ECR})\ p\ \&\ {\sim}\mathrm{K}p \vdash {\sim}\ En.$$

Since the premise is obviously consistent and possible for some values of p, these two sequents will demonstrate a knowability-like paradox, and Williamson thinks he can prove these sequents without employing any non-Cartesian premises.

Here is how the proof goes. By (3) above, Tennant's restricted knowability claim according to which all Cartesian truths are knowable, we get:

$$p\ \&\ (\mathrm{K}p \rightarrow En) \vdash \Diamond\mathrm{K}(p\ \&\ (\mathrm{K}p \rightarrow En)). \tag{4}$$

Since ${\sim}p \vdash p \rightarrow q$, we can prove:

$$p\ \&\ {\sim}\mathrm{K}p \vdash p\ \&\ (\mathrm{K}p \rightarrow En). \tag{5}$$

(4) and (5) together allow us to prove:

$$p\ \&\ {\sim}\mathrm{K}p \vdash \Diamond\mathrm{K}(p\ \&\ (\mathrm{K}p \rightarrow En)). \tag{6}$$

Williamson then attempts to show that

$$\mathrm{K}(p\ \&\ (\mathrm{K}p \rightarrow En)) \vdash En. \tag{7}$$

If he can show that (7) is correct, then by the modal principle $p \rightarrow q \vdash \Diamond p \rightarrow \Diamond q$, he can derive

$$\Diamond\mathrm{K}(p\ \&\ (\mathrm{K}p \rightarrow En)) \vdash \Diamond En, \tag{8}$$

which together with (4) by hypothetical syllogism gives us

$$p\ \&\ {\sim}\mathrm{K}p \vdash \Diamond En. \tag{9}$$

[18] Tennant, *The Taming of the True*, p. 262.

The proof finishes by attempting to derive (CR) from (9). Once this derivation is complete, we can start the proof all over, this time with $\sim E$ in place of E, yielding a final inference to (ECR) in place of (CR), giving us a different knowability-style paradox that is not solved by Tennant's restriction.

So the keys to the proof are these: first, the proof of (7) above, and second, the proof of (CR) from (9). The proof of (7) goes as follows: suppose the premise of (7) ($\mathrm{K}(p \,\&\, (\mathrm{K}p \rightarrow En))$) implies two claims—$\mathrm{K}p$ and $\mathrm{K}p \rightarrow En$; then that antecedent implies En by a simple application of modus ponens. Furthermore, the premise of (7) does imply both of the claims. It implies $\mathrm{K}p$ using the distribution principle for the K-operator K-Dist followed by &-Elimination, and it implies $\mathrm{K}p \rightarrow En$ by the same distribution principle together with KIT, the principle according to which knowledge implies truth. So (7) is correct.

The proof of (CR) from (9) is a bit more problematic. The move involves dropping the possibility operator from the conclusion of (9), and, in general, such a move would be a mistake, since actual truth does not normally follow from possible truth. Williamson thinks the present situation is different, and argues for the difference as follows:

> For it is not contingent whether n is even. Since I can count the books on my table, it is decidable whether n is even; hence n is either odd or even; but if n is odd, it could not have been even, for the mathematical properties of numbers are not contingent; thus n could have been even only if it is even.[19]

Williamson here argues that because the mathematical claims in question are necessarily true if true at all, the distinction between actual and possible truth disappears for such claims. Hence, if the possible truth of such a claim is proven, then so is the actual truth of that claim.

Recall, however, that we began our discussion with the assumption that consistency and possibility are distinct, on pain of trivializing Tennant's proposal. Once we assume this distinction, there can be cases where p and q are true at precisely the same possible worlds, and yet the first but not the second is consistent with some further claim r. The compossibility of each with r must be the same, but if consistency is assumed to be distinct from possibility, then no inference from compossibility to consistency can be made.

Williamson's argument shows that the possible truth of a mathematical claim is compossible with the given premise if and only if (and necessarily so) the actual truth of that mathematical claim is compossible

[19] Williamson, "Tennant on Knowable Truth", pp. 111–112.

with the premise. What he needs to show, and does not show, is that the same holds for consistency and inconsistency.

Williamson might reply that the argument can be resuscitated by noting the particular nature of the claim in question, namely, *En*. Even though consistency and possibility must be assumed to be distinct, the claim that a certain number is even or odd will be easy to show inconsistent if impossible. For to be even is merely to be divisible by two without remainder, and that operation is an easy one to perform on any number. So Williamson may simply need to rephrase the argument from (9) to (CR). He could say that removing the possibility operator on the conclusion is provable because one can derive inconsistency in this particular case. We can thus move from (9) to (CR) by assuming the conclusion of (9), which is $\Diamond En$, and deriving the conclusion of (CR), which is *En*. The argument runs: if $\sim En$, then *En* is inconsistent, hence $\sim\Diamond En$; hence *En*.

This inconsistency claim forms the core of Tennant's reply to Williamson. In summary form, Tennant claims that whether

$$p \,\&(\mathrm{K}p \rightarrow En)$$

is Cartesian depends on the truth-value of *En*, for if *En* is false, an inconsistency will be derivable from the assumption that this claim is true. The proof relies on demonstrating that *En* follows from assumed knowledge of

$$p \,\&(\mathrm{K}p \rightarrow En),$$

and the proof is simple: $\mathrm{K}(p \,\&\, (\mathrm{K}p \rightarrow En))$ implies $(\mathrm{K}p \,\&\, \mathrm{K}(\mathrm{K}p \rightarrow En))$, which in turn implies $(\mathrm{K}p \,\&\, (\mathrm{K}p \rightarrow En))$, from which *En* follows. Since we have already had to assume that *En* is inconsistent if false, it follows that

$$p \,\&(\mathrm{K}p \rightarrow En)$$

is Cartesian only if *En* is true.[20]

Recall that for Williamson to derive a contradiction, he has to run through his proof twice, once deriving *En* and a second time deriving $\sim En$. Given Tennant's results, it follows that at most only one of Williamson's proofs will be allowable, since either

$$p \,\&(\mathrm{K}p \rightarrow En)$$

or

$$p \,\&(\mathrm{K}p \rightarrow \sim En)$$

[20] Neil Tennant, "Is Every Truth Knowable? Reply to Williamson", *Ratio* 14.3 (September 2001), pp. 269–271.

will be anti-Cartesian (since assuming that each is known will generate a contradiction in one of the two cases). Hence, Williamson has failed to show that a paradox can be derived even given Tennant's Cartesian restriction.

It is worth reminding ourselves exactly where we are here. We have been arguing that in order for Williamson's argument to work, he must assume that *En* is consistent only if true, for without that assumption he cannot derive *En* from $\Diamond En$. If we assume that *En* is consistent only if true, however, then one of the two claims with which his pair of proofs begins will fail to be Cartesian, exonerating Tennant's approach from Williamson's challenge. So the crucial claim in the discussion is that we can derive a contradiction from *En* unless it is true.

We noted the importance of this assumption for Williamson to be able to derive (CR) from (9), but perhaps we were a bit hasty. Why does Williamson need to show that (CR) is derivable from (9); why isn't it enough for him to show that (9) must be true if (CR) is true? One reason is that it is this task that Williamson sets for himself—he says he will show, using the logic that Tennant prefers, a knowability paradox even given Tennant's restrictions.[21] What is not clear is why Williamson adopts this task.

With Tennant, let us call a proposition that is necessarily true if true at all a *polar* proposition. Some polar propositions, like *En*, will be consistent only if true, but others will lack this property. For example, consider the identity claim that water is H_2O. Since it is an identity claim, it is a polar proposition; but it is false that it is consistent only if true—it is not a matter of logic that it is true. Williamson could have used a polar proposition such as this one in his proof, and instead of claiming that (CR) follows from (9), he only needed to claim that (CR) could not be false if (9) is true—that is, the move from (9) to (CR) is *semantically valid* even if it is not *syntactically derivable*. On this basis, a paradoxical result follows, even given Tennant's restriction.

The only way out of this alteration of Williamson's strategy for Tennant would be to deny that there are any polar propositions that are not also consistent only if true. That denial entails that there is no such thing as metaphysical necessity that is not simply logical necessity, with the implications that there are no a posteriori necessities and that identity statements need not be necessary. Those implications are false, however: *water is* H_2O is a polar proposition because it is an identity

[21] Williamson, "Tennant on Knowable Truth", p. 111.

claim, but it is known as a result of scientific investigation rather than logical acumen. It is an a posteriori necessity, and any position forced to say that it is not is mistaken.

So I think something similar to Williamson's argument can be used to show that there are technical problems for Tennant's approach. His approach faces other problems as well. It is hard to see how his proposal provides a substantive response to the paradox, since, as I will argue, it is *ad hoc* and unprincipled.

Is Tennant's Proposal Substantive?

Tennant distinguishes two forms of the universal knowability principle, the *actualist* and the *necessitarian* forms. The actualist form applies to truths *simpliciter*, while the necessitarian form applies to all possible truths and asserts that they would be knowable if they were true. This distinction is captured formally by distinguishing (1), which asserts only the knowability of what is true, from its stronger cousin (4):

$$\forall p(p \rightarrow \Diamond \mathrm{K}p).^{22} \qquad (4)$$

Standard anti-realism affirms (4) and not just (1). After all, anti-realism is committed to more than the mere claim that truth *happens* to be epistemic; it is *essential* to the nature of truth that it be epistemic.[23] Thus the anti-realist must, it seems, endorse (4). Unlike the actualist reading of the knowability claim, however, (4) is subject to counterexamples belonging to the second type of anti-Cartesian proposition, existentially inconsistent ones like *No thinkers exist*, as well as those that generate the knowability paradox. Thus the anti-realist must seek solutions to both these problems.

[22] By endorsing (4) restricted to Cartesian claims, Tennant commits himself to (5) $\Diamond p \,\&\, (\mathbf{K}p \nvdash \bot) \Rightarrow \Diamond \mathbf{K}p$, but does not attempt to establish that this claim is true. To do so, he will have to provide a formal system for $\Diamond$ that is complete with respect to sentences $\mathbf{K}p$. This task is daunting. For instance, p's knowability consists in the fact that there is a procedure by means of which a subject could come to recognize that p is true. This requires some very delicate idealization, however, for the procedure may be one that no subject can in fact execute (because of its length, for example), and no one has ever given a precise enough account of this idealization to permit a formal result like (5). So Tennant's endorsement of (5) is something akin to a realist who endorses bivalence because there aren't any counterexamples to it.

[23] For example, Crispin Wright's endorsement of superassertibility as a truth predicate maintains such, inasmuch as superassertibility is itself an epistemic notion. Wright's pluralism about truth predicates may leave open whether other truth predicates are epistemically unconstrained, but Wright does not endorse such a claim nor does he hold that such additional truth predicates would conceive of truth anti-realistically. See *Truth and Objectivity*, Harvard, 1992, especially chapters 1 and 2.

The anti-realist might suggest denying (4) but endorsing its restriction to propositions that do not create this problem, in just the same way that an anti-realist might respond to the knowability paradox by denying (1) but affirming its restriction to propositions that are not subject to the paradox. Since (4) is at issue in the paradox as well as this other problem, the anti-realist may feel an impulse to deal with both problems at one fell swoop by imposing a disjunctive condition on (4): *all truths except those that cause either the knowability paradox or the problem of existentially inconsistent propositions are, necessarily, knowable.*

This proposal is strikingly similar to Tennant's own approach. Such similarity raises two related problems. First, the approach grants far too much to the realist. What the move proposes is to endorse the realist's view of truth in connection with existentially inconsistent statements and statements that generate the knowability paradox. These statements are true or possibly true, but their truth appears radically non-epistemic. Anti-realists should find this conclusion intolerable. After all, if truth would have been radically non-epistemic had there been no knowing subjects, why must it actually be otherwise? Anti-realistic arguments establish, if they are successful, that truth *could not be* other than epistemic. If this means biting the bullet and holding that 'No thinkers exist' is necessarily false, then so be it. Not without reason does the term 'idealism' come to mind when anti-realism is under discussion. Call this problem for anti-realism the *idealism problem*.

Second, the approach above, which says that all truths except those that lead to the idealism problem and the knowability paradox are necessarily knowable, is obviously *ad hoc* and unprincipled, and its similarity to Tennant's own proposal suggests strongly that the charge applies equally well to his view. Tennant even endorses the idea that there is a distinction between principled and substantive responses to the paradox and unprincipled and arbitrary ones. He writes,

> One's immediate intuition is that one would not wish the antirealist's principle of knowability to be deprived of its necessitarian import just because of tricky examples like 'No thinkers exist'. One's reaction would therefore be to restrict the principle of knowability, *but in a principled way*, without depriving it of its philosophical bite.[24] (italics added)

Tennant here makes explicit the prospect of approaching the problems facing (4) in an unprincipled way. We have seen what one unprincipled

[24] Tennant, *The Taming of the True*, p. 274.

solution would look like, one enough like Tennant's own proposal to raise the suspicion that his, too, is *ad hoc* and unprincipled. It is no less unprincipled to say that p is necessarily knowable except when the assumption that p is known is inconsistent than to say that p is necessarily knowable except when it is not. In the former case the resulting knowability principle is no longer as obviously vacuous as in the latter case, but it hardly constitutes a *principled* way for the anti-realist to restrict (1) in order to avoid the two problems. Anyone proposing it would be guilty of constructing an alternative to (4) geared precisely to avoid the idealism and knowability problems, especially when one notes that there are only two kinds of possible truth for which the assumption of knowledge is inconsistent—namely, those propositions that lead to the idealism problem and the knowability paradox. It is as if the anti-realist were to respond to counterexamples by asserting that anti-realism holds for all propositions except the counterexamples.

What goes wrong in these cases is that these proposals do not cite some feature of truth that calls for the restriction in question. Instead, the counterexamples are viewed as tricks up the realists' sleeve, suitable for rejection by even the most *ad hoc* devices.

It is interesting to see how Tennant misses the proposal's lack of principle. Immediately after the above passage occurs the following.

> One would look for some general property F of propositions such as 'No thinkers exist', a property F whose possession by any proposition ϕ would make it abundantly clear why ϕ should be exempted from the intended scope of the anti-realist principle of knowability. The latter principle could then take the restricted form that, necessarily, all true propositions lacking property F are knowable.
>
> It is clear that the anti-realist . . . already has good reason, in the form of propositions such as 'No thinkers exist', to restrict the scope of his principle of knowability . . . What the anti-realist should try to provide, therefore, is some uniform characterization F . . . to restrict his principle in the way just indicated.[25]

Tennant proposes to use anti-Cartesianhood as this property F.

We can agree with the strategy while rejecting the proposal. Tennant is right that a unitary principle excluding the problematic propositions is what is needed. In other words, the search for "some general property F" is the task of providing a unified account of the failure of the anti-realist principle (4), one that allows an explanation of (4)'s failure.

[25] Ibid.

Tennant makes two mistakes, however, about this general property. The first mistake is that he imposes no restrictions on the general property *F* that is supposed to relieve anti-realism of the counter-examples, when it is clear that restrictions are essential to the success of any restriction of (4) if such a restriction is to avoid the charge of being *ad hoc*. After all, the obviously *ad hoc* strategy of maintaining that (4) is true except where it isn't cites some general property that renders (4) immune from counterexample. As pointed out above, the appropriate restrictions on the general property *F* require that the nature of the restriction be motivated by one's general conception of truth, and it is precisely such a motivation that is missing in Tennant's strategy. The second mistake is that involved in his claim that the "anti-realist already has good reason" to restrict (4). The good reason in question is that the anti-realist's (4) appears subject to counterexample. That reason would ordinarily be viewed as a reason to abandon anti-realism, rather than a mere technical difficulty for the anti-realist program, but that is precisely how Tennant interprets the difficulties that plague (4). Because he sees the problems as mere technical difficulties ('trick' is his chosen term of art), the arbitrariness of his approach eludes him.

In fairness to Tennant, he believes that his discussion of the operator "wondering whether" has already given him the requisite reason for thinking that (4) is in need of restriction. That is, he thinks that the knowability paradox is merely an instance of a more general phenomenon in which a variety of claims of the same form as (4) require technical emendation.[26] There are two points to note here, however. The first summarizes what we learned in Chapter 1, that Tennant's argument regarding the operator in question is a failed argument. The other point to note, however, is the enigmatic form of inference involved in this argument. Suppose Tennant were right about "wondering whether". If so, the proper conclusion to draw would seem to be: a fairly plausible claim about wondering—that any truth can be wondered about—would have been false, and "general wonderment" philosophy, defined in terms of acceptance of this claim, would have to be rejected. Tennant thinks otherwise, however. He thinks that general wonderment philosophy is still correct; it is just that we have made a technical mistake in its formulation, just as the anti-realists have made a technical mistake in formulating their position in terms of (4). Moreover, he thinks that

[26] This point is clear both in Tennant's original discussion in the *The Taming of the True* and also in his "Is Every Truth Knowable? Reply to Hand and Kvanvig", *Australasian Journal of Philosophy* 79.1 (March 2001), pp. 107–113.

because the same phenomenon occurs in two areas strengthens his case for claiming that there is a formulation error rather than false doctrine. I, for one, can't see why. It appears to me, instead, that if Tennant were right about "wondering whether", there'd be an interesting relationship between refutations of anti-realism by the knowability paradox and refutations of general wonderment philosophy.

In any case, Tennant's argument about "wondering whether" fails, and so the point is moot. The difficulties for (4) thus are not mere technical problems of formulation, so if (4) is to be restricted in a principled way, one cannot simply exclude by *fiat* propositions that make trouble for it. Instead, one needs to cite something about one's conception of truth that calls for the restriction in terms of property *F*, and this Tennant does not do.

There is another way to make this same point. If we can view the difficulties plaguing (4) as mere 'tricks,' solvable by even the most *ad hoc* emendations, we have a nice recipe for solving many philosophical paradoxes. For example, consider Russell's paradox, which threatens the claim that any grammatically predicative expression defines a set. We can apply Tennant's strategy and avoid the paradox simply by saying that any such expression defines a set except when the assumption that it does so yields a contradiction. Such an approach to Russell's paradox is clearly *ad hoc*. What it does not provide, but what is needed, is some account of the nature of sets that precludes problematic ones and a defense that this account is the proper account to give of what a set is. The same is true of the difficulties plaguing the anti-realists' conception of knowability. What is needed is *an account of the nature of truth* that prevents problematic ones from making trouble for anti-realism. Tennant's proposal simply does not do that.

Tennant is aware of the danger that his account is trivial or unprincipled and attempts to head off the objection. He writes,

> Does our restriction of [(1)] to [(3)] render the knowability principle thereby expressed toothless? Does the restriction to Cartesian truths secure the truth of the knowability principle by sheer stipulation? Does it amount to claiming that the principle holds except where it doesn't? Not at all. Think of all the Cartesian propositions of mathematics and empirical science, propositions that involve no mention, within them, of epistemic notions. To claim that every such truth is in principle knowable is still to forswear metaphysical realism.[27]

[27] Tennant, *The Taming of the true*, p. 275.

Here Tennant argues that his proposal is not trivial because it preserves some room for dispute with realists. In particular, 'all the Cartesian propositions of mathematics and empirical science' remain in contention.

Construed as an argument, these remarks are a failure. To claim that every proposition lacking epistemic notions is in principle knowable is not to forswear metaphysical realism. Realism comes in many varieties, just as does anti-realism; to adopt the slogan that truth is radically non-epistemic does not commit a theorist to any particular claim about any particular kind of truth, including ordinary ones of math and science that contain no epistemic notions. To hold that such truths are in principle knowable is to forswear some versions of realism, but that is all.

There is a further problem here. Tennant holds that 'to claim that every [ordinary truth of mathematics and science] is in principle knowable is still to forswear metaphysical realism.' This claim implies that realism is incompatible with the hypothesis that there could be an omniscient individual. It is not.

By 'omniscient individual' we need not mean one who knows of every statement *p* whether it is true or false. To posit the possibility of such an individual would be to assume the principle of bivalence, thus committing one to a stance on at least one common construal of the realism question. To avoid such a commitment, we can understand omniscience simply in terms of knowing everything that is true. This hypothesis simply does not decide the realism/anti-realism matter. Nothing in it prevents one from holding that truth is conceptually non-epistemic, nor from holding otherwise. Realists can affirm the principle of bivalence and say that truth outstrips our ability to grasp it, and still hold that there could be an individual with infinite cognitive ability for whom every truth is an open book, thus ruling out unknown truths. They can even hold that such an individual is essentially omniscient, grasping every truth in every world in which it exists. In fact, they can even hold that this individual is necessarily existent, so that unknown truths are *impossible*, without betraying their realism. Even the hypothesis of an essentially omniscient necessarily existent individual remains independent of the realism/anti-realism dispute in question, for it neither implies bivalence nor the claim that truth is within the grasp of beings with at most finite extensions of our abilities.[28] Realists and anti-realists may disagree about which statements of mathematics and science are true or

[28] *Contra* Alvin Plantinga, "How to be an Anti-Realist", *Proceedings of the American Philosophical Association* 56, pp. 47–70, where Plantinga argues that anti-realism is true precisely because God is necessarily existent and essentially omniscient.

about which such statements can be known by individuals with nothing more than finite extensions of our abilities, but they need not disagree on the principle that all the *true* statements of mathematics and science are *jointly knowable*. Thus even if Tennant's argument for the non-triviality of his proposal were valid (which it is not), its premise is false.

Tennant has another reply to the charge that his proposal is *ad hoc*, claiming that his proposal is but an instance of a philosophically respectable strategy.[29] He believes there is nothing *ad hoc* about his strategy of throwing out counterexamples that are provably inconsistent with a general claim, and cites Tarski's theory of truth as a case in point for throwing out the Epimenides counterexample "This sentence is false" to the following truth principle:

$$\text{(Truth)}\ \forall p(p \leftrightarrow \text{it is true that } p).$$

Tarski's theory seeks to avoid the paradox of the Liar by stratifying languages so that no language can contain its own truth predicate, thereby avoiding the Epimenides counterexample.

Two points are in order about this example. First, Tarski's theory of truth is far from a simple restriction strategy on (Truth) that replaces it with a more limited claim where the value for p in (Truth) implies no contradiction (the analogue of Tennant's restriction strategy on the knowability claim). Tarski develops a theory of truth that implies such a restriction on (Truth), but he does not propose that principle itself as the solution to the liar sentence. So the Tarski example is not analogous to Tennant's own proposal.

Second, even though Tarski's theory makes an important and serious effort to deal substantively with the Liar Paradox, it is worth reminding ourselves that it has been criticized quite regularly for being an *ad hoc* and unmotivated response to that paradox. I do not adopt a view here on whether these criticisms are justified, but they obviously have some degree of credibility, even though Tarski's work takes serious steps beyond the kind of restriction strategy adopted by Tennant. All things considered, then, the Tarski example gives Tennant no support in claiming that his own proposal is not *ad hoc*.

There is, however, a response Tennant could have made that would be more to the point. At first glance, the proper answer to the worry about

[29] Neil Tennant, "Is Every Truth Knowable? Reply to Hand and Kvanvig", *Australasian Journal of Philosophy* 79.1 (March 2001), pp. 107–113. This reply is to a paper by Michael Hand and myself, "Tennant on Knowability", *Australasian Journal of Philosophy* 77.4 (December 1999), pp. 422–428.

triviality is that the restriction itself is cast in syntactic terms employing the language of derivability, whereas the consequent of the knowability involves only the modal concept of possibility, unlike the proposal that all truths are knowable except when they are not. This point alone does not secure immunity from the triviality charge, but it does place the defender of the triviality charge in a position of needing an argument to buttress the charge.

What might such an argument be? Suppose that a conditional is asserted with an exception clause: All A are B except when the A is also a C. Further suppose that C and ∼B are a priori logically equivalent, and the a priori nature of the equivalence takes no deep or penetrating thought to discover, in the way, for example, that $7 + 5 = 12$ is not terribly difficult to come to know a priori. In such a case, the triviality charge will be difficult to avoid. One such case in the general schema above is when C is just "not in the complement class of non-B" (for the intuitionists in the crowd, we can stipulate that the predicates here are decidable ones such as 'is odd', 'is even', etc., to avoid worries about the intuitionistic rejection of double negation).

I do not claim that the equivalence between *B* and *not non-B* is analytic, for two reasons. First, the distinction between analytic and synthetic is a troublesome distinction to draw. Second, if there is a distinction to be drawn here, there will need to be a further distinction between deep and superficial analytic truths, in order to preserve the distinction between *Everything is self-identical* and *Every even integer greater than 2 is the sum of two primes (counting 1 as prime)*. The better path is to talk of fairly obvious a priori logical equivalence and let the chips fall where they may regarding analyticity. The point I am making is that the following proposal is trivial in virtue of a fairly obvious logical equivalence: All A's are B except when the A's are not in the complement class of the non-B's.

Is Tennant's proposal like this trivial one? The answer depends on what kind of possibility is involved in the concept of possible knowledge. It is well known that there are a multitude of concepts of possibility, and anyone using the language of possibility owes us an explanation of just which notion is intended. In the context of Fitch's proof, the weaker the notion of possibility, the more troubling the consequence derived, and in terms of weak notions of possibility, two options come to mind: logical or metaphysical possibility. The notion of logical possibility can be clarified semantically in terms of the consistency of some maximal proposition, but if it is clarified in terms of consistency in this way, it is

obvious that it is a priori equivalent to a denial of the idea of there being a logical derivation of the falsity constant. After all, consistency is nothing but the idea that no contradiction can be derived.

So that leaves the concept of metaphysical possibility, where that concept is understood so as to leave it open a priori whether or not metaphysical possibility is logically equivalent to logical possibility. In such a case, Tennant's proposal is not trivial, but is subject to refutation by Williamson-style arguments that play on the divergence between logical and metaphysical necessity. One can think of the Williamson argument discussed above as a reductio on the idea that Tennant's proposal can survive the failure of logical equivalence between the class of logically necessary truths and the class of metaphysically necessary truths. But if the Williamson argument proves that Tennant's proposal can only succeed by employing the notion of logical possibility to explain the concept of knowability, then Tennant's proposal is reduced to a triviality. For it says that knowledge of p can be consistently added to the description of any situation in which p is true, except when the addition of that knowledge assumption generates a contradiction.

To conclude this discussion of Tennant's proposal, I should note that Tennant distinguishes hard from soft anti-realists according to whether they claim independent grounds for holding that p is incompatible with $\sim Kp$. The knowability paradox leaves hard anti-realism intact (because that view simply denies the second premise of the paradox, the one that says some truths are not known), but the paradox does touch soft anti-realism, which refuses to endorse the conditionals $p \rightarrow Kp$ and $\sim Kp \rightarrow \sim p$. Tennant holds that his proposal gives the soft anti-realist a way around the paradox.

With the unrestricted knowability principle the hard anti-realist would have to insist on a highly constructive interpretation of the existential quantifier and on highly constructive contents of assertions, generally, in order to reconcile himself or herself to the result established by the paradox: that every true proposition is inconsistent with the failure of it being known. But with the restricted principle, one is deprived of that proof (and, it would appear, of any proof) of that result, and the threat of this uncomfortable actualism about knowledge of truth recedes. The anti-realist can maintain that every (Cartesian) truth is knowable, without thereby being committed to saying that every (Cartesian) truth is known.[30]

[30] Tennant, *The Taming of the true*, pp. 275–276.

What Tennant has *not* shown is precisely what needs to be shown, if the Scylla and Charybdis of hard anti-realism and realism itself are to be negotiated. He has not articulated a *position* that makes truth out to be epistemic but nonetheless treats the knowability paradox by restricting (1) instead of abandoning (2). There was never any doubt that (1) can be restricted in a way the blocks the paradox. The fundamental metaphysical issue is a different one: can (1) be restricted in a way that the principled anti-realist will not find unpalatable? Tennant has not answered this question. He has not formulated an anti-realistic conception of truth that can abide restrictions on universal knowability, and his proposed restriction on that thesis is unhappy—it floats freely, cut off from familiar anti-realist conceptions of truth and lacking one of its own.

DUMMETT AND HAND

Recently, Michael Dummett has entered the camp of those wishing to restrict the knowability claim.[31] Dummett claims that the anti-realist's mistake leading to the knowability paradox is in giving a blanket characterization of truth in terms of knowability. He claims that such a blanket characterization is a mistake, and that the anti-realist ought to give an inductive characterization of truth, whose base clause involves basic statements, to which the knowability characterization applies. All the rest follows the usual pattern: a conjunction is true if and only if each conjunct is true; a disjunction is true if and only if one of the disjuncts is true, etc. Dummett thus maintains that the anti-realist is committed, in terms of knowability, only to the claim that all basic truths are knowable.

Dummett's proposal has been criticized on the same grounds as the criticism of Tennant above. Berit Brogaard and Joe Salerno, for example, write, "As it stands, Dummett's treatment of the paradox is unprincipled. The only reason we are given for restricting the knowability principle to basic statements is that it blocks Fitch's result."[32] And Neil Tennant claims that Dummett's account fails to give us any "rationale" for restricting the knowability principle as drastically as the account does, in particular, more drastically than Tennant's own account does.[33]

[31] Michael Dummett, "Victor's Error", *Analysis* 61, pp. 1–2.

[32] Berit Brogaard and Joe Salerno, "Clues to the Paradoxes of Knowability: Reply to Dummett and Tennant", *Analysis* 62.2 (April 2002), p. 144.

[33] Neil Tennant, "Victor Vanquished", *Analysis* 62.2 (April 2002), p. 142.

Michael Hand answers these complaints by developing a restriction strategy very similar to Dummett's, appealing to the concept of a verification procedure. By generating the account in this way, Hand's proposal cannot be accused of being unprincipled in the way Dummett's account can.

We can think of a verification procedure as a kind of program whose steps are commands of various natures. Using this idea, Hand distinguishes the procedure itself from the performance of it, using this distinction to rescue anti-realism from a commitment to the idea that all truths are knowable. According to Hand, to claim that all truths are knowable is to claim that the procedure itself is executable or performable, and a procedure might be unperformable even though the proposition associated with that procedure is nonetheless true.

How might this happen? Hand's idea is that there are two different ways in which a procedure might be unexecutable. The first is for the series of commands itself to display interference, for it to be structurally incoherent. For example, consider the verification procedure that includes "Look to the left" and "Never look to the left." This series of commands is structurally incoherent. It is impossible to carry out in virtue of its contents. Because it is impossible to carry out for this reason, any proposition whose meaning is identified with the series will be one that is necessarily false.

Other series of commands cannot be carried out, though for a different reason. In order to specify this other type of unperformability, we need to distinguish between basic sentences and their verification procedures, and complex sentences and their verification procedures. Consider, for example, the proposition *Some thinker exists*. If we distinguish between basic and complex sentences, we can specify the verification procedure for this quantified sentence in terms of the verification procedures of its instances. The verification procedure for the quantified sentence is executed by executing the verification procedure for its instances. Thus, for *Some thinker exists*, we verify it by examining existent things, and the procedure terminates when a positive instance is found. Even if we assume that the procedure for the instances is executable, it won't follow that the procedure for the quantified formula is executable, however. Consider a universal statement regarding an infinite domain: examining an instance can be done, but no stopping point will ever be achieved if the statement is true.

Hand holds that something similar to this kind of unperformability occurs when we consider claims such as *No thinkers exist*. The individual steps of the verification procedure are not inconsistent in virtue of their

content, either internally or together with the other individual steps, but cannot be executed if the claim is true for a different reason (the reason being that there is no one to execute the steps). The individual steps are still executable, and hence the sentence has a meaning and might be true, since it is inconsistency among basic steps that implies logical impossibility for the claim in question.

Hand holds that performability of a procedure is only required for basic procedures:

> The epistemic nature of truth does not reside in the necessary existence of epistemic agents, but only in the performability in principle of the various "initial" verification procedures and the operations on procedures that yield new procedures for more complication propositions, that is, the verificatory *steps* that make up the procedures in general. That is, the epistemic nature of truth is due to the effectiveness of these procedures.[34]

One way of explaining this idea would be to merge Hand's appeal to "initial" verification procedures with Dummett's distinction between basic and non-basic statements so that the initial verification procedures are those for basic statements. If so, then Hand's and Dummett's proposals are twins, though Hand's is motivated in a way that Dummett's is not. Hand's view may also be viewed as an extension of Dummett's, if basic statements are conceived of as able to be associated with a multi-step verification procedure. There is some discomfort in this extension, however, for Hand's idea seems to be that performability only comes into play in the smallest bit of complex procedures. If basic statements might have multi-step procedures, then there is no reason a priori to think that these procedures have to be executable, since the combination of steps might introduce "interference" of just the sort Hand uses to argue against the idea that the heart of the anti-realist position involves a commitment to the performability of verification procedures. So I think the best view to adopt is that the "initial" procedures are one-step procedures, and such procedures are precisely the sort possessed by basic statements, even though this latter claim is not found in Hand's own description of his theory. Viewed in this way, Hand's view is Dummett's combined with a decent motivation so that the distinction between basic and non-basic statements is no longer central to the account.

I think, nonetheless, that the proposal is not wholly successful. The issue involves the connections between the concepts of the *effectiveness* of

[34] Michael Hand, "Knowability and Epistemic Truth", *Australasian Journal of Philosophy* 81.2 (June 2003), p. 218.

a procedure and the *executability* of that procedure. The proposal Hand wishes to avoid is a standard operational account of effectiveness:

PROPOSAL 1 A procedure is effective iff it is executable (i.e., it can be executed by a suitably placed epistemic agent) and were it executed, it would reveal the truth-value of the related proposition.

Hand's proposal is a more limited, non-global requirement:

PROPOSAL 2 A procedure is effective iff each step of the procedure is executable (i.e., could be executed by a suitably placed epistemic agent), and were the procedure executed, it would reveal the truth-value of the related proposition.

Hand motivates this restricted alternative to Proposal 1 by a discussion of recursion theory and the apparent way in which concepts of recursion theory, such as computability, seem to involve something akin to Proposal 1:

> The idea that effective procedures are performable by properly qualified epistemic agents is often used to motivate the ideas of recursion theory, and can be seen in many of the technical terms of that theory, such as 'computable' and 'decidable'. These terms suggest performability by an agent, but in fact the theory itself has no such presupposition. To be a recursive function is to have a definition in terms of various functions (the so-called initial functions) and relations among functions (composition, recursion, minimization. Calling a function computable presupposes no computing agents, but is merely a statement of how the function related to the specified initial ones. No doubt the choice of initial functions and of the permitted operations on functions can be motivated by reference to what computing agents can do, but the abstract study of these functions need not presuppose the existence of any agents available to perform calculations with them.[35]

Hand remarks here that one can motivate the ideas of recursion theory with reference to what computing agents can do, but that way of motivating the ideas of theory is no part of the theory itself, properly conceived. Just so, the account of verification procedures relies on the effectiveness of the procedures in yielding a truth-value, not in the performability of the procedures by suitably placed epistemic agents, as Proposal 1 requires.

This argument is a sound one for rejecting Proposal 1, but it is less successful as an argument for Proposal 2. The analogy with recursion theory should lead to a proposal in place of Proposal 1 in which the role of computing or epistemic agents is relegated to the pedagogical role of

[35] Ibid., p. 218.

motivating the ideas involved in the abstract specification of various routines or functions—that is what the recursion theory analogy supports. Proposal 2, however, allows the intrusion of agents into the account of the initial procedures, and in this way, fails to be modeled strictly on the recursion theory example.

A different way of pursuing the analogy is as follows. The problem with Proposal 1 is that it tries to operationalize, in the context of epistemic agents, the notion of effectiveness for a procedure. The problem with such operationalizing can be brought out by considering the analogy with recursion theory. The operational elements by which Proposal 1 attempts to define effectiveness of a procedure play a much different role in recursion theory—they are used to motivate the ideas of the theory, not to define them. As such, these operational elements have *no place whatsoever* in the theory, properly conceived. Just so in an appropriate account of the effectiveness of a verification procedure—the operational elements used in Proposal 1 have *no place whatsoever* in the theory of effectiveness of a verification procedure, properly conceived. If such elements have no place whatsoever, however, we must reject not only Proposal 1 but Hand's own Proposal 2 as well. For Proposal 2 does not expunge the theory in question of the operational elements in Proposal 1, it only limits their role to the initial steps of verification procedures.

In sum, if the problem with Proposal 1 is its operational flavor, a natural antidote to the mistake would be a proposal that simply refuses to operationalize at all. Proposal 2, however, is a halfway house between full operationalizing and no operationalizing. The problem is that the example used to argue that operationalizing is a mistake lends no support to the halfway house of operationalizing only the initial procedures, and thus Hand's proposal must be viewed as unsupported by the argument he gives for restricting the knowability claim.

We have, to this point, examined four attempted restriction strategies on the universal knowability claim and found problems with all of them. Even though none of these restriction strategies are successful, we have no reason at this point for thinking that no restriction strategy can be successful. That would be unfortunate if our investigation of the paradox required drawing a conclusion on the matter, for any such conclusion would be less than compelling. We don't need such a conclusion, however, for there is a deeper flaw with the idea that a suitable approach to the paradox involves restricting the knowability claim. I will argue that pursuing restriction strategies does not address the fundamental paradoxicality involved in the knowability paradox.

THE FUNDAMENTAL FLAW OF RESTRICTION STRATEGIES

So far, we have seen a number of attempts to save anti-realism from the paradox by refusing to endorse the knowability claim. Instead of holding that all truths are knowable, restriction strategies insist only that certain kinds of truths need to be knowable. Perhaps only actual truths need to be knowable; perhaps it is only basic or atomic truths that need to be knowable; or perhaps it is only truths that do not yield a contradiction when we assume that they are known that need to be knowable.

This direction of the literature on the paradox depends on a certain assumption regarding the nature of the paradox being addressed. That assumption is that the fundamental problem raised by the paradox is the threat it poses to positions inclined to endorse the knowability claim. This assumption views the paradox as a local problem for anti-realists, rather than the global problem I have argued exists for everyone.

This point deserves careful explanation. The explanation will take us on a slight detour through a theistic response to the paradox, one that I will argue fails to appreciate the true nature of the paradox. I will then show that restriction strategies adopt a position on the paradox that is similarly inadequate.

The knowability paradox is typically thought of as deriving from two assumptions. The first is that all truths are knowable and the second is that some truths are unknown. Upon generating a contradiction from these two assumptions, we are required to discharge, leaving us to conclude that the knowability claim is false (and anti-realism thereby threatened).

Suppose, however, that we characterize the paradox in terms of there being a proof that there are unknown truths, for it begins by assuming that a particular truth is known (the truth that p is an unknown truth), and derives from that claim the impossibility of knowing this conjunctive truth. For philosophers of a theistic bent, this characterization of the paradox may disturb, since it threatens the idea that there is an omniscient being. Given such a disturbance, theistic philosophers may see themselves as having a strong reason to find some flaw in the proof, hoping thereby to prevent the knowability paradox from refuting their theistic perspective.

Such a response, however, involves confusion. After more sober reflection, the theistic philosopher may see the flaw in this reaction to the paradox. The theistic philosopher may come to see that the above proof

is no threat to theism unless it is true that p is an unknown truth, for if p is not an unknown truth, then omniscience does not require that it be known that p is an unknown truth. Yet, if there is an omniscient being, then there aren't any unknown truths!

Hence, if one thinks there is an omniscient being, the above proof can be dismissed. It is no more interesting a challenge to theism than any argument that presumes an omniscient being must know what is false. So the theistic philosopher can move on to other interesting areas of philosophy, knowing that the knowability paradox is of no concern.

So characterized, the supposed reaction by the theistic philosopher is both right and wrong. It is right in that the proof above does not threaten the claim of omniscience, but it is wrong in supposing that nothing paradoxical remains about which the theistic philosopher need be concerned.

One way to see this point is to notice that the paradox does not depend simply on whether one accepts the two assumptions in question, the knowability assumption and the non-omniscience assumption. The central perplexity involved in the paradox does not depend on the idiosyncracies of one's favored philosophy, but rather on a perplexing lost logical distinction between what is actually the case and what might be a case. It is obvious that knowledge implies the possibility of such, but what is not obvious is what the Fitch proof attempts to demonstrate: that, to put it carelessly, possible knowledge implies actual knowledge. Should that distinction disappear, it would be fitting to find ourselves in a state of perplexing philosophical stupor. How could it be that there is no logical distinction between actuality and possibility?

We might try for equilibrium by reminding ourselves that there are philosophical domains in which the distinction between actuality and possibility is lost. For example, modal logicians have long been comfortable with the idea that what is actually necessary is not logically distinct from what is possibly necessary. The comfort experienced by this thought will not last long, however. We are comfortable with the lost distinction in this domain because we have a semantical theory to which to appeal to explain why there is no logical distinction here, and we became comfortable with denying the distinction here only after the development of the semantical theory that makes intelligible the loss of such a distinction. Nothing similar can be said when we return to the context of the knowability paradox, however: we have no semantical basis whatsoever for being sanguine about a lost distinction between actual and possible knowledge.

So even if our theistic philosopher should dissent from the assumption that there are unknown truths, said philosopher has as much reason as anyone to view Fitch's proof as establishing a very troubling conclusion. For there is nothing about theism that yields an explanation as to why actual and possible knowledge are not logically distinct. It is for this reason that the theistic response to the paradox is a red herring, even though such a philosopher can take refuge in holding that theism itself is not at stake. The imagined theistic philosopher denies that there are unknown truths, and it thereby achieves serenity in the face of the paradox. Such serenity is warranted, however, only if the theistic perspective does more. It will need to explain why, to speak again with the philosophically vulgar, there is no logical distinction between actual and possible knowledge.

We can now apply this understanding of the failure of this theistic response to the anti-realist restriction strategies. Anti-realists, I maintain, do something similar to what the imagined theist has done. The imagined theist denies the assumption of non-omniscience, thereby claiming to avoid any perturbation from the paradox. The now-dominant anti-realist strategy is to deny the knowability claim, substituting for it some careful emendation with weaker implications, also thereby claiming to avoid the reach of the paradox. Yet, if the theistic response to the paradox is a red herring, one should wonder why the anti-realist restriction strategy isn't as well. What reason can an anti-realist give on behalf of a restriction strategy that will render respectable such a response to the paradox in contrast to the theistic response?

Here is the reason an anti-realist has to give. The claim above is that the heart of the paradox concerns a lost logical distinction between actuality and possibility. This characterization, however, needs to be made more precise; we need to speak, not in the loose and popular sense here, but with clarity and precision. The lost distinction is not one between actual and possible knowledge, for even false (contingent) claims are objects of possible knowledge (in worlds where they are true). Hence, one can't derive that *p* is actually known from the claim that it might be known.

So suppose we try to be more careful here. The lost distinction, the anti-realist can claim, is a lost distinction between actual known *truths* and possible known *truths*. That is, a careful presentation of the lost distinction is

$$(\mathrm{LD})\ \forall p((p\&\Diamond Kp) \leftrightarrow (p\&Kp)).^{36}$$

[36] Here I will ignore a complication that I will discuss later. Intuitionists deny the double negation rule needed in Fitch's proof to finish by concluding that all truths are

The proof from p&Kp to p&◇Kp is trivial, depending only on the modal principle that what is actual is possible. So, the anti-realist can claim, the heart of the paradox is found in demonstrating that p&Kp follows from p&◇Kp. That proof, however, requires assuming that all truths are knowable. So (LD), the careful expression of the heart of the knowability paradox in terms of a lost logical distinction between actuality and possibility, is derivable only on the assumption that all truths are knowable. Hence, employing a restriction strategy on the knowability claim, weakening it in such a way that (LD) can no longer be derived, is a perfectly respectable strategy in dealing with the paradox. So long as the claim that truth is epistemic involves only some weaker commitment than would allow a derivation of (LD), the anti-realist may claim to be entitled to a serenity for having disarmed the paradox to which the theistic philosopher is not.

This account fails to appreciate the full generality of the paradox, settling instead for a local characterization most relevant to those of anti-realist persuasion. What is valuable in it is the demand for precision in formulating the lost logical distinction that is at the heart of the paradox. The failure to appreciate the full generality of the paradox is demonstrated in thinking that (LD) above is the only precise formulation of the lost distinction. If we consider again the most common approach to explaining the paradox, it involves two assumptions, the assumption that all truths are knowable and the assumption that some truths are not known. The proof then shows how to derive a contradiction from these two assumptions, thereby demonstrating that the knowability assumption implies omniscience. As noted already, the proof in the other direction is trivial, thereby generating a logical equivalence. The proper representation of the logical equivalence here is not (LD), however, but rather

$$(\mathrm{LD}^*)\ \forall p(p \rightarrow \Diamond Kp) \leftrightarrow \forall p(p \rightarrow Kp).$$

Furthermore, whereas (LD) is not a theorem, but instead depends on the knowability assumption, (LD*) is a theorem (if we grant the validity of Fitch's proof). Carefully articulated, it claims that there is no logical distinction between universally knowable truth and universally known

known. Instead, all that is derivable by intuitionistic logic is that nothing unknown is true, or that everything not known is false. The issue I'm raising here could be put in a way that remains neutral on this dispute between classicists and intuitionists, but doing so would only make for cumbersome presentation, since this dispute is not relevant to the point I'm making in the text concerning the heart of the paradox.

truth. This more careful articulation still codifies a lost logical distinction between actuality and possibility.

As noted already, this lost distinction is, purportedly, a theorem—the logical representation of it, (LD*), is apparently provable from no assumptions whatsoever. As a result, pursuing a restriction strategy on the knowability claim will be irrelevant to any felt need to preserve such a logical distinction. Anti-realists may still find comfort in undermining (LD) by pursuing a restriction strategy, but they should not pretend that undermining (LD) has any implications for the paradox itself. To claim such will be to adopt a strategy toward the paradox that is, in a word, a red herring.

The important point to note about the distinction between (LD) and (LD*) is that the latter depends on no assumptions whatsoever, either realist or anti-realist. It is therefore a formulation of the central paradoxicality engendered by Fitch's proof that is more general than the formulation provided by (LD). On this more general understanding of the paradox, what is threatened is not the logical distinction between known and knowable truth, but rather the distinction between universally known and universally knowable truth. This difference, however, is insignificant. The fundamental paradoxicality here is that of a lost distinction between actuality and possibility, and that loss is as perplexing on either rendition of the paradox.

Once this point has been appreciated, the question of the basic commitments of anti-realism is simply irrelevant to the paradox. So if a restriction strategy should be found that keeps anti-realists from having to affirm the claim that all truths are knowable, that result would be interesting to the tenability of anti-realism but irrelevant to the paradox. To put it simply, paradox remains whether or not anti-realism is defensible, for the heart of the paradox is about a lost logical distinction between possibility and actuality. If anti-realism is true, this lost distinction is between known truth and knowable truth. If anti-realism is false, the distinction is still lost, though in a different place. Without anti-realism, the paradox still threatens the logical distinction between universally known truth and universally knowable truth, and that loss is no more palatable than the loss implied by anti-realism.

CONCLUSION

In the last two chapters, therefore, we have seen the scope of commitment to the major assumption of the paradox and an attempted line of

defense of the viewpoint expressed (perhaps poorly) by that assumption. The topic of these chapters is interesting for the question of the plausibility of anti-realism and other views that appear to be committed to the knowability claim, and here our results give cause for concern for those who wish to defend such views. From the perspective of the paradox itself, however, the discussion is not relevant at all. It is easily explicable psychologically why an anti-realist seeing the proof would immediately worry about the defensibility of anti-realism, but such theory-driven responses to the paradox should not blind us to the fact that there is something deeply paradoxical here, whether or not the paradox undermines anti-realism and other views that appear to share a commitment to the knowability claim. Once we see that the heart of the paradox is the lost distinction in the domain of truth between universal knowledge and universal knowability, it becomes clear that the question of the plausibility of anti-realism is not fundamental to the paradox.

In the next chapter, we will turn to an approach to the logic of the paradox that is relevant. There we will consider whether the extra-logical rules used in the derivation of Fitch's result can be maintained.

4

Rules for the Knowledge Operator

As we saw in Chapter 1, the variety of ways of generating a paradoxical result from the claim that all truths are knowable relies on no obvious flaw in the application of first-order classical logic, leaving those who wish to avoid abandoning the knowability claim looking elsewhere for a culprit. A natural suspicion one might have concerns the rules for the knowledge operator, (KIT), and (K-Dist). The first rule codifies the idea that knowledge implies truth, and the second allows knowledge of a conjunction to yield knowledge of the conjuncts. I will begin with a discussion of (KIT), but the greater part will be devoted to (K-Dist), since that rule raises the issue of whether, in slogan form, knowledge is closed under deduction. To be a bit more perspicuous, the issue is whether the set of things known also includes all of the logical consequences of that set, and if not, whether there is some appropriate restriction on this closure claim that is defensible. We begin, however, with the less controversial rule, (KIT).

DENYING (KIT)

This principle encodes the idea that the knowledge operator is factive:

$$(\text{KIT})\mathrm{K}_p \vdash p$$

As I argued in Chapter 1, the heart of the knowability paradox involves operators that are factive, so (KIT) is a crucial principle in the derivation of Fitch's result.

It is hard to find in the history of epistemology reasons for questioning (KIT). The only arguments for a rejection of this principle derive from ordinary language considerations, in which people sometimes use the term 'knowledge' and its cognates in ways that do not require truth. For example, upon being surprised by a grade on an exam, a student might

lament by saying, "I just *knew* I was going to do better than that!" Other examples can be found in careless descriptions of the science of a particular era. Even some philosophers are prone to language of the following sort: "According to the scientific knowledge of the ancients, the earth was flat."

It cannot be denied that there are appropriate uses of the verb 'knows' and its cognates that violate (KIT). Those who speak in the above ways are not always misspeaking. Even so, there are straightforward ways to explain away these examples while retaining (KIT). In order to see how to do so, we need a distinction between semantical and pragmatic aspects of language.

The clearest example of this distinction involves the difference Paul Grice investigated between semantical and conversational implication.[1] We often draw conclusions from what people explicitly say, where the conclusion drawn is a different claim from what was said. For example, a friend might say, "Jim is a bachelor", and we might conclude from what is said that Jim is not married. This inference is semantic in character, since it is part of the conventional meaning of 'bachelor' that one is unmarried. Not all conclusions drawn are semantic in character, however. My daughter wishes to go to a party, to which I say, "You can't go to the party unless you clean your room." She goes to her room and cleans it, concluding from what I said that if she cleans her room, she can go to the party.

It is important to recognize that this inference is not semantical in character. For one thing, it is not truth-preserving. What I said is logically equivalent to the claim that she can't go to the party if she doesn't clean her room, which is of the form $\sim c \rightarrow \sim p$—in slogan form, "No clean, no party." She concludes "Clean room, party hearty", which is of the form $c \rightarrow p$, and this latter claim clearly does not follow from the former.

Some philosophers think "unless" should be translated as "iff not", and will find the example in the text unpersuasive for that reason. There are relatively clear cases, however, in which this translation is unsatisfactory. The dour queen who, in the face of criticism by an advisor that she sentences too many people to death, says, "Well, at least I don't guillotine anyone unless they have a head", semantically implies that if a person has no head, she doesn't chop it off. But she does not imply that

[1] Paul Grice, "Logic and Conversation", in *Studies in the Way of Words* (Cambridge, Mass., 1968); page references are to the reprint of the article in *Pragmatics: A Reader*, edited by Steven Davis (New York, 1991), pp. 305–315.

if you have a head she chops it off—this lack of implication helps explain why her advisor doesn't run for the door in terror of being beheaded. So, at least sometimes, "unless" means only "if not", and a semantical theory is more elegant if it can find ways to treat logical connectives as *constants*. The beauty of Grice's distinction between conversational and semantical implication is that it allows us to treat "unless" as a constant, explaining the other uses by recourse to the various maxims that Grice delineated as governing conversational contexts. The relevant one here falls under the category of Quantity, and the maxim in question is "Make your contribution as informative as is required (for the current purposes of the exchange)."[2] In the context of my conversation with my daughter, the purposes of the exchange are transactional: she wants to determine what to do to be able to go to the party, and I want to secure the completion of some chores. In that context, to honor this maxim of Quantity, I need to be forthcoming about all the chores that need to be completed in order for her to be able to go to the party. Thus, by specifying only the cleaning of the room, I conversationally, but not semantically, imply that if she cleans her room, she can go to the party.

The precise character of the distinction between semantics and pragmatics is controversial,[3] but we need not possess a careful account of the distinction to recognize particular examples of the phenomenon such as those above. Once we recognize the existence of the distinction, we have resources to explain the appropriateness of uses of 'know' and its cognates that violate (KIT) without abandoning it. The uses are appropriate for efficiently conveying a content that differs from the semantic content of the utterance itself. In the example, "I just *knew* I was going to do better than that", the utterance conveys the maximal conviction of a person, together with the psychological aspect of closure characteristic of knowledge—that further inquiry into the issue would have been a waste of time. So even though the statement is strictly false—because it violates (KIT)—it is nonetheless an appropriate use of language because the choice of words is particularly well-suited to an efficient means of communicating, via conversational implication, the thought in question.

There are other reasons as well to be unimpressed with the idea that knowledge might not imply truth. When we inquire concerning the

[2] Grice, "Logic and Conversation," p. 308.

[3] For a thorough discussion of the issues involved, see Jason Stanley and Jeffrey C. King, "Semantics, Pragmatics, and the Role of Semantic Content," in Zoltan Szabo, ed., *Semantics vs. Pragmatics* (New York, forthcoming).

value of knowledge or the value of truth, the two topics are obviously connected, so that if one denigrates the importance of truth, one thereby denigrates the value of knowledge. It is hard to imagine a pragmatist rejection of the value of truth[4] while at the same time embracing the singular value of knowledge as some unsurpassable cognitive achievement, for knowledge is valuable only if finding the truth is valuable. (KIT) is part of the explanation of why the value issues are related in this way.

One final point is in order. Even if one wishes obstinately to reject all of the foregoing, we can still point to theoretical positions under attack that require some operators that are truth-implying. Verificationism, for example, is committed to the idea that (a suitable kind of) verification implies truth, and semantic anti-realists wish to define truth in terms of our intellectual capacities, requiring that some cognitive achievement is logically sufficient for truth. So even if a critic wishes to quibble with the above arguments for (KIT), it will still be an indefensible position to hold that nothing in the neighborhood of (KIT) can be used to threaten the anti-realist positions under attack by the knowability paradox. So, in light of the arguments for (KIT), the simplest route for defenders of anti-realism to follow is to grant (KIT) and hope to find mistakes elsewhere. Hence, if there is a logical problem with the derivation, it is not traceable to (KIT).

DENYING (K-DIST)

The situation is quite different when it comes to the distribution rule (K-Dist). For there are well-known epistemological theories that deny that knowledge is closed under known implication, and if knowledge is not closed under deduction, it would be natural to wonder why (K-Dist) should be accepted. The motivation behind such a rejection is the desire to avoid skepticism, and the idea is that if knowledge is not closed under known implication, then many skeptical challenges simply turn out to be irrelevant to the question of whether we know what we ordinarily believe. For example, we know that Bush is President, and that claim implies that there is no Secret Service plot to fool the public and maintain social stability in the face of the recent death of Bush by using a

[4] For such a position, see Richard Rorty, *Philosophy and the Mirror of Nature* (Princeton, 1981).

Bush-imposter to maintain the public's impression that Bush is still alive. Do we know that there is no such plot? Some think not, and claim that our failure to know it undermines our claim to know that Bush is President. Yet, if knowledge is not closed under known implication, no conclusion follows about whether we know that Bush is President. Perhaps we can know that Bush is President and yet fail to know that the many skeptical hypotheses incompatible with this claim are false. Skepticism may be irrefutable, but at least it has been rendered impotent for skeptical purposes if this account succeeds. One of the best-known defenders of this approach to skepticism is Robert Nozick,[5] though the original discussion of the closure issue is developed by Fred Dretske.[6]

The question in our context is whether any of these concerns about the purported failure of closure undermine (K-Dist). We can begin first with Dretske's approach.

DRETSKE'S ACCOUNT

Dretske's investigation of epistemic operators proceeds by distinguishing between *penetrating* and *semi-penetrating* operators. A penetrating operator Ω is such that, if p entails q, then Ωp entails Ωq. Semi-penetrating operators are ones that penetrate to some of the logical consequences of what they govern, but not to all. Dretske argues that epistemic operators such as "S knows that", "S can prove that", "there is evidence that", etc. are semi-penetrating operators, and thus that knowledge is not closed under known logical implication. But such operators are not non-penetrating, for they do penetrate to some very important logical consequences of what they govern. Most relevant here is what Dretske says about the rule of simplification:

> . . . [I]t seems to me fairly obvious that if someone knows that *P* and *Q*, has a reason to believe that *P* and *Q*, or can prove that *P* and *Q*, he thereby knows that *Q*, has a reason to believe that *Q*, or can prove (in the appropriate epistemic sense of this term) that *Q*.[7]

Dretske here explicitly endorses (K-Dist). Thus, even though he rejects the view that knowledge is closed under known logical implication, there are some rules of implication in which knowledge penetrates

[5] Robert Nozick, *Philosophical Explanations* (Cambridge, Mass., 1981).

[6] Fred Dretske, "Epistemic Operators", *Journal of Philosophy* 67 (1970), pp. 1007–1023.

[7] Dretske, "Epistemic Operators", p. 1009.

to what is implied. According to Dretske, simplification is one such rule; hence Dretske's position provides no comfort whatsoever to those who wish to escape the force of the knowability paradox by denying (K-Dist).

NOZICK'S APPROACH

For those hoping to find a way out of the paradox by denying (K-Dist), greater comfort can be found in Nozick's approach. According to Nozick, to know something is to be a tracker of its truth, i.e., roughly, to be such that one would believe the claim if it were true and one would not believe the claim if it were false. Nozick then concludes that knowledge cannot be closed under known logical implication, for even if p implies q, counterfactual situations in which p is false do not co-vary with counterfactuals situations in which q is false. Thus, one might be a truth tracker with respect to p, and yet not be a truth tracker with respect to q. That is, q might bear the following relation to a person who knows p: if q were false, the person in question might or might not believe it.[8]

It is important to note that this rejection of the thesis that knowledge is closed under known implication is completely theory-driven, and thus the argument against the closure thesis is no more secure than Nozick's theory itself. In order to be persuasive, we need either a persuasive argument for the theory or some other intuitive basis for the rejection of (K-Dist). Nozick's own dialectic does not rely solely on the implications of his theory. He says,

> S's belief that $p\&q$ tracks the fact that $p\&q$; if it were true he would believe it, and if it were false he wouldn't believe it. It may be that if the conjunction $p\&q$ were false, it is the first conjunct p that would be false, and in that situation the person wouldn't believe p and so wouldn't believe $p\&q$. However, it does not follow that his belief in q tracks the fact that q; for if q were false (which is not what would or might be the case if the conjunction were false—p would then be the culprit) he might still believe q . . .
>
> So, we must adjust to the fact that sometimes we will know conjunctions without knowing each of the conjuncts. Indeed, we already have adjusted. Let p be the statement that I am in Emerson Hall, not-SK be the one that I am not on Alpha Centauri floating in that tank; since p entails not-SK, p is

[8] Nozick, *Philosophical Explanations*, pp. 204–211.

(necessarily) equivalent to p & not-SK. I know that p, yet I do not know that not-SK.[9]

There are two points to note about this rejection of (K-Dist). First there is the motivation used in the first paragraph, a motivation that is only as persuasive as Nozick's theory of knowledge. Second is the skeptical problem noted, where we may know where we are at present, but not know that we are not in a skeptical scenario. In order to ascertain whether the knowability paradox can be avoided by denying (K-Dist), we will need to examine both motivations, the motivation provided by Nozick's theory itself as well as the intuitive basis in avoiding skepticism that Nozick cites in motivating a denial of closure and, in particular, a denial of (K-Dist).

This intuitive basis for rejecting closure is found as well in Colin McGinn's explanation for rejecting closure.[10] McGinn argues that the fundamental kind of knowledge is *discriminatory* knowledge, i.e., knowing one thing from another. This starting point gives, he claims, a pretheoretic ground for rejecting the closure principle:

> I want to suggest that the natural intuitive argument [against closure] introduces considerations of discrimination in a way that conforms with my earlier contentions about knowledge. The following seems an intuitively correct principle: one can *know* that p only if one can *tell whether* p—I can know that (e.g.) it is raining outside only if I can tell whether it is raining outside.[11]

McGinn then suggests that this pretheoretical intuition about the need for discrimination offers a good argument against the closure principle:

> It seems to me, furthermore, that the discrimination account of knowledge is hovering in the background of this intuitive reasoning; for what is critical to the argument [against closure] is the idea of a capacity to *distinguish* one state of affairs from others. To know that there is a table there requires discriminating tables from chairs and dogs and empty corners; to know you are not a brain in a vat requires discriminating this state of affairs from envatment—and a person could have the former discriminative capacity without the latter.[12]

McGinn analyzes the core of the intuition that knowledge is not closed under known implication in terms of the incapacity to discriminate

[9] Ibid., p. 228.

[10] Colin McGinn, "The Concept of Knowledge", *Midwest Studies in Philosophy IX 1984: Causation and Causal Theories*, Peter French, Theodore E. Uehling, Jr, and Howard K. Wettstein, eds., (Minneapolis, 1984), pp. 529–554.

[11] Ibid., p. 543.

[12] Ibid., pp. 543–544.

envatment from non-envatment, and there is surely something right about this account. Moreover, the emphasis on discriminatory capacities in McGinn's thought and the related idea of the sensitivity of a belief in Nozick's epistemology both threaten closure through the same intuitive basis, that we are not in a good enough position to tell whether we are envatted, even if we are in a good enough position to know that we have hands.

I want to argue first that the burden of the argument against (K-Dist) is going to have to be carried by Nozick's theory or its close cousins rather than by the anti-skeptical motivation cited by Nozick and McGinn. So I turn first to address the intuition behind the skeptical scenario in order to rebut the impression that the skeptical scenario by itself gives grounds for a rejection of closure. After canvassing the growing consensus in contemporary epistemology that closure is compatible with a non-skeptical attitude, we will turn to the task of evaluating whether Nozick's theory and its cousins provide sufficient motivation for denying (K-Dist).

SKEPTICAL SCENARIOS AND THE QUESTION OF CLOSURE

One response to Nozick's use of the brain-in-a-vat hypothesis is simply to insist that we do know that we are not brains-in-a-vat. This reaction is, I believe, the common reaction to far-fetched skeptical scenarios invented by epistemologists. Those more skeptically inclined may challenge this knowledge claim, and an adequate answer to the challenges may be hard to find, but the absence of an adequate response to skeptical challenges in no way threatens the claim that we do in fact know that we are not in skeptical scenarios.

Even so, it is worth noting what kinds of explanations might be offered to rebut the skeptic's concerns. One approach is broadly Moorean, according to which we infer the falsity of the skeptical scenario from our obvious knowledge of various claims of common sense. Another, more recent approach, defended by Timothy Williamson, is that the central claim underlying appeal to skeptical scenarios is false, the claim that one's qualitative experience is identical to what they are in the actual situation.[13]

[13] Timothy Williamson, *Knowledge and its Limits* (Oxford, 2000), chapter 8. See also John McDowell, *Mind and World* (Cambridge, Mass., 1994).

There is a deeper problem here, however. If the only problem were that some skeptical folk are tempted to think that we don't know that the skeptical scenarios do not obtain, one of the approaches just mentioned might be sufficient to explain why skeptical scenarios present no counterexample to closure principles. The deeper problem is that skeptical scenarios are only one instance of a more general phenomenon, for apparent lapses of closure are not limited to cases involving far-fetched imaginary cases constructed by philosophers.

Here are some examples of this phenomenon. Suppose I've been invited to give a talk in St. Louis next month and have agreed to go, and my brother calls to ask if he can visit the very same date as the talk. I reply that there is a problem since I won't be home that day. That I won't be home that day is something I know. Yet, I also know that tragedies sometimes happen—people have heart attacks, strokes, auto accidents, etc.; and I also know that were I to suffer some of these tragedies, I wouldn't be in St. Louis at the time in question. So it looks like the basis of my response to my brother implies something about immunity from tragedies, and surely I don't know that I won't suffer these tragedies.

Lottery situations provide other examples. My daughter sees a million-dollar mansion on television and wants to know if we could buy a house like that. I reply that I'll never have enough money for such a house, and this claim is something I know to be true—after all, I'm a philosophy professor, happy in this role with no intensions of changing careers, and I have no rich relatives from whom I might receive a large inheritance. I bought a lottery ticket, however, and I know that if I win the lottery, I would be able to buy such a house. I may be justified in believing that I won't win the lottery, but I would not claim to know that I won't win the lottery.

It is important to note that not all such cases involve knowledge of the future, but that aspect is not essential to the issue at hand. One of the original examples in the literature on this problem involves ordinary perception. You visit the zoo in San Diego with your children, and report your knowledge that the animal you are presently observing is a zebra. You also know that if it is a zebra, it is not a cleverly disguised mule painted to be indistinguishable to the ordinary observer from a zebra. You do not know, however, that the animal you are looking at is not a cleverly disguised mule.[14]

[14] See Fred Dretske, "Epistemic Operators", "Conclusive Reasons", *Australasian Journal of Philosophy* 49 (1971), pp. 1007–1023, and "The Pragmatic Dimension of Knowledge", *Philosophical Studies* 40 (1981), pp. 363–378.

Cases such as these are cases that appear to threaten closure principles about knowledge, and a significant literature has built up over the past thirty years or so about such cases.[15] In each of such cases, there is an *ordinary proposition*, the proposition regarding which we naturally ascribe knowledge, and a *lottery proposition*, the proposition implied by the ordinary proposition and regarding which we are not ordinarily willing to ascribe knowledge. It is worth noting, however, that there are approaches to such cases that attempt to accommodate the idea that we know, for example, that we won't be able to afford a mansion and yet do not know that we will not win the lottery. One such approach is Contextualism, according to which the truth-value of attributions of knowledge depends on the standards in place at the context of utterance. Some contexts are high standards contexts and some are ordinary standards contexts, leaving open the possibility that shifts in context can occur to explain how it could be true to say that one knows that one won't be able to afford a mansion and yet not to be true to say that one knows one will not win the lottery. The two most prominent defenders of Contextualism, Stewart Cohen and Keith DeRose, both maintain that one of the benefits of their view is that it preserves closure.[16] The alternative defended by John Hawthorne, which he terms a version of Invariantism, also allows the truth-value of knowledge ascriptions to vary, where the variability traces to the situation of the individual whose knowledge is in question rather than to the situation of the individual ascribing knowledge, as it does for the Contextualist.[17]

By appealing to shifting standards in these ways, both views are able to avoid abandoning a closure principle about knowledge. An appropriate closure principle will be restricted to a situation in which the standards for the truth of knowledge ascriptions remain the same. So when shifting standards explain how an ordinary proposition can be known even though the related lottery proposition is not known, the shifting of standards prevents such cases from being counterexamples to an appropriate closure principle concerning knowledge.

[15] See John Hawthorne, *Knowledge and Lotteries* (Oxford, 2004) for a discussion of the problem and a thorough guide to the literature.

[16] See, e.g., Stewart Cohen, "How to be a Fallibilist", *Philosophical Perspectives* 2 (1988), pp. 581–605; "Skepticism, Relevance, and Relativity", in *Dretske and his Critics* (Oxford, 1991), pp. 17–37; "Contextualist Solutions to Epistemological Problems: Scepticism, Gettier, and the Lottery", *Australasian Journal of Philosophy* 76. 2 (1998), pp. 289–306; Keith DeRose, "Contextualism and Knowledge Attributions", *Philosophy and Phenomenological Research* 52 (1992), pp. 913–929; and "Solving the Skeptical Problem", *The Philosophical Review* 104, 1 (1995), pp. 1–52.

[17] Hawthorne, *Knowledge and Lotteries*.

The point to note in our context of evaluating whether (K-Dist) should be rejected is this. The best case to be made for denying this rule depends on arguing that knowledge is not closed under deduction, and the above theories all present substantive bases on which to deny that the phenomenon used by Nozick to defend failure of closure in fact establishes the conclusion he seeks. In fact, each of these approaches claims to offer a much more attractive account of the phenomenon involving ordinary and lottery propositions, in part because these accounts are consistent with closure principles about knowledge. On the basis of these alternatives, we lack a basis at this point for using Nozick's argument to undermine (K-Dist).

It is worth noting that Invariantism of the sort Hawthorne defends need not be conceived of in terms of a semantical theory that treats the term 'knows' in terms of a set of relata, one of which is a specification of the standards in place regarding the individual in question. Contextualists emphasize various ways in which these standards can be raised, one important item being the salience of the possibility of error.[18] If we think of the concept of knowledge as involving the concept of justification, details of the theory of justification can yield the suggested implication without any recourse to the more complex semantics posited by Contextualist theories. One promising approach to the theory of justification treats justification as a function of positive evidence and known defeaters: justification amounts to adequate positive evidence not ultimately undermined by information one possesses that defeats the justificatory force of the positive evidence. In order to accommodate the Contextualist emphasis on salience of the possibility of error, one need only treat the defeater condition so that concerns by the individual in question about the likelihood of error are themselves defeaters. That is, if the possibility of error is salient for an individual regarding a particular claim they are considering, then the individual has a defeater of the positive evidence possessed for the claim in question. The benefit of such an Invariantist treatment is that the semantics for 'knows' can remain in

[18] Cohen is most direct about the importance of such salience, and this feature of standards-raising is appealed to in a way that generates, to my mind, disastrous consequences by David Lewis in "Elusive Knowledge", *Australasian Journal of Philosophy* 74. 4 (1996), pp. 549–567, some of which are noted in later discussion in the text. According to Lewis, merely reflecting on the question of whether we have met the requirements for knowledge causes the salience condition to raise the standards to the impossible heights, so that such reflection always and everywhere destroys the possibility of knowledge—more carefully, it raises the standards so that every statement of the form 'S knows that p' is false.

accord with a standard epistemological approach to the nature of knowledge in terms of a subtype of justified, true belief.

One perceived benefit of treating these factors semantically rather than in terms of a special feature of the concept of defeat is the capacity to preserve a closure principle about knowledge, for when the semantic factors regarding the standards for knowledge are kept fixed, a closure principle can be maintained without denying that, e.g., we can know we have hands even though we do not know that we are not brains-in-a-vat. It might appear that treating these factors as a special feature of the concept of defeat fails to preserve this perceived benefit, but that is a mistake. The closure principle, on such a theory, will need to be restricted in terms of these special features of the concept of defeat, but a closure principle will still be available. The only difference is that the elements in question will enter into a specification of the closure principle on the one theory whereas they will enter into the semantical treatment of 'knows' on the other theory, and that difference presents no substantive basis for preferring one approach to the other.

In sum, there is a variety of approaches to these puzzles that allow for some closure principle regarding knowledge. One approach is the skeptical approach, which maintains closure and concludes that we do not know any of the ordinary propositions. Another approach is Moorean, an approach which bites the bullet, maintaining that we know lottery propositions because they are implied by ordinary propositions that we know to be true. In between are approaches that find knowledge attributions sensitive to context, either that of the speaker or that of the individual to whom knowledge is attributed or denied. The first two approaches simply deny Nozick's intuition that he knows where he is but doesn't know that he is not a brain-in-a-vat. The latter two approaches agree with Nozick's intuition, but deny that the intuition threatens closure.

So the point to note is that Nozick's intuition, by itself, cannot show that closure must be abandoned. It remains to be seen exactly what kind of closure condition can be defended on these alternative approaches, and whether (K-Dist) would be sustained in its simplicity by a full defense of some closure principle. For example, one promising approach to the question of closure is to begin from the idea that knowledge is closed under competent deduction.[19] This approach needs amending, since in the process of deduction one may acquire additional defeating information either about the premise of the deduction or its

[19] This approach is defended by Timothy Williamson in *Knowledge and its Limits*.

conclusion.[20] (Alternatively, we might put this point in terms of factors arising in the process of deduction that result in the standards for knowledge being raised so that it is no longer true that the premise of the deduction is known.) Once the necessity of this additional complexity is noted, the simplicity of (K-Dist) is no longer defensible. Instead, that rule would have to be replaced by a rule that allowed the use of &-Elimination in knowledge contexts when no additional information arose that would undermine knowledge of the premise or the deduced conclusion. Such complications, however, give no reason for thinking that no adequate closure principle can be found to sustain the inference used in the knowability paradox.

One possible exception here is the contextualist theory of David Lewis.[21] According to Lewis, any reflection on whether we know raises the standards for knowledge to heights impossible to achieve. So, if we deduce a claim with a knowledge operator in it, the standards will have been raised during the deduction so that any knowledge of the premise of the deduction will disappear in the process of completing the deduction. Since one application of the closure principle in the paradox uses &-Elimination on the claim that p is unknown, Lewis's view would prevent the application of a closure principle of the sort described in the text in the context of the paradox. Of course, the premise of the deduction already has a knowledge operator in it, so if Lewis is correct, consideration of the conjunction itself would raise the standards to impossible heights. So Lewis's view implies directly that there are truths that cannot be known, simply in virtue of the fact that some truths have knowledge operators in them. Lewis's approach, therefore, blocks the derivation of a lost distinction between actual knowledge and possible knowledge on the basis of a proof of Fitch's result.

The difficulty is that this aspect of Lewis's view is unacceptable. For example, the following conversation is incoherent on Lewis's view:

FATHER Your best friend is a thief!
DAUGHTER You don't know that! You have no evidence for that.
FATHER Yes I do, I saw her shoplift at WalMart and get arrested by the cops.
DAUGHTER But you can't be sure that it was *her*!
FATHER Yes, I can; I know her well, and it *was* her.
DAUGHTER OK, she's a thief, but maybe she had a good reason for stealing . . .

[20] For more on the needed qualifications, see the debate between Fred Dretske and John Hawthorne, on the question "Is Knowledge Closed Under Deduction?", *Contemporary Debates in Epistemology*, Matthias Steup and Ernest Sosa, eds. (Oxford, 2004).

[21] David Lewis, "Elusive Knowledge."

In this conversation, the daughter acquiesces to the father's claim to know that her friend is a thief, but on Lewis's view, the proper response would be:

DAUGHTER Don't you know that as soon as the concept of knowledge and standards of evidence appear in the conversation in a way that requires us to reflect on them that the standards for knowledge become impossibly high! So you have to acquiesce—haven't you heard: knowledge is elusive!

I suggest that the first dialogue has the more realistic ring to it (and not just because you have to have read Lewis to answer in the second way!), and thus that Lewis's theory cannot do service in an attempt to avoid the paradox in virtue of any defensible features of Contextualist epistemology.

We return then to the earlier point, that whatever complications are needed to a closure principle in terms of additional information acquired in the process of deduction, these complications provide no grounds for thinking that the application of &-Elimination in a knowledge context will be out of order. So if we are to find an argument against closure, we will have to move beyond the phenomenon involving ordinary and lottery propositions to the particular theory Nozick develops and close cousins of it.

COUNTERFACTUALS AND NOZICKIAN EPISTEMOLOGY

Nozick's theory of knowledge clarifies the conditions for knowledge beyond that of true belief by appeal to counterfactuals, and there are four counterfactuals on which a theory of the sort he develops might focus in giving a counterfactual-based account of knowledge. Where 'p' is a sentence, 'B' is the operator 'S believes that', and '$\Box\!\!\rightarrow$' the symbol for counterfactual implication, the four candidates are, with identifying descriptions:

1. $p \Box\!\!\rightarrow Bp$ (the transparency condition)
2. $Bp \Box\!\!\rightarrow p$ (the safety condition)
3. $\sim p \Box\!\!\rightarrow \sim Bp$ (the sensitivity condition)
4. $\sim Bp \Box\!\!\rightarrow \sim p$ (the idealism condition)

Neither the idealism condition nor the transparency condition have been popular in recent epistemology, and for good reason. The transparency condition would limit what we know to things that wouldn't be

true unawares. Yet, there is good reason to think that much of what we know requires attention on our part. The fact that we don't know what is going on around us when we are asleep doesn't imply that we don't know what is going on around us when we are awake and attending to the features of our immediate environment. Moreover, much of what we know about our immediate environment involves subtle features of it that could easily have escaped our attention.

A similar problem affects the idealism condition. We can know what is going on around us when we are awake and attending to the features of our immediate environment, and yet the idealism condition limits what we can know to what would be false when not believed. That would require that anything about our immediate environment that is unchanged when we are asleep would fail to be a candidate for knowledge. It would also require the destruction of subtle features of our environment which we know about, which we ordinarily assume would still be there even if they had gone unnoticed by us.

Such considerations explain the focus of recent epistemology on the other two conditionals, the safety and sensitivity conditions. Nozick puts the sensitivity condition to primary use, while other epistemologists, such as Ernest Sosa, prefer the safety condition.[22] Both approaches are Nozickian in the sense that they distinguish knowledge from true belief on the basis of subjunctive conditionals.

I will begin by considering whether there are good reasons to prefer safety over sensitivity as the preferable form of Nozickian epistemology, since a primary motivation for preferring safety over sensitivity has been that safety is thought to preserve closure. I will argue that this issue is a red herring, for the move from sensitivity to safety is of no help on this issue, and thus that both sensitivity and safety approaches to knowledge threaten (K-Dist). After establishing this claim, we will be in a position to consider whether either approach to the nature of knowledge is sufficiently promising that any appeal to (K-Dist) would be unwarranted.

Closure

A theory of knowledge respects closure when it adopts the idea, in some form, that the logical consequences of what we know are also known.

[22] See Ernest Sosa, "Proper Functionalism and Virtue Epistemology", and "Postscript to 'Proper Function and Virtue Epistemology' ", both in Jonathan L. Kvanvig, ed., *Warrant in Contemporary Epistemology* (Totowa, N.J., 1996), pp. 253–270 and 271–280; and Timothy Williamson, *Knowledge and its Limits* (Oxford: Oxford University Press, 2000).

This rough closure claim is certainly false, otherwise we'd all be logically and mathematically omniscient. Theories that wish to respect closure defend the view by restricting this rough form of the view. One way to restrict it is to hold that when we know p and know p entails q, then we know (or are in a position to know) q.

The project of refining the closure claim so that it is defensible is not a simple and easy project, but there is intuitive plausibility to the general idea, for we often appeal to some kind of closure condition in order to explain why a claim to knowledge is mistaken. When identifying a particular suv as a Ford Explorer, my assertion was challenged by one of my children with the question, "How do you know it's not a Nissan?" The point of the question is that Nissan has a suv with the same body style as the Ford Explorer, and the presupposition of the question is that I could only know that it was an Explorer if I knew it was not a Nissan. This presupposition appears to involve some sort of closure principle about knowledge, one that I accepted in the context, because I replied that I saw the Ford logo on the front hood. Closure is initially attractive precisely because it provides a convenient explanation of the type of discussion this example illustrates.

Nozick noticed and embraced the fact that the sensitivity requirement does not preserve closure with respect to knowledge, using this feature of his account to respond to skepticism. For example, on Nozick's view, it is possible to know that some object is a hand, know that if that object is a hand then no deception by a Cartesian evil demon is occurring, and yet not know that no such deception is occurring. Not many have been impressed with this aspect of Nozick's view, and some have claimed that replacing sensitivity conditions with safety conditions gives hope for upholding closure.[23]

Consider the following two arguments:

1. p entails q	1. p entails q
2. $\sim p \Box\!\!\rightarrow \sim Bp$	2. $Bp \Box\!\!\rightarrow p$
3. So, $\sim q \Box\!\!\rightarrow \sim Bq$	3. So, $Bq \Box\!\!\rightarrow q$

The first column represents the argument that would need to be valid in order for sensitivity to preserve closure, and the second column the argument need for safety to preserve closure. We know that the first

[23] A clear and helpful discussion of these issues can be found in Steven Luper, "The Epistemic Closure Principle", *The Stanford Encyclopedia of Philosophy* (Spring 2002 Edition), Edward N. Zalta (ed.) <http://plato.stanford.edu/archives/spr2002/entries/closure-epistemic/>.

argument fails, on the basis of Nozick's example about being in Emerson Hall and not being a brain-in-a-vat. A bit simpler example to follow concerns having hands and being in a Cartesian evil demon world. Both safety and sensitivity counterfactuals are true: if I had no hands, I wouldn't believe that I do; and if I were to believe that I have hands, I would have hands. My having hands entails that no demon is active who makes it always and everywhere false that anyone has hands. But if there was such a demon, I wouldn't believe that there was, so this belief is not sensitive to the facts even though it is entailed by one that is. Such an example does not constitute a counterexample to the safety requirement, however, for if one were to believe that there is no demon, there would be no demon, (given the operating assumption that the actual world is not a demon world), so this belief is safe as well. So the standard examples used to show that sensitivity is not closed under entailment lend support to the idea that safety is so closed.

The important point to notice here is the way in which sensitivity requires the truth of a knowledge claim to answer to what happens in extremely remote possibilities, such as the remote possibility in which an evil demon is deceiving one. Safety theories, however, maintain that what happens in such remote possibilities is irrelevant to whether one knows. What is relevant, instead, is what happens in close possibilities, possibilities which are very much like the way things actually are. Because of this difference, safety theorists can claim that closure is preserved in many of the kinds of cases that sensitivity theorists use to show that closure is violated.

This evil demon case also highlights one of the features of the logic for counterfactuals. One of the truth-preserving operations on material conditionals is transposition (contraposition), but this rule is one of the standard rules that are invalid for subjunctive conditionals.[24]

Still, concluding that safety is closed under entailment would be incautious. First, note that it would be an amazing coincidence if every case in which sensitivity fails the closure test constitutes a counterexample to transposition on counterfactual conditionals, and yet that is precisely that to which defenders of closure for safety are committed. This intuitive point against the claim that safety involves closure can be further supported using the standard semantics of possible worlds and similarity relations for counterfactuals, for this apparatus makes it very

[24] See, e.g., David Lewis, *Counterfactuals* (Oxford, 1973).

easy to construct a countermodel to both arguments above. Stipulate that in the actual world α the following is the case:

$$\text{True-in-}\alpha\text{: p, Bp, q, } \sim\text{Bq, p entails q.}$$

This stipulation guarantees the truth of the first premise of the above arguments. Further stipulate that in all worlds W most similar to α, we have:

$$\text{True-in-W: } \sim\text{p, } \sim\text{Bp, } \sim\text{q, Bq.}$$

On this model, the premises of both arguments are true, and each conclusion is false. For the first argument, the conclusion is false since in the most similar world where q is false, Bq is true. For the second argument, the conclusion is false as well, since in the closest world in which Bq is true, q is false.

In order to resist this countermodel, one will have to show some feature of the standard semantical treatment of counterfactuals that makes this model impossible. It is not enough simply to argue that the model is unusual, that in ordinary cases, some W will be less similar to α than some other world. To rule out such cases, one will need to be able to show that it is impossible for the situation to be as described above, and there simply are no semantical resources in the standard semantics for counterfactuals to sustain such a claim.

This result should not surprise us. We know that contraposition fails for counterfactual conditionals, but that fact is compatible with such contrapositives nearly always having the same truth-value. So it shouldn't surprise us if the contrapositives defining safety and sensitivity have the same truth-value in some special context, unless we had reason to think that such a context was one where counterexamples to contraposition were likely to occur. It is hard to find any basis for such a suspicion in the special context of closure issues, so we shouldn't expect safety to be immune from whatever closure failures bear on sensitivity.

One may still resist such a conclusion by pointing out general difficulties with the standard semantics for counterfactuals, and claiming that more intuitive readings of subjunctive conditionals on a case-by-case basis fail to produce counterexamples. The upshot of such a position is that even though we can't see some intuitive connection between closure difficulties for sensitivity and counterexamples to contraposition, the case-by-case examination shows that there is nonetheless a connection between them. If these claims were true, that would give some reason for maintaining the closure claim about safety and

blaming the standard semantics for formal failures such as the one described above.

The problem is that the case-by-case claim made above is false. Suppose Bill is in fake barn country where all the fake barns are red, and the rare remaining real barn is green. Then consider the claim that the building in the pasture is a green barn, and the related claim that the building in the pasture is a barn. Suppose that Bill is asked whether the object in question is a green barn, and Bill answers, "Yes", believing that the object in question is a green barn. His belief is safe: anytime he forms such a belief in an environment relevantly like the present one, he will be correct. Bill believes also that the object in question is a barn. This belief, however, is not safe: If he were to form such a belief in an environment relevantly like the present one, he would be mistaken.[25] Hence, safety does not preserve closure.

So safety theorists have no more right to claims about closure than do sensitivity theorists; the difference is that defenders of sensitivity have admitted failure of closure and safety theorists such as Sosa deny that their view has this implication. I think we can explain this difference in terms of a confusion between two closely related alternatives to sensitivity. I think defenders of closure using the language of safety have confused safety with a closely related view that is best expressed using the language of reliable indication. If one begins by thinking about sensitivity and safety in terms of subjunctive conditionals, one may be tempted to think of reliable indication as expressible in terms of a subjunctive conditional where the reliable indicator is the belief itself. Thinking in this way generates the following closure argument, which I juxtapose with the failed safety closure argument:

1. p entails q	1. p entails q
2. Bp $\Box\!\!\rightarrow$ p	2. Bp $\Box\!\!\rightarrow$ p
3. So, Bp $\Box\!\!\rightarrow$ q	3. So, Bq $\Box\!\!\rightarrow$ q

Given this first argument, one could then claim that having such a reliable indicator of the truth of q will put one in a position to know that q is true, needing only for one to add the belief that q in order to come to know that q.

Notice that this first argument, the argument representing closure of reliable indication, has no counterexamples, either intuitive or formal.

[25] I thank Stew Cohen for this point, who noted that the example is borrowed from what are now referred to as Saul Kripke's "Nozick-bashing" lectures at Princeton in the mid-1980s.

For the second premise to be true, we must consider some situation in which both Bp and p are true, and the first premise requires that in any such situation q will be true as well. Thus, such a situation cannot be one where Bp is true while q is false.

Reliable indication, however, is a different condition from the safety condition, and defenses of closure under reliable indication do not support closure under safety. So no mere substitution of a safety condition for Nozick's sensitivity condition will save the view from the untoward failure of closure. Instead, what will be needed is a replacement of the kind of theory Nozick favors with a version of reliabilism, if one wishes to follow the lead found in the pair of contrasting arguments above. In doing so, one might attempt to highlight the resemblances between the two views by clarifying reliable indication in terms of counterfactuals, but one will be hard pressed to rely on the resemblances for long. For example, if believing that p is a reliable indicator of p, why think that it is the only kind of reliable indicator that can function in one's theory of knowledge? Why not any reliable indicator, whether a belief or not? For example, perhaps sensory experience itself is a reliable indicator of the truth of certain propositions. If so, why should the theory rule out this kind of reliable indicator as suitable in the account of knowledge?

Reliable indicator theorists may respond by claiming that the reliable indicator view is simply the deeper explanation of the truth of the safety view, in the following way. They may say, "If something is a reliable indicator of p, then believing p is also a reliable indicator of p, so reliable indication explains what is correct about the safety view." The idea behind such a claim is to keep the counterfactual relationship between beliefs and propositions central to the view, so as to maintain in some form an analogy between the safety view and the reliable indication view. The problem is that such a response generates a new violation of closure, for it commits the reliable indicator theorist to the following argument form:

1. p entails q
2. X $\Box\!\!\rightarrow$ p
3. So, X $\Box\!\!\rightarrow$ q
4. So Bp $\Box\!\!\rightarrow$ p (using "if X is a reliable indicator of p, then so is believing p.")
5. So, Bq $\Box\!\!\rightarrow$ q (again using "if X is a reliable indicator of p, then so is believing p", with q as the value for p)

This argument implies a restricted form of closure for safety: it says that if you have a reliable indicator for p, then the belief that p is safe

and so is any belief entailed by the belief that p. The barn example above shows that one cannot endorse such a view: even though (4) is true when p is about green barns, (5) is false when q is about barns. (1)–(4) may all be true and (5) still false.

So if an interest in preserving closure leads one to endorse a reliable indication theory, the best thing to do is simply embrace reliabilism and stop trying to pretend that one is defending a view of the same family as Nozick's. Restricting the reliable indicators to beliefs makes the theory insufficiently general, and once one generalizes, no basis is left for maintaining that the reliable indicator view is some small variant of the safety view.

Though our primary concern is with the question of closure and the way in which Nozickian theories threaten it, it is worth noting that not much comfort can be taken by defenders of closure by attempting to substitute the reliable indicator theory described above for a safety or sensitivity view. The problem is that this version of reliabilism is too problematic for that. Put simply, it has strongly counterintuitive consequences. Suppose X is a reliable indicator of p, and consider whether it is possible for X to obtain (and be a reliable indicator of p) and p to be false. If we consider the ordinary probabilistic conception of reliability, such would seem possible, but if reliable indication is defined in terms of the counterfactual *If X were to obtain, p would be true*, we get a different result. In that case, the obtaining of X as a reliable indicator of p is incompatible with the falsity of p.

We must be careful regarding this point, for I am not claiming that the combination of the obtaining of X and p is impossible. A counterfactual relationship between two items does not imply such an impossibility. What I am claiming is something different, which we might formulate in terms of the language of possible worlds. Suppose we find a world where X obtains and p is false. In such a world, the counterfactual if X were to obtain, p would be true is also false, and hence X is not a reliable indicator of p in that world. So a counterfactual theory of reliable indication has the following troubling consequence: the obtaining of a reliable indicator of p is a logical guarantee of the truth of p, for no falsehood is ever reliably indicated to be true.

Compare this result with ordinary process reliabilism, a crude version of which claims that knowledge requires production of belief by a process that generates true beliefs most of the time. On such a view, one can maintain that reliable beliefs are not always true beliefs, but on the reliable indicator view above, that claim cannot be maintained, if we

mean by "reliable belief" a belief the content of which is reliably indicated to be true. Ordinary process reliabilism accords better with our ordinary use of the language of reliability, and this result gives us some evidence for thinking that the counterfactual account of reliable indication at least partially stipulative.

The conclusion to draw is that substituting a safety condition for Nozick's sensitivity condition cannot be motivated by an attempt to preserve closure, for safety does no better on that score than does sensitivity. This conclusion leaves some hope for those who wish to follow in Nozick's path in hopes of avoiding the knowability paradox, but we are not yet in a position to draw the conclusion that Nozickian epistemology is a haven against the paradox. Defenders of safety and closure have not yet breathed their last.

Further Complications on Behalf of Safety and Closure

A natural move at this point for those who wish to focus on safety rather than sensitivity is to wonder whether the closure of safety can be rescued by inserting talk of probabilities into these approaches. Instead of the above conditions, a theorist might say that a belief is safe when most of the time the claim in question is true in the presence of belief,[26] and one might say that a belief is sensitive when most of the time one would not hold the belief if the claim were false.

Both of these claims contain a scope ambiguity, depending on whether the probability operator is taken to govern the entire conditional or only the consequent of the conditional. Disambiguating gives us the following:

Wide scope probable safety: $Pr(Bp \Box\!\!\rightarrow p)$
Wide scope probable sensitivity: $Pr(\sim p \Box\!\!\rightarrow \sim Bp)$
Narrow scope probable safety: $Bp \Box\!\!\rightarrow Pr(p)$
Narrow scope probable sensitivity: $\sim p \Box\!\!\rightarrow Pr(\sim Bp)$.

The first thing to notice is that closure does no better on the narrow scope probabilistic readings than on the non-probabilistic versions of safety and sensitivity. If Bill were to believe that the object in question is a barn, in all likelihood he'd be wrong—after all, the real barns in the area are extremely rare. So if our goal is to preserve closure, neither the

[26] It is interesting to note that Sosa makes such a move in "The Place of Truth in Epistemology," in Zagzebski and DePaul, eds., *Intellectual Virtue: Perspectives from Ethics and Epistemology* (New York, forthcoming).

move from sensitivity to safety, nor the move from narrow scope probabilistic sensitivity to narrow scope probabilistic safety will help.

What about wide scope readings? Such readings generate the following arguments:

1. p entails q
2. Pr(∼p □→ ∼Bp)
3. So, Pr(∼q □→ ∼Bq)

1. p entails q
2. Pr(Bp □→ p)
3. So, Pr(Bq □→ q)

These arguments are harder to assess for in large part because of the variety of interpretations proposed for the Pr operator, and the accompanying difficulties for each. For example, if we take a frequency approach to probability, these arguments will fail since it will be a contingent matter whether the claim within the scope of the probability operator in the conclusion is true most of the time when the premises are true. A similar point holds for other interpretations of probability, such as subjective theories and nomological ones, namely that the truth of the conclusion in the face of the truth of the premises will be contingent at best. This point undermines the validity of these arguments, for arguments are valid only when the premises necessitate the conclusion.

The only approach that might allow the validity of these arguments is also one of the more mysterious ones, the logical theory, which maintains that any true probability judgment is necessarily true. If the conclusions are true, the above arguments will be valid since any argument with a necessarily true conclusion is valid.

Even so, there is cause for concern here. First, notice that even given this interpretation of probability, it is far from obvious how such an interpretation can be used to show that either of the arguments above is valid. Second, recall that the defender of safety is claiming superiority for that view over the sensitivity view on the basis of closure results. So even if the second argument above is valid, that will not be a reason to favor safety over sensitivity unless the first argument is invalid. How would a defender of safety show that the first argument is invalid (assuming the logical theory of probability)? The best that can be done, I think, is to show that the information in the premises leaves the likelihood of the embedded sentence of the conclusion utterly inscrutable, i.e., they provide us with no information from which we can obviously conclude that the conclusion must be true. The same holds, however, for the second argument. So the challenge to the defender of safety is that it is not enough to maintain that the second argument is valid. Some

difference must be found between it and the former argument if the superiority of safety is to be sustained.

Moreover, the move to any probabilistic reading raises a further problem for the idea that some safety condition explains the difference between knowledge and true belief. Consider how beliefs in lottery cases can be probabilistically safe. No adequate theory of knowledge can imply that we know that our ticket in the lottery will lose, no matter how large the lottery is, but if some probabilistic safety condition distinguishes knowledge from true belief, then such knowledge is possible. In the case of a relatively large lottery where one believes that one's ticket will lose, one's belief is safe and probably safe as well. The lesson is that probabilistic safety theorists will need some further condition to give an adequate account of the nature of knowledge.

The probabilistic move is not the only complication that has been introduced to avoid the problem of failure of closure for safety conditions. Sosa's latest account of safety is:

> S's belief that p is safe iff S holds it on a basis that this belief would not have without being true.[27]

He offers this new account to avoid the Kripke red-fake/green-barn counterexample. Here's how it works with that case. S believes that there's a green barn in the field, and deduces that there's a barn in the field. In that case, the first belief is safe, having a perceptual basis, and the second is safe as well, having a deductive basis. Had S simply formed the perceptual belief that there's a barn in the field, however, it would not have been safe.

Even so, Sosa believes that we don't know that the object in the field is a barn. We have that intuition, according to Sosa, because of the way our perceptual phenomenology works: we are appeared to barn-in-the-field-ly, and also are appeared to greenly, so that the basis of the belief that there's a barn in the field is the first appearance state. Were the phenomenology different, so that we are appeared to green-barn-in-the-field-ly, no violation of safety would occur either for the belief that there's a green barn in the field or that there's a barn in the field.

To appreciate Sosa's proposal, we must first distinguish between what is in our visual field and what we attend to within this field. It is possible for one's visual field to contain a green barn, and yet one attend only to

[27] John Greco, ed., *Ernest Sosa and his Critics* (Boston, 2004), p. 292.

the barnishness of the experience. But Sosa needs more than this point. He needs there to be a distinction between unified noticing or attending, where one attends to the green-barnishness of the experience, and conjunctive noticing or attending, where one attends to the barnishness of the experience and in a separate attending, to the greenishness of the experience.

It is here that the account is less than fully convincing. We can easily make sense of sequentially noting the barnishness of an experience and the color of it, and also of simultaneously noticing two aspects of our experience, where one part has a property and another part has a different property, such as when there are two objects in my visual field. But can we abstract away from color in a way that would result in a simultaneous, independent noticing of the greenish character of my experience of an object and the barnishness of that same object? I doubt that such is possible.

Moreover, even if it were possible, it is hard to see why one phenomenological account over the other should somehow indicate a difference in what is known. Take two barn-believers, one of whom has the conjunctive abstraction experience Sosa indicates, and the other of whom has a unified experience. One might think that a difference in knowledge should be expected, since one of the two will have to draw an inference that the other does not have to draw, but that is mistaken. There is no such limitation on cognitive architecture, for belief forming processes that are non-inferential can go from either input to the output belief in question. Given that the issue is simply one of the kind of cognitive architecture in question, the mere difference in phenomenology should be insufficient to establish, by itself, a difference on the issue of whether knowledge is present.

We should judge, therefore, that these complications do not help, but leave us rather at the same point we were at the beginning of this section, which is that the threat to (K-Dist) arising from a Nozickian approach to knowledge in terms of counterfactuals is not assuaged by changing the focus from a sensitivity condition for knowledge to some other counterfactual condition. The fact is that we can't count on the differences between safety and sensitivity conditions to provide a means of escaping the threat to closure posed by Nozickian approaches within the theory of knowledge, and hence we will not make progress in assessing the threat to (K-Dist) by pursuing the question of which subjunctive conditional account of knowledge can preserve closure, if any.

DEEPER PROBLEMS FOR SAFETY AND SENSITIVITY CONDITIONS

The dominant response to the threat to closure posed by Nozickian epistemology has not been positive, especially when the implications of the theory are that one can know that something is a green barn, but not that it is a barn. Such a position is truly deserving of DeRose's appellation, "the abominable conjunction."[28] My own intuitive response is not nearly as hostile to the suggestion that we can know we have hands even though we do not know that we are not brains-in-a-vat being stimulated to experience the world in precisely the way we are actually experiencing it. I am not saying that I find the latter conjunction unproblematic. I am saying, though, that I find it less problematic than the former conjunction, and it may be that there is a way to develop a safety or sensitivity approach to reduce the kinds of failures of closure to those that are the least problematic kind.

There is a further consideration to note as well, found in the difficulties encountered in recent attempts to formulate a defensible closure condition.[29] In light of the difficulties in formulating an acceptable closure condition, we should be suspicious of categorical rejections of a Nozickian theory based only on closure considerations. So, if we are to conclude that Nozickian epistemology and its close cousins pose no threat to (K-Dist), we need arguments beyond merely the intuitive idea that there is something wrong with a theory that implies that one can know that something is a green barn without knowing that it is a barn.

I will argue, however, that there are deeper problems with Nozickian approaches to the nature of knowledge and that because of these failures, one cannot appeal to Nozickian approaches to find fault with (K-Dist). To the extent that a Nozickian approach to knowledge is inadequate, to that extent it lacks the resources needed to cast doubt on (K-Dist). I will argue that the extent of the inadequacy of a Nozickian approach has two aspects. First, I will argue that appeals to safety or sensitivity can play, at most, only a limited role, that neither sensitivity nor safety approaches are suited for explaining the nature of knowledge, but instead can explain only part of what we know. The limited roles I will

[28] Keith DeRose, "Solving the Skeptical Problem", *The Philosophical Review* 104 (1995), pp. 1–52.

[29] For a comprehensive account of the issues involved in this problem, see John Hawthorne's *Knowledge and Lotteries* (Oxford, 2004).

argue for are these: sensitivity is more naturally suited to explaining perceptual knowledge, whereas safety is more naturally suited for explaining statistical or inductive knowledge, and neither is well suited for explaining the other category. I will also argue that there are cases where no appeal to safety or sensitivity can help at all in explaining the difference between knowledge and its lack. After presenting these arguments, I will urge that we have arrived at the point where it is reasonable to conclude that Nozickian epistemology lacks the resources to threaten the use of (K-Dist) in Fitch's proof.

Consider, first, the way in which safety and sensitivity are more at home in cases of perception and statistical or inductive knowledge, respectively. Sosa argues against the sensitivity requirement and in favor of the safety requirement in just this way. He imagines a case where a person has dropped one's garbage down a chute in an apartment building, and that person believes that the garbage will soon be in the basement. The long track record of correlation between the two events makes the belief safe: if one believes, in this situation, that the garbage will soon be in the basement, it would be true. Notice, however, that the belief is not sensitive. If the garbage would not soon be in the basement, it likely would somehow have gotten caught in the chute, in which case one would still believe that the garbage will soon be in the basement.

What is instructive here is that Sosa's defense appeals to a case of inductive knowledge, and sensitivity is not well suited to such contexts. It is equally instructive to notice that things go in the other direction in perceptual cases. Suppose one is color blind to the extreme: everything looks brown. Suppose also that in one's environment, everything is a shade of brown. Then, when looking at a brown thing, one's belief is safe: if one believes it is brown, it would be brown. But one's belief is not sensitive, and its insensitivity is the important point here, for such a person does not know the color of the thing he is looking at even if the environment is so unbelievably friendly to him that he is always right.

In perceptual cases, our attention generally focuses on powers of discrimination, and here sensitivity is central. A substantial track record of truth may show that a belief is safe, but it won't show that it counts as knowledge when one lacks the discriminatory capacity to differentiate truth from error. Lacking such a capacity makes one's beliefs insensitive, and this feature dominates the safe character of the belief, for in such cases, knowledge is lacking.

If I am right, neither sensitivity nor safety are necessary conditions for knowledge. Sensitivity is more at home in the realm of perceptual

knowledge and safety in the realm of inductive or statistical knowledge. Of course, it is open to a theorist to opt for one of the two, and simply deny that there is knowledge of the other sort. Such a route would be a very difficult one for a defender of the safety condition, for that would require holding that inductive knowledge is possible even though perceptual knowledge isn't. More likely is the other route: to hold that perceptual knowledge is possible, but statistical or inductive knowledge is not. I will not pursue either line here, except to point out what I take to be the severest counterexample to the idea that inductive knowledge is not possible. One of the deepest aspects of the human condition is coping with our own impending death, and among the most important items of self-knowledge that we have is that we will die. Only theory-clouded judgment can maintain that I can know what color the sky is but not that I will die.

The other option is to bifurcate one's theory of knowledge, maintaining that one of the two counterfactual conditions we've examined always explains the difference between knowledge and true belief. This suggestion will not work either, and this leads to the second point I wish to argue for in this section—namely, that there are distinctions involving knowledge and its lack which neither safety nor sensitivity can address. Such cases can be found routinely in the literature surrounding the Gettier problem, one of which is the following. You see Tim steal a book from the library, and believe as a result that Tim stole the book. You report Tim to the police, and when they go to Tim's home to arrest him, Tim's mother swears his innocence. She says that Tim has a twin brother, Tom, who is a kleptomaniac, and that Tim is on a safari expedition in Africa. Where is Tom? She claims to have no idea, of course.

This case splits into two. In one case, the police have no past history with Tim's mother. Her claims are persuasive and she appears honest, so the police must check out her story before proceeding with their plans to find Tim and arrest him. In the other case, the police have a long history of acquaintance with Tim's mother, knowing that she has lied repeatedly to save Tim from trouble. In fact, she has used this precise lie in the past, and the police have investigated the story and learned that Tim does not have a twin but that he is an only child. In such a case, they laugh off the tale and carry on with their plans to arrest Tim.

In the second case, both you and the police have knowledge about who stole the book that is unaffected by the mother's testimony, but not in the first case. In both cases, the mother's testimony is a defeater of

your evidence that Tim stole the book, but it has differential significance in the two cases. In the first case, it undermines your knowledge, since it is a possibility that the police cannot rule out. But in the second case, it doesn't undermine your knowledge.

If this particular case isn't convincing, imagine different contexts for the mother to say what she says. Imagine, for example, that the mother's testimony is part of a theater production in which she was acting. Or imagine her saying it in her sleep, as part of a dream, or saying it to convince herself: "He *has* a twin brother! He really does! Tim's in Africa, he *really* is!" And imagine any of these possibilities with no one around at all to hear what she says. Knowledge can't be so fragile that it disappears just because a normally competent speaker denies what we believe.

Notice that neither safety nor sensitivity can explain the difference between the two cases. In both cases, the mother is failing to tell the truth, and whatever is the proper account of the relevant difference in the two cases, that difference has nothing to do with the safety or sensitivity of your belief.

Furthermore, notice that there will be cases like this one for any type of belief one wishes to consider. The above case involves a perceptual belief, but it could just as easily have been a statistical or inductive belief, a belief based on memory, or a testimonial belief.

Safety and sensitivity may be part of the story of what distinguishes knowledge from true belief, but they are not enough of the story to fulfill the promise of Nozickian epistemology in the present context of giving a good reason for denying (K-Dist). First, the denial of closure is not itself a virtue of such approaches, and is taken to be a reason not to adopt such an approach or to find a particular version of this approach that is not anti-closure. Second, the extent of the inadequacy of a Nozickian approach must not be underestimated, for the inadequacies identified here show that it would be totally unwarranted to rely on the hope for some future development of the approach to avoid these problems and reinstate grounds for objecting to (K-Dist).

The above discussion is sufficient for blocking any rejection of (K-Dist) through appeal to Nozickian epistemology, but there is another way as well to show that Nozick's theory is of little help to those wishing to avoid the knowability paradox. For even if we agree with Nozick that (K-Dist) is an invalid rule of inference, it is nonetheless possible to acquire knowledge by deducing it from previously known information. Nozick agrees with this point, and offers a theory as to what makes the difference between cases of deduction in which knowledge is not

transmitted and cases in which it is.[30] The crucial feature of his account for present purposes is, where q is deduced from p, the following condition is the additional necessary component in order for knowledge of the premise to be transmitted to knowledge of the conclusion:

> Transmission of Knowledge Principle (TKP): $\sim q \,\Box\!\!\rightarrow \sim$(S believes that p).

In the context of the knowability paradox and the use made of (K-Dist) in it, the first point to note is that the following sequents are not licensed by (TKP):

$$\mathrm{K}(p\&q) \vdash \mathrm{K}p$$

$$\mathrm{K}(p\&q) \vdash \mathrm{K}q.$$

The following sequents, however, are licensed by the claim that (TKP) is all that is needed beyond competent deduction for knowledge to be transmitted from premises to conclusion:

$$\mathrm{K}(p\&q),\ \sim p \,\Box\!\!\rightarrow \sim \mathrm{B}(p\&q) \vdash \mathrm{K}p$$

$$\mathrm{K}(p\&q),\ \sim q \,\Box\!\!\rightarrow \sim \mathrm{B}(p\&q) \vdash \mathrm{K}q,$$

where 'B' is read "it is believed that".

This principle allows a way to patch up the demonstration of the knowability paradox even if (K-Dist) is rejected. Crucial to patching up the demonstration is the truth or falsity of the following instances of TKP:

$$\sim p \,\Box\!\!\rightarrow \sim(\text{it is believed that } (p\&\sim \mathrm{K}p)) \qquad (1)$$

$$\sim \mathrm{K}p \,\Box\!\!\rightarrow \sim(\text{it is believed that } (p\&\sim \mathrm{K}p)). \qquad (2)$$

For if (1) and (2) are true, then from K(p&∼Kp), we can deduce Kp and K∼Kp, from which a contradiction quickly follows by KIT. For (1) to be false, i.e., to block the inference that (K-Dist) licenses, it must be false that close situations in which p is false are situations in which the relevant cognizers believe that p is true but that they do not know that it is true. Note first that it is not impossible to believe that some claim is true and that you do not know it: Pyrrhonists claim that such is the case for all our beliefs. Even so, it is highly unusual to find oneself in such a position; that is, it is highly unusual to believe that certain particular claims are

[30] Nozick, *Philosophical Explanations*, p. 231.

true but that you lack knowledge of them. Moreover, it is obvious that there will be no general argument stemming from the precise content of $p\&\sim Kp$ that will force such a pyhrronist belief on everyone. But if there is not, then there is no reason to think that close worlds in which p is false will fail to be worlds in which such pyhrronist beliefs are not held. So, there is no reason whatsoever for thinking that (1) and (2) will be generally false; in fact, the unusual nature of pyrrhonist beliefs suggests that there is every reason to think that instances of (1) and (2) will be generally true. However, if they are generally true, then we will be able to find possible worlds containing possible individuals for which the inference licensed by (K-Dist) is valid, precisely because the relevant instances of (1) and (2) above are true.

In sum, then, the Nozickian response to the paradox simply will not work. But perhaps it is the details of Nozickian epistemology that are at fault here. Perhaps if we were to consider others who reject the claim that knowledge is closed under known logical implication, we will find resources for blocking the paradox.

Is Closure Necessary?

There is a stronger point to be made here as well. There is a way of formulating the knowability paradox due to Timothy Williamson that allows the demonstration to proceed without appeal to (K-Dist) at all.[31] Suppose we say that a claim p is *conjunctively unknown* if no conjunction is known of which p is a conjunct. Some truths are conjunctively unknown, and all truths are knowable, let us suppose. Yet, no one can know the following: p is true and it is conjunctively unknown that p is true. For if someone did, it would not be conjunctively unknown that p is true. More formally,

$$\text{p is conjunctively unknown iff } \sim \exists qK(p\&q). \quad (1)$$

Assume that p is an arbitrarily chosen conjunctively unknown truth, i.e.,

$$p\&\sim\exists qK(p\&q). \quad (2)$$

To show that knowledge of (2) is impossible, assume

$$K(p\&\sim\exists qK(p\&q)). \quad (3)$$

[31] Timothy Williamson, "Verificationism and Non-Distributive Knowledge", *Australasian Journal of Philosophy* 71.1 (1993), pp. 78–86.

Let $r = \sim\exists qK(p\&q)$; then an alternative way of expressing (3) is

$$K(p\&r), \tag{4}$$

from which it follows that

$$\exists qK(p\&q), \tag{5}$$

by existential generalization. Since (2) follows from (3), however, (5) can't be true because it contradicts the second conjunct of (2). We are thus able to discharge line (3), yielding

$$\sim K(p\&\sim\exists qK(p\&q)). \tag{6}$$

Since (6) depends on no assumptions, by the Rule of Necessitation we get:

$$\Box\sim K(p\&\sim\exists qK(p\&q)), \tag{7}$$

to which we apply one of the Dual rules to yield

$$\sim\Diamond K(p\&\sim\exists qK(p\&q)). \tag{8}$$

By the knowability of all truth, since (2) is true, it can be known:

$$\Diamond K(p\&\sim\exists qK(p\&q)). \tag{9}$$

So we must abandon either the claim that all truths are knowable or the claim that some truths are conjunctively unknown.

This argument relies on KIT (to prove line (6) in a way that does not depend on any assumptions, we relied on the fact that (3) implies (2)) but not on (K-Dist). Moreover, the paradoxical result here is a knowability paradox, the only difference being the addition of the adverb 'conjunctively' to the second assumption. The same paradoxical phenomenon of a lost distinction between actual known truth and possible known truth occurs here, since we must hold that there is no logical distinction between what is conjunctively knowable and what is conjunctively known.

The proper conclusion to draw, then, is that there is no escape from the paradox by rejecting (K-Dist). Theories that deny closure are generally not successful in blocking Fitch's proof, and they are not especially plausible theories of knowledge anyway. Furthermore, even if we sidestep the plausibility issue, the paradox can still arise in a slightly amended form, showing that there is not much of a future in trying to block the paradox by questioning the (K-Dist) rule.

CONCLUSION

Thus, there are several reason for pessimism about the prospects of avoiding the paradox by challenging the special rules for the K operator. First, (KIT) is about as well entrenched a principle as one will find anywhere in philosophy, and for good reasons. Second, the arguments against (K-Dist) depend either on a faulty association between anti-skeptical motivations and closure, or they depend on theories of knowledge that are problematic. Finally, the essential elements of the paradox can be preserved by defining the concept of conjunctive knowledge, yielding the same paradoxical results without using the distribution principle at all. The case against the possibility of escaping the paradox by following this route is, therefore, airtight. If there is a way of escape it simply must be found elsewhere.

5

Reservations about the Underlying Logic

The last way to resist the proof of Fitch's results that underlie the paradox is to find problems with the logic used in the demonstration. The logic assumed is classical logic supplemented with modal and epistemic operators of the usual sort. The primary vantage point from which this logic has been challenged is that of intuitionism. In this chapter, we'll explore the challenge to the proof presented by intuitionism and the prospects of this viewpoint for avoiding the paradox that results.

INTUITIONISM AND ITS IMPLICATIONS

The distinctive features of intuitionistic logic derive from its treatment of negation, motivated by intuitionism in mathematics which limits proofs of existential statements to *constructive* proofs of such. A constructive proof of an existential claim cannot proceed via a reductio, but must instead be derived by generalization from a proven instance. The result of this restriction for a treatment of negation is that the rule of Double Negation is abandoned, and with it the Law of Excluded Middle, which is interderivable intuitionistically with Double Negation.

Because of this heritage in the philosophy of mathematics, some intuitionists, such as Crispin Wright, balk at the second premise of the paradox.[1] Wright claims that such an existential claim would be unacceptable to anyone with constructivist inclinations (since one cannot produce a provable instance of it), and posits instead that the truth aimed at in (2) is not the existential sentence but rather the denial of the universal claim, i.e., $\sim\forall p(p \rightarrow Kp)$.

[1] Crispin Wright, *Realism, Meaning, and Truth*, 2nd ed (New York, 1993).

Such a maneuver cannot help the anti-realist avoid Fitch's result. In that derivation, certain assumptions are made, from which a contradiction is derived. The assumptions on which the derivation depend include the claim that all truths are knowable and that some truths are not known. Anyone facing this contradiction has a choice of which assumption to discharge, and if one is a committed anti-realism, understood in terms of a commitment to the knowability claim, one will have no choice but to discharge the assumption that some truths are not known. At no point in this procedure does the question ever arise as to whether one accepts, even provisionally, the claim that some truths are unknown: it is merely an assumption for purposes of the proof.

In this context, the important point to note is that in the derivation of the contradiction in question, no distinctively classical rules of inference are used. Rather, all first-order rules used are acceptable to both classicists and intuitionists. The rules of which the intuitionist disapproves—Double Negation, Law of Excluded Middle, and the interdefinability of the quantifiers–play no role in the derivation of the contradiction. Instead, the only place disputed rules are used is in the discharge step, where the claim that all truths are known is inferred on the basis of the denial that some truths are unknown.

We therefore have two aspects of the logic of the proof to consider. One aspect involves what is to be learned from deriving the contradiction in question, for it is only in the discharge of the assumptions that any distinctively classical rules are employed. The other aspect is the intensional (i.e., modal and epistemic) context of some of the steps in the proof, since intuitionists might be troubled by intensional inferences that would not bother classicists. Let us look at the modal issues first, since they can be dispatched more quickly.

MODAL PRINCIPLES IN THE INTUITIONISTIC CONTEXT

In addition to intuitionistic reservations about first-order rules used in the proof, one might also express intuitionistic qualms about the interpretation of the possibility operator. To this point, we have said very little about the notion of possibility involved in the claim that all truths are knowable, and some concerns by anti-realists have been voiced that may lead us to think that silence is not the best policy. There are two points to note here. First, in one version of the proof from Chapter 1, we

employed a dual rule relating necessity and possibility, of which there are four versions:

$$\Box p \dashv\vdash \sim \Diamond \sim p \text{(Dual)}$$

$$\Box \sim p \dashv\vdash \sim \Diamond p \text{(Dual)}$$

$$\sim \Box p \dashv\vdash \Diamond \sim p \text{(Dual)}$$

$$\sim \Box \sim p \dashv\vdash \Diamond p \text{(Dual)}.$$

Several authors have questioned whether such rules could be acceptable to intuitionists. Intuitionists notoriously refuse to allow all of the related rules for existential and universal quantifiers, which are:

$$\forall x \phi \dashv\vdash \sim \exists x \sim \phi \ (\forall\exists\text{Dual})$$

$$\forall x \sim \phi \dashv\vdash \sim \exists x \phi \ (\forall\exists\text{Dual})$$

$$\sim \forall x \phi \dashv\vdash \exists x \sim \phi \ (\forall\exists\text{Dual})$$

$$\sim \forall x \sim \phi \dashv\vdash \exists x \phi \ (\forall\exists\text{Dual}).$$

In particular, intuitionistic logics refuse to countenance the third and fourth versions of this rule. This point is important in the modal context since one common way of understanding the notions of possibility and necessity is in terms of universal and existential quantification over possible worlds: $\Box p$ means "in every possible world p is true", and $\Diamond p$ means "in some possible world p is true" on this approach. Given such an understanding, it would not be surprising to find intuitionists objecting to the dual rules regarding these modal operators.

In addition, intuitionists may have a general suspicion about the dual rules for modal operators arising from their first-order theory. Intuitionists identify truth with provability, which is a modal notion, and there is a way to read double negation as a kind of possibility notion, since the double negation of a formula says or implies that no proof is available that the formula is unprovable.[2] So intuitionists will have multiple grounds for concern about the dual rules here.

A second concern is raised by Neil Tennant, who claims that the anti-realist needs some interpretation of the $\Diamond$K operator that is factive, so

[2] For discussion of this issue, see David DeVidi and Graham Solomon, "Knowability and Intuitionistic Logic," *Philosophia* 28 (2001), pp. 319–334.

that $\Diamond Kp$ implies p.[3] Others have taken this claim as grounds for thinking that the notion of possibility that intuitionists employ must be quite different from that employed in usual modal systems such as S4 and S5.[4]

There are good reasons to focus more on first-order worries, however, rather than on concerns about the notion of possibility used in the paradox. First, no intuitionist can get much mileage out of denying that there is a respectable notion of logical possibility on the interpretation of which $\Diamond Kp$ fails to be factive.[5] It may be that there is another notion of possibility that yields factivity, but it would be an objectionable pattern of argument to deny the existence of other notions of possibility simply because one needs there to be one such factive notion. For example, even though it is false that George Washington was born in New York, it might have been true, and might have been known to be true (e.g., by his parents). To account for such a case, there will have to be a non-factive reading of "$\Diamond$K(George Washington was born in New York)", even if there is another notion of possibility on which $\Diamond K$ is factive. Furthermore, the paradoxicality of the derivation is more troubling the weaker the notion of possibility it employs, for if there were available a reading of $\Diamond Kp$ that is factive, it would not be terribly surprising to find that this factivity resides in the redundancy of the possibility operator in the presence of the knowledge operator (much as the characteristic S5 axiom canonizes a similar redundancy for the possibility operator in the presence of the necessity operator, i.e., $\Box p \dashv\vdash \Diamond\Box p$). Should the factivity in question reside in this kind of redundancy, then $\Diamond Kp$ would imply both p (because it is factive) and Kp (because in the presence of the K operator, the possibility operator is redundant), and the existence of both of these implications would render it unproblematic that there is no logical distinction between knowable and known truth, in the sense of knowability so delineated. When we move to weaker notions of possibility on which the presence of the possibility operator is not redundant, the loss of this logical distinction is more troubling, and so if we

[3] Neil Tennant says, "For, very importantly, one has to secure the 'factive' nature of the complex $\Diamond K$: one can always infer φ from $\Diamond K\phi$." "Victor Vanquished", *Analysis* 62.2 (2002), p. 140.

[4] See, for example, Berit Brogaard and Joseph Salerno, "Clues to the Paradoxes of Knowability: Reply to Dummett and Tennant" *Analysis* 62 (2002), pp. 143–150.

[5] The details of the modal logic for this notion of possibility is explored in Timothy Williamson, "On Intuitionistic Modal Epistemic Logic", *Journal of Philosophical Logic* 21, (1992), pp. 63–89. We need not go into the precise details of such a logic, since the paradox does not depend on such modal details.

treat the possibility operator in terms of the weakest notion of possibility (either logical or metaphysical possibility), the derivations become the most paradoxical. So if there is a strong notion of possibility on which $\Diamond Kp$ is factive, the paradox may still be troubling (for there is no guarantee that the possibility operator will need to be treated as redundant in the presence of the knowledge operator); but the paradox is certainly troubling when we read the possibility operator in its weaker sense of mere logical or metaphysical possibility. Hence, regardless of what an intuitionist needs to say about possibility to provide room for an anti-realist theory of truth that identifies truth with knowability, there is a weaker notion of possibility that such a project provides no reason to reject and which can be used in a perfectly legitimate way in the derivations involved in the knowability paradox.

Second, to whatever extent a notion of possibility is encoded in intuitionistic double negation, it can't be identified with the weak notion of logical possibility that is adequate for Fitch's proof. The reason is simple: $\sim\sim p$ contradicts $\sim\sim\sim p$, both intuitionistically and classically, but $\Diamond p$ does not contradict $\Diamond\sim p$, on any understanding of negation, when the notion of possibility in question is a weak notion of logical or metaphysical possibility. So even if negation encodes some notion of possibility, it won't be the notion at work in the most troubling interpretations of Fitch's proof.

There is still the issue of the dual rules for the operators, however, and here two points are in order. First, it is far from clear that the formal semantical treatment of the operators in terms of quantification over worlds should be taken seriously enough that it provides intuitionists a reason to reject some of the dual rules. After all, as Quine has taught us, quantification has serious ontological import, and the mere fact that our formal semantics adverts to such devices should not in itself be taken as a reason to draw ontological conclusions directly from the formal semantics. There is, for example, the response that the formal semantics is merely a heuristic device rather than a metaphysically serious one.[6]

This point becomes more telling when we notice analogous cases where the dual rules are unproblematic and there is no temptation to treat the operators in question as implicitly quantifying over worlds (even if the semantics for the operators adverts to such worlds). Consider, for example, the moral inference from a denial of obligation to

[6] For more on the issues here, see Alvin Plantinga, *The Nature of Necessity* (Oxford, 1974), especially his discussion of the difference between pure and depraved semantics, as well as David Lewis, *On The Plurality of Worlds* (Oxford, 1986).

permissibility to refrain: if I have no obligation to give a student a higher grade just for asking, then it is permissible for me not to do so. For another example, not that if it is not beyond my control to try to honor my promises, then it is within my control not to do so.[7] In both of these cases, intuitionists have no special reason to balk at the inferences in question, and this fact reveals that if they balk in the case of necessity and possibility, the source of their objection is the connection between such operators and quantifiers in first-order theory. It is a mistake, however, to treat the connection between formal semantics and ontology in a way that will compel this worry, and there are a number of approaches to the language of possible worlds that simply refuse to treat them as objects of quantification.

Moreover, even if we do read off our ontology from our formal semantics, it is hard to adopt constructivist attitudes toward possible worlds. For a constructivist, no existential claim is allowed without a construction of, or an effective method of finding, an instance, but it is very hard to see what that would be like when we are talking about possible worlds. The closest I can come to finding such is through a consistency proof for a set of propositions, but even here problems remain. For one thing, a proposition can be impossible without there being proof rules that allow the derivation of an inconsistency from it—think of claims such as *this object is red and green all over*, or *some bachelors are married*. It takes auxiliary hypotheses to derive an inconsistency here, and these hypotheses will not be provably true unless they are all taken as axiomatic.

A final point provides the nail in the coffin of the attempt to evade the proof by denying some of the dual rules. The only dual rule needed in the proof from Chapter 1 is the rule that take us from $\Box\sim$ to $\sim\Diamond$, and the corresponding dual rule for quantifiers (from $\forall\sim$ to $\sim\exists$) is acceptable to constructivists. So even if we had to give up some of the dual rules, such as the one allowing us to go from $\sim\Box$ to $\Diamond\sim$, the only dual rule needed is not one that intuitionists and other constructivists have any reason to find objectionable.

So if there is something intuitionistically unacceptable about the underlying logic, the place to find it is going to be in the first-order

[7] This example is a bit more controversial than the first, since it appeals to a version of the Principle of Alternative Possibilities, which Harry Frankfurt attacked in "Alternative Possibilities and Moral Responsibility", *Journal of Philosophy* 66 (1969), pp. 829–839. The principle cited in the text does not succumb to Frankfurt's objections, however, and suggests a direction for finding a correct version of the principle.

theory. One of the primary motivations for intuitionism in the philosophy of mathematics and also for intuitionistic logic has to do with the issue of decidability and the existence of undecidable sentences, and a good place to enter into the fray of the dispute between classicists and intuitionists is by investigating this ground for rejecting classical logic.

DECIDABILITY ISSUES

In order to motivate intuitionistic logic, the intuitionist must give us some reason to refuse endorsing some of the classical rules. One maneuver here involves remaining neutral on the Law of Excluded Middle expressed as the schema (p v ∼p) on the basis of a skepticism about the decidability of all propositions. A mathematical statement is decidable, we may say, just when either it or its negation is provable. Generalizing to all statements, we may say that a statement is decidable just when the application of an effective decision method will establish its truth-value.

Once we see the derivation of Fitch's results, a problem emerges for defenses of intuitionism on the basis of skepticism about decidability. Intuitively, such skepticism is an appeal to the appropriateness of some type of epistemic humility concerning the reach of our intellectual powers. The derivation of Fitch's result, however, shows tension between such expressions of epistemic humility and any viewpoint that requires knowability for truth.

Note that the above account of decidability, according to which a statement is decidable just when the application of an effective decision method will establish its truth-value, makes the class of decidable statements at least a subclass of the class of knowable statements, for any effective method that will establish the truth-value of a claim will also yield knowledge. We should also note the danger of allowing the opposite inference, that anything knowable is also decidable, for then the tension between skepticism about decidability and a commitment to universal knowability devolves into outright contradiction. In such a case, no appeal to the knowability paradox would be needed to undermine this particular version of anti-realism, since it would be blatantly inconsistent at its core.

The important question, then, is whether we can find an argument for maintaining that the class of decidable claims is a proper subclass of the class of knowable claims. Neil Tennant argues for this view, claiming

that logical truth is, in a technical sense, undecidable but knowable (since axiomatizable).[8]

The inference is hasty, however. From the fact that a set of consistent sentences is axiomatizable it does not follow that they are knowable. Most important is the fact that the knowability claim will be true only if the axioms themselves are knowable, so merely pointing out axiomatizability facts doesn't settle the question of knowability. On the formal side, we can move from axiomatizability to recursive enumerability of the theorems, from which it follows that there is an effective method that will find a proof for any theorem in a finite number of steps. But knowability is lamentably psychological when viewed from the pristine realm of logic, and the possibility of knowledge does not follow from these logical points. The mere existence of an effective method for finding a proof does not show that it is possible for anyone to find that method or to employ it.

Once we leave the realm of logical truth, the problems become even more severe for trying to distinguish decidability from knowability. For one thing, the technical notion of decidability is not quite appropriate here. For ease of reference, we can use Tennant's characterization of the notion, which is:

> A domain of discourse D is decidable iff there exists an effective method F, having a value for every sentence of D, that, for every sentence of D, has the value T only if the sentence is true and has the value F only if the sentence is not true.[9]

An effective method is, according to Tennant, a mechanical procedure,[10] i.e., a procedure for computing an answer, which if followed correctly, is guaranteed to give a correct answer in a finite number of steps.[11]

We cannot use this notion for two reasons. First, we must distinguish between the formal notion just defined (even if imprecisely) and the use to which it might be put, for there may be such a function F without anyone being capable of seeing its effectiveness for the domain of discourse in general or being able to use it. To function as an epistemic notion, the formal notion must be supplemented with qualifications

[8] Neil Tennant, *The Taming of the True* (Oxford, 1997), p. 163. For the undecidability of first-order theory, see Alonzo Church, "A Note on the Entcheidunsproblem", *Journal of Symbolic Logic* 1 (1936), pp. 40–41.

[9] Tennant, *The Taming of the True*, p. 174.

[10] Ibid., p. 173.

[11] Geoffrey Hunter, *Metalogic An Introduction to the Metatheory of Standard First Order Logic* (Berkeley, Ca., 1973), p. 14.

about the knowability of, or ability to competently deploy, the effectiveness of the function to generate knowledge. So, just as before, there is, from the point of view of logic, a lamentable psychological aspect that cannot be ignored.

Before considering the other reason against using Tennant's characterization, it is important to forestall an objection. The concerns in the last paragraph should not be thought of in terms of the following argument: even though we might have thought that applying an effective method would yield knowledge, that view is mistaken since we might not know the results of applying function F because we don't know that applying F is effective. Such an argument is guilty of a levels confusion, claiming that Kp is false because KKp is false.[12] What I am arguing is that applying an effective method need not yield knowledge. A belief can be guaranteed to be true and yet not be known to be true (e.g., believing that $2 + 2 = 4$), and a method might be completely verific in spite of evidence to the contrary (clairvoyance might be impeccably accurate in spite of our skepticism about it). Levels confusion occurs when we infer that knowledge is absent on grounds that we are not sure the conditions for knowledge have been met. In such a case, the right conclusion to draw is that we do not know that we know, not that we do not have knowledge. In the present case, the claim is that mere reliability, even the most impeccable reliability, is not sufficient for knowledge, as the clairvoyance claim clearly shows.[13]

So the first point to note is that we must distinguish between purely formal concepts of decidability and epistemic ones. To use Tennant's account of decidability would be to ignore the significance of this issue in a way that is unacceptable.

The most important point, however, is that we must distinguish between purely logical concepts of decidability and empirical ones. Note in the above definition the requirement that the method in question provides a logical guarantee of a correct answer. Such a requirement implies that hardly any empirical claims are decidable, for there simply are no empirical methods that guarantee correct answers.

In order to provide a suitable extension to the empirical domain, then, the notion of a logical guarantee (of determining the truth-value of the statement in question) will have to be replaced with a weaker notion,

[12] See William P. Alston, "Levels Confusion in Epistemology", *Epistemic Justification*, (Ithaca, 1989), pp. 153–171.

[13] The clairvoyance example is first found in Laurence BonJour, *The Structure of Empirical Knowledge* (Cambridge, Mass., 1985).

such as confirmation, since nothing as strong as the notion of a proof is applicable in this realm. Instead of a guarantee of truth, the procedure cannot be required to guarantee anything more than confirmation of the truth-value of the sentence. To be effective, the procedure must get us to the truth, of course, but that will be an additional feature rather than something guaranteed by an application of the method. These thoughts can be accommodated by leaving Tennant's definition of decidability as is and replacing his account of the conditions under which a method is effective. The new understanding will be this: an effective method is a procedure for computing an answer, which produces a correct answer and, if followed correctly, is guaranteed to produce (at least *prima facie*) confirmation for that answer in a finite number of steps.

One might wonder why the method needs to be guaranteed to produce confirmation. I think the answer has to do with the intuitive notion of decidability and its connection to the notion of a decision procedure. Without such a guarantee, we could not, for example, construct a priori a list of preferred methods to use in our search for truth. We would be at the mercy of our environment, where the environment determined whether or not a method would yield confirmation. The intuitive idea of decidability, however, places the locus of control within the individual rather than within the environment—to call a sentence "decidable" is to report that we have within our resources a way of approaching that sentence so as to ascertain its truth-value. Flipping a coin that was, to all appearances, a fair coin would not constitute such a method, even if our environment were controlled by a meticulous and benevolent angel who ensured perfect calibration between states of the world and results of the coin flip and who also gerrymandered experiences so as to yield confirmation for what we conclude from the flip. No, the concept of decidability requires that the adequacy of resources and methods used not be at environmental behest, and if we are forced to abandon the idea that such resources are available, it is hard to see how an intuitionist has any hope for extending the idea of decidability to the empirical domain and using it to argue for a revision of the logic underlying the knowability paradox. The methods will not, of course, guarantee truth, but if they do not guarantee at least *prima facie* confirmation, then our methods are epistemically insufficient for classifying sentences into two groups, the decidable ones and the undecidable ones. Weaker methods that provide confirmation without a guarantee might allow us to categorize a sentence as decidable for the type of environment for which that method was adequate, but

would not undergird the intuitive idea underlying decidability of having within our own resources, independent of the environment in which we find ourselves, ways of appropriately investigating whether the sentences in question are true.

Those familiar with the history of epistemology of the last century will immediately call to mind a difficulty here. We have defined the broader notion of decidability, which applies to the empirical realm as well as to the more formal realm of logic and mathematics, in terms of a method that guarantees confirmation (thereby implying that there are epistemic principles that are necessarily true). The history of epistemology of the last century is highly suspicious of the claim that there are any such methods. The dominant epistemological viewpoint on the nature of confirmation or justification says that there are no such methods because there are no necessarily true epistemic principles. Without a suitable extension of the formal concept of decidability to the empirical domain, no case can be made from the general character of undecidables to the conclusion that classical logic needs to be revised, so the challenge to our characterization of the broader notion of decidability must be faced before proceeding further. I will argue that the objections from this dominant viewpoint can be met.

To be clear about the order of inquiry here, I will be arguing later that there are other more telling objections to be raised to Tennant's approach here, so in one sense, the present discussion is a detour from the main point to be argued. It is important, however, not only to reach the proper conclusion, but also to reach it in the right way, and since the dominant epistemological viewpoint of the last century—holistic coherentism—provides such an obvious route to denying Tennant's strategy, it is appropriate to disarm this objection before proceeding further. In the process, we will learn important lessons about what can be said for and against such a holistic viewpoint.

A CHALLENGE FROM (AND TO) HOLISTIC COHERENTISM?

In the language of contemporary epistemology, the assumption made in our account of decidability is that there are some epistemic principles that are necessarily true. Theories that deny the assumption that there are epistemic principles that are necessarily true are holistic theories of justification, such as holistic coherence theories, and such theories

have enjoyed considerable popularity over the last half-century or so. As William Lycan says,

> When I was a graduate student in the 1960s, the pantheon of American philosophy enshrined just four *theoi*: in alphabetical order, Roderick Chisholm, Nelson Goodman, W.V. Quine, and Wilfrid Sellars. Of these, the last three (allowing a bit of interpretation in Goodman's case) defended coherentist epistemologies that have had enormous influence.[14]

In describing his own relationship with Chisholm—the paragon of a defender of such necessary epistemic principles—Keith Lehrer identifies his attraction to coherentism with an intuitive sense that Chisholm's principles are, for the most part, true, but only contingently so.[15] Thus, there is a deep concern that the arguments just presented are inadequate and perhaps irrelevant because they assume the falsity of holistic coherentism.

I will argue, however, that such theories do not present any substantive challenge to the interpretation of decidability employed above and the direction our discussion is taking. There is an objection to holistic coherence theories concerning the basing relation that I will explain. This objection is thought by many to undermine such holistic theories, but they are mistaken. The way out of the objection, however, pulls the teeth from holistic coherentism as a threat to our assumption that there are necessarily truth epistemic principles.

The objection I have in mind arises from critics of coherentism who have depicted it so that it founders on the distinction between warrant for the *content* of a belief and warrant for the *believing* itself. According to this objection, one might have warrant for the content of what one believes without basing one's belief properly, without holding the belief because of what warrants it. When the first kind of warrant obtains, I will say that a belief is propositionally warranted; when a belief is both propositionally warranted and properly based, I will say it is doxastically warranted.[16]

The critics I have in mind are Alvin Plantinga and John Pollock. Plantinga characterizes coherentism so that it must deny that there are

[14] William G. Lycan, "Plantinga and Coherentisms", in *Warrant in Contemporary Epistemology*, Jonathan L. Kvanvig, ed. (Totowa, N.J., 1996), p. 3.

[15] Keith Lehrer made these remarks at the session of the Pacific Division of the American Philosophical Association in which Peter Klein and Ernest Sosa gave talks in celebration of Lehrer's work in 1996.

[16] Here I ignore differences between the epistemological concepts of rationality, justification, and warrant. There may be differences between these concepts that are important in some contexts, but the differences are irrelevant here.

non-basic warranted beliefs,[17] and John Pollock thinks that an important kind of coherentism cannot explain the difference between a properly and an improperly based belief.[18] I will show a connection between these two claims, but more important, I will argue that holistic coherentism can avoid these objections. In the context of our present discussion of decidability and knowability, the result of answering this objection will be that holistic coherentism ceases to pose a threat to the course of our discussion concerning the motivation for a revision of classical logic on the basis of the possibility of undecidable sentences.

Plantinga's Characterization

Plantinga characterizes coherentism as follows:

> Current lore has it . . . that the coherentist does not object to circular reasoning at all, provided the circle is large enough . . .
>
> But why saddle him with anything so miserably implausible? There is a much more charitable way to construe his characteristic claim. He should not be seen as endorsing circular reasoning . . . His suggestion, instead, is that coherence is the sole source of warrant. *He is instead pointing to a condition under which a belief is properly basic* . . . On his view, a belief B is properly basic for a person *S* if and only if *B* appropriately coheres with the rest of *S*'s noetic structure . . .[19]

Plantinga here proposes that coherentism is a special kind of foundationalism because it implies that there are properly basic beliefs. Even worse, coherentists are foundationalist zealots—not only are some warranted beliefs properly basic, all of them are.

These claims are shocking. Coherentists are wont to assert, not that *all* warranted beliefs are properly basic, but rather that *none* of them are. Coherentists would react strongly to being characterized as special kinds of foundationalists, inclined perhaps to treat such pronouncements as projections by harried foundationalists instead of insight into coherentism. We might suspect, then, that Plantinga has unusual senses of the crucial terms employed in his characterization of coherentism. The evidence, however, will show otherwise.

[17] Alvin Plantinga, *Warrant: The Current Debate* (Oxford, 1993). Page references in the text to Plantinga's views are to this work.

[18] John Pollock, *Contemporary Theories of Knowledge* (Totowa, N.J., 1986). Page references in the text to Pollock's view are to this work.

[19] Plantinga, *Warrant: The Current Debate*, pp. 77ff.

For Plantinga, which basic beliefs are properly basic is a matter for each particular theory to clarify,[20] and Plantinga endorses two principles about the basing relation itself:

(1) If one belief is based on another, then the second is a cause of the first; and
(2) If one explicitly infers a belief from another, then one bases the first belief on the second. [21]

One obvious implication of claim (2) is that no basic belief is inferential in character, for by definition basic beliefs are beliefs that are not based on other beliefs. Since inferential beliefs are not basic beliefs, Plantinga implies that coherentists must deny that any warranted belief is ever explicitly inferred from other beliefs. That is, coherentists deny the possibility of warranted inferential beliefs.

Recall that Plantinga's explicit motivation for his construal of coherentism was charity, wanting to avoid saddling the coherentist with the "miserably implausible" claim that circular reasoning can sometimes generate warrant. Such charity will leave the coherentist wishing Plantinga had already given at the office. No coherentist I know of accepts such a characterization, and if some do, they are foolish indeed. For it is patently obvious that warranted beliefs can be inferred ones.

So, did Plantinga err in divining the character of coherentism, and if so, where? If we look carefully at the quotation above, we find two adjacent sentences that give a clue. He says, correctly, that the coherentist's "suggestion . . . is that coherence is the sole source of warrant". But in the very next sentence, he also claims that the coherentist "is . . . pointing to a condition under which a belief is properly basic". Since his discussion of coherentism focuses on this second claim, Plantinga may think that the two claims are equivalent, or that the second follows from the first.

No such connection exists. The first remark (that coherence is the sole source of warrant) is a claim about propositional warrant, that kind of warrant that accrues to the content of what one believes (or to the content of a claim one does not believe). According to a coherentist, only coherence with an appropriate system is capable of generating such warrant. The concepts of basing and proper basing employed in the second claim above (that the coherentist is pointing to a condition under which a belief is properly basic) are quite another thing. These concepts

[20] Ibid., p. 70. [21] Ibid., p. 70.

have to do with doxastic warrant, the warrant a belief has when its content is propositionally warranted and the belief is properly based. So, Plantinga has made a mistake in characterizing coherentism, one that arises from failing to recognize the distinction between propositional and doxastic warrant.

Somewhat surprising, it is not clear that Plantinga believes what he writes anyway. He remarks later on two types of coherentism:

> The pure coherentist holds that all warranted propositions in a noetic structure are basic in that structure; no warrant gets transmitted. The impure coherentist holds that some propositions may get their warrant by virtue of being believed on the basis of others; but the ultimate source of the warrant in question is coherence. Both accept the view that coherence is the *only* source of warrant; and this is the central coherentist claim.[22]

This quote stands in contrast to the earlier quote from Plantinga, for here we have specified a version of coherentism (impure coherentism) according to which some beliefs are warranted but not properly basic, because they get their warrant from other beliefs. What makes the view coherentist is its insistence that the ultimate source of warrant is still coherence. Elsewhere Plantinga describes the view as follows: "Global coherentism is compatible with local foundationalism; the view that coherence alone is the source of warrant is compatible with the view that warrant is sometimes transmitted."[23]

So Plantinga thinks of warrant transmission as a distinctively foundationalist idea, though it can be abducted for coherentist purposes as long as it is kept under tight wraps. Perhaps a charitable way to interpret Plantinga is to take his earlier remarks about coherentism as applying only to pure coherentism. Then we could say that he characterizes coherentism so that coherence is the sole source of warrant, and that this claim, when unsullied by impious idolatry at the foundationalists' altar, yields the pure view that all warranted beliefs are properly basic.

This interpretation removes the contradiction but leaves the mystery, for we still have a characterization of one kind of coherentism that no coherentist should accept and that is still susceptible to the response that such a characterization arises only by ignoring the distinction between propositional and doxastic warrant. The mystery could be solved in a way that absolves Plantinga if coherentist purity undermines any appeal to this distinction in such a way that the only choice left to

[22] *Warrant: The Current Debate*, pp. 79–80. [23] Ibid., p. 79.

coherentists is to agree that all warranted beliefs are properly basic, despite what they might wish to say. Plantinga gives no such argument, but John Pollock does.

Pollock's Criticism

Pollock says that there is a distinction between what I call propositional and doxastic warrant, and that any correct theory must allow for it. But, he claims, (holistic positive) coherence theories (the most plausible version of the view, on which justification is holistic rather than linear and requires positive reasons rather than merely the absence of reasons against) cannot explain the distinction, for such a distinction would require a coherentist to maintain that every belief is based on every other belief because of the holistic picture of warrant adopted by coherentists.[24] So holistic coherentism should be rejected.

This criticism can support Plantinga's characterization of coherentism in the following way. If Pollock is right, some coherentists must hold that no belief ever fails to be warranted by failing to be properly based. To claim otherwise requires holding the hopelessly implausible claim that everything in one's belief system is causally responsible for belief in order for the belief to be doxastically warranted. Yet, if the coherentist must maintain that no belief ever falls into epistemic disfavor by being improperly based, the coherentist must admit that every belief is automatically based properly, that is, that it passes any legitimate basing test for (doxastic) warrant. So we get the Plantingian characterization if Pollock is right: for a true coherentist, every warranted belief is properly basic.

The important question is whether Pollock's criticism of (one kind of) coherentism can be sustained. In a word, it cannot. Pollock assumes that the coherentist must clarify the concept of proper basing in terms of a causal relation between that which propositionally warrants a belief and the holding of the belief. This assumption involves two requirements. The first is that the basing relation is a species of causal relation; I will grant that point in what follows. The second concerns what the basing relation is a relation between. According to Pollock, it must be a relation between *that which propositionally warrants a belief* and *the belief itself*. I will focus on this second claim.

Coherentists need not accept this claim. Propositional warrant may be as systemic an affair as you please, and yet doxastic warrant depend

[24] *Contemporary Theories of Knowledge*, p. 81.

nonetheless on some special components of the system. The alternative view, underlying Pollock's criticism, myopically ignores the vast array of logical relationships that can obtain between (the contents of) beliefs, yielding that some beliefs are relevant to the epistemic status of other beliefs even though the former beliefs do not impart any warrant to them. One such logical relation borrows from the work of J. L. Mackie on causation.[25] According to Mackie, a cause of an effect is an INUS condition: it is an Insufficient but Non-redundant element of larger condition which is itself Unnecessary but Sufficient for the occurrence of the effect. For example my throwing the ball causes the window to break, even though the first even is hardly sufficient in any strong sense for the breaking of the window (an equal and opposite force on the other side of the window at the time the ball strikes the window would have prevented the breakage). Nonetheless, according to Mackie, there is a larger condition including the causal field in which my throwing of the ball occurs, and that larger condition is sufficient, though not necessary, for the breaking of the window. Furthermore, my throwing of the ball is a non-redundant component of that larger condition: take that condition without my throwing the ball, and the window will not break.

The concept of an INUS condition can be put to use in epistemology as well. In particular, a coherentist can claim that some beliefs are INUS conditions for the warrant of other beliefs. An INUS condition for warrant is insufficient for the imparting of warrant, and hence is not a warrant-imparter, or defeasible reason, for belief. It is, however, a non-redundant part of a larger condition which is itself a reason for belief (a warrant-imparter). Such an INUS condition for warrant might be a cause of belief, and there is no reason a coherentist cannot appeal to such conditions in an account of proper basing. Such an appeal amounts to a denial of the claim that the basing relation is a relation between that which propositionally warrants a belief and the belief itself; according to the holistic coherentist, that relation obtains only between the entire relevant system and a given belief. Nonetheless, such INUS conditions for warrant are clearly relevant to the epistemic status of belief—take away all the INUS conditions for warrant, and in normal circumstances warrant itself will disappear (just as taking away all the causes of an event will normally result in the absence of the event as well). So the coherentist can deny that the basing relation must be a relation between items that stand in the warrant-imparting relation by

[25] J. L. Mackie, *The Cement of the Universe: A Study of Causation* (Oxford, 1974).

developing a concept of epistemic relevance that includes INUS conditions for warrant and allowing that proper basing can occur when one bases a belief on something epistemically relevant to it.

Some sketchy examples of what such a theory might look like may prove helpful. Consider subjective versions of coherentism. Such versions can maintain that one's system of beliefs contains a (subjective) theory of epistemic relevance that places constraints on appropriate basing. On such a theory, having an appropriately based belief, (i.e., a belief that is not disbarred from candidacy for doxastic warrant because of some defect with regard to how it is based), might require being aware of just which elements in the system count as epistemically relevant to the belief. Alternatively, a theory of this sort might require that the explanation (the best one and perhaps the one the person in question accepts or would accept on reflection) of why one holds the belief conform to that theory of relevance.

Imagine, for another example, a Bayesian account of warrant in terms of degrees of belief. The version I imagine employs diachronic dutch books as a constraint on warranted degrees of belief, so that one's warranted degree of belief tomorrow is a function of one's conditional degrees of belief today, conditional on what future experience might teach.[26] Such a theory is fully holistic because warrant obtains on such a theory only when the entire set of beliefs is probabilistically coherent. What is important about the view in the present context is that one's conditional probabilities today contain an implicit theory of epistemic relevance, a theory that implies that some new information is relevant to some degrees of belief and not others. By containing such an implicit theory of relevance, a Bayesian account of warrant can add a basing requirement without adopting the absurd viewpoint that the entire system of (degrees of) belief must be causally responsible for every properly based degree of belief.

Pollock might wish to classify these theories as non-holistic, but his account of that distinction fails to yield that result. He says that a linear coherence theory "embraces essentially the same view of reasons and reasoning as a foundations theory," one according to which "P is a reason for S to believe Q by virtue of some relation holding specifically between P and Q. A reason for a belief is not automatically the set of *all* one's beliefs." He says a holistic coherence theory claims that, "in

[26] For explication and discussion of such versions of Bayesianism, see Bas van Fraassen, *Laws and Symmetry* (Oxford, 1989), especially Part II.

order for S to have reason for believing P, there must be a relationship between P and the set of *all* of his beliefs (where this relationship cannot be decomposed into simple reason relationships between individual beliefs)."[27]

This distinction fails to show that the above theories are non-holistic, for an INUS condition for warrant for *p* is simply not a reason, defeasible or otherwise, for *p*. A reason for *p* is something that imparts (perhaps defeasible) warrant to *p*, and an INUS condition for warrant is Insufficient for the impartation of any warrant whatsoever. Thus, a coherentist can affirm a holistic view of reasons, where only systemic relations count in favor of the warrantedness or positive epistemic status of belief, and can nonetheless isolate INUS conditions of positive epistemic status with reference to which one defines proper basing. Consider the Bayesian view above again. One's probability for *p given q* constrains one's opinion about *p* upon learning *q* in part because the condition cited is not simply *q*, but rather *q* plus all of one's background information. So what is sufficient for the warrant *p* acquires when *q* is learned is some larger condition (*q* plus background information) of which *q* is a non-redundant component (because the background information alone does not warrant *p*). Further, the larger condition, though sufficient for the warrant of *p*, is not itself necessary. So the Bayesian view is already three-fourths of the way toward *q* being an INUS condition for warrant with respect to *p*; all that needs to be affirmed is that *q* on its own never imparts any degree of warrant. Certainly, that option is open to Bayesians, and if it is taken, the conditional probabilities of today that constrain future opinion do so by specifying INUS conditions for warrant. By taking this option, the Bayesian can avoid the charge of holding a linear coherence theory and still use INUS conditions for warrant in order to define proper basing.

One might worry that this INUS condition response will not work on grounds that everything in the relevant system will be such an INUS condition. Such a worry can be allayed easily. I'd have the same warrant for thinking that I exist even if I didn't believe my grandmother is quirky, or even if I believed the opposite. So the latter belief plays no important explanatory role regarding the warrant of the former belief, and hence is not a Non-redundant part of the belief system. So not every element of a belief system is an INUS condition of warrant for every belief.

[27] *Contemporary Theories of Knowledge*, p. 73.

One might worry that *too much* of a belief system will be INUS conditions, and hence that the coherentist will have to hold that a belief will have to be based on a substantial part of the system. This concern highlights the need for something I won't be able to provide here: a good theory of the basing relation. The force of the criticism can nevertheless be blocked, for the difficulty is not unique to coherentism. According to Pollock's own theory and other versions of foundationalism,[28] *p* can be a reason for *q*, be subject to defeat by *d*, where this defeater is itself overridden by a further claim *o*. In such a case, there are at least two reasons for believing *q*: *p* and *p*&*d*&*o*. Furthermore, this heirarchy of defeaters and overriders is potentially unlimited, yielding the result that there can be a potentially unlimited number of reasons for believing any particular claim. The question for such theories is, which reasons should the belief be based on? This question raises precisely the worry faced by the coherentist that too much of the system will be INUS conditions for warrant. Furthermore, both kinds of theorists will answer the worry in one of two ways: either isolate some of the INUS conditions or defeasible reasons as privileged when it comes to basing, or argue that there is nothing especially implausible about insisting that one base one's belief on the entire collection of INUS conditions or defeasible reasons (for, as all should recognize, belief formation is a very complex thing and it should not surprise us if very much of our belief system is causally responsible in one way or another for belief). We can thus legitimately ignore the worry that too much of the system will be INUS conditions for warrant, for the difficulty posed is not unique to coherentism and the coherentist can approach the problem in much the same way as other theorists.

So the point stands that in the subjective and Bayesian examples above, accounts of the basing relation are possible that are not psychologically implausible. The adequacy of any such theory is, of course, another question. But the mere possibility of formulating them shows that Pollock is mistaken.

[28] Pollock denies that his theory is foundationalist, on grounds that foundationalists and coherentists are committed to the Doxastic Assumption, according to which all beliefs are warranted by their relation to other beliefs. Even though there are subtle issues to be addressed regarding the relationship between coherentism and this assumption, it is patently obvious that foundationalists are not committed to it. Typically, foundationalists hold that some beliefs are justified directly by sensory experience, thus denying the assumption. Pollock says he is not a foundationalist because he denies the doxastic assumption. Once the connection between the two is broken, however, it is obvious that Pollock is just a foundationalist with an incorrect metatheory about it.

With the failure of Pollock's argument goes Plantinga's construal of coherentism, for the only hope for that construal rested on denying that coherentists can help themselves to the distinction between doxastic and propositional warrant. So Plantinga has misconstrued coherentism and Pollock has misapprehended its weaknesses. Coherentism is the view that limits the sources of propositional warrant to coherence itself, but this characterization leaves coherentists free to impose basing requirements of their choice on doxastic warrant. Contrary to Pollock, coherentists can distinguish between propositional and doxastic warrant, and because they can, contrary to Plantinga, coherentism does not imply that any warranted beliefs are properly basic.

Lessons Learned

The point of this discussion concerned the assumption in our account of the broader concept of decidability—the extension of the formal concept to the empirical domain—that there are necessarily true epistemic principles, and this assumption seemed to presuppose that holistic coherentism is false. Some versions of holistic coherentism are false, because they succumb to the above objection concerning the proper treatment of the basing relation. Others, however, have resources to address such objections. When they do, however, they will be forced to construct epistemic principles of their own. They will not be the same type of epistemic principles foundationalists offer, for they will specify, perhaps, only INUS conditions for warrant. They will, however, have the same modal status as foundationalist principles, and it is the modal element in the account of broad decidability that is at stake in this discussion. When defining broad decidability, I wrote of methods guaranteed to produce *prima facie* confirmation. If holistic coherentist language is preferred, we can substitute talk of methods that guarantee production of INUS conditions for warrant. In either case, the intuitive idea is maintained of having resources within our control which, independent of our environment, allow us to investigate in an adequate way whether a sentence is true. In what follows, I will continue to speak of methods that guarantee production of *prima facie* confirmation for the answer generated to the question of the truth or falsity of a given sentence, with the understanding that if holistic coherentism is the preferred theory, some minor translation will be needed in our discussion. Minor translation, however, presents no logical obstacle to the attempt to use the concept of broad decidability to argue for a revision of classical logic.

In the attempt to provide such an argument, we have already seen the demise of that project when decidability is identified with knowability. One way to respond is to suggest, with Tennant, that the class of decidables is a subclass of the class of knowables.

I want to argue, however, that this approach will not work, and we can begin to move toward this conclusion by seeing our way past Tennant's argument. Tennant argues, it will be recalled, that the class of decidables is smaller than the class of knowables, yielding an implication from decidability to knowability, but not the reverse. We must remember, however, that what is undecidable by Church's theorem is a *set*, not a *sentence*. The set of theorems of first-order theory is undecidable, but it is not *sets* that are knowable. What is knowable or not is a *sentence* (or some other truth-bearer). So Tennant's argument gives us no grounds to deny an inference from the premise that a certain sentence is knowable to the conclusion that this sentence is decidable.

If we limit our discussion to sentences, note what we should say. If a claim, logical or empirical, is known to be true, then it would be a mistake to say that the truth of the sentence remains undecided. The truth-value of the sentence has been decided, in the sense that a method has been used that both yields the answer "true" for the sentence (and hence yields the answer "true" if the sentence is true and "false" if the sentence is false) and guarantees confirmation for the answer generated. To hold otherwise would require adopting a construal of decidedness on which the fact that a sentence remained undecided would not be a reflection of our epistemic limitations at all, for it is not an epistemic limitation for there to be a sentence that is undecided in some technical sense and yet known to be true. So knowledge implies decidedness.

Suppose, then, that a particular sentence is not known to be true, but that it could be known to be true. This possibility of knowledge implies the possibility of decidedness for the sentence in question. If not, then there would be a possible world in which the claim in question is known to be true, even though that claim is undecidable and hence not decided in any world. But if the claim in question is not decided in any world, then the fact that particular sentence is known to be true is irrelevant to the issue of whether the truth-value of the sentence has been decided, i.e., knowledge fails to imply decidedness. So knowability implies decidability because (i) knowledge implies decidedness, and (ii) the modal argument "if p implies q, then $\Diamond p$ implies $\Diamond q$" is impeccable.

These points remain if we attend carefully to the details of the account of broad decidability. Recall that, for that account, we replace the idea

of a guarantee of truth with the idea of a guarantee of confirmation. To be effective, the procedure must get us to the truth, and it must do so in a way that, if followed correctly, is guaranteed to produce confirmation for that answer in a finite number of steps.

Thus, a decidable sentence is one for which there is a method that can produce, at worst, an accidentally true belief for which there is confirmation. As is well known, knowledge is more than accidentally true belief for which one has confirmation—that is the lesson of the Gettier literature. Still, if a belief can be confirmed and true in an accidental fashion, that same belief can be confirmed in the same way and be true in a non-accidental fashion. So, since decidable sentences are ones for which there is a guarantee of confirmation, they are also sentences which can be known to be true through the application of the method which guarantees confirmation.

There is a further complication. It is easy to mistake the guarantee of confirmation in the above account of decidability for a guarantee that the belief is, all-things-considered, confirmed. This mistake would be disastrous for the account of effective methods, for there are no empirical methods that are guaranteed to produce an all-things-considered confirmation. This point is a caution against the idea of trying to regiment evidential relationships as rules of inference akin to those we find in logical theory: such a practice ignores the defeasible character of evidential connections, imposing monotonic conditions on a non-monotonic subject matter.[29] The counter to this tendency is to keep in mind that applying an effective decision procedure in the empirical realm may be guaranteed to yield some confirmation for the claim in question, but the guarantee provided is only for *prima facie* confirmation, confirmation which is an all-things-considered confirmation when every defeater present is addressed adequately by some reinstater.[30]

[29] Compare, on this point, John Pollock's Oscar program. That program involves regimenting confirmation relations as something akin to natural deduction rules, but the program is built in such a way that, when the rules are defeasible ones, Oscar is sent out to look for defeating information, rather than directly to form a belief. My objection in the text is not to formulating confirmation relations akin to natural deduction rules, but rather to formulating them without at the same time structuring the theory to honor the defeasible character of such reasoning. On Pollock's work and the Oscar project, see the website for the project at <http://oscarhome.soc-sci.arizona.edu/ftp/OSCAR-web-page/oscar.html>.

[30] A defeater of the confirmation provided for p by e is a further piece of information d such that the conjunction $d\&e$ does not provide confirmation for p. A reinstater is some still further piece of information r such that conjunction $d\&e\&r$ provides confirmation for p.

This qualification raises the need for a further inference, but it is not problematic. For when we have an accidentally true but confirmed claim, as discussed in the last paragraph, that is no guarantee that the confirmation is sufficient for knowledge. The confirmation might be present and yet defeated by further information. Even so, given the definition of *prima facie* confirmation, we can claim the following: if an effective method produces confirmation that is subject to defeat in a particular case, it is possible for that method to produce confirmation that is not subject to defeat or is accompanied by a reinstater. Hence if the confirmation present is not an all-things-considered confirmation, it might be, and if it were, it might be knowledge. So application of an effective method in the sense articulated implies knowability.

This argument does not consider knowledge that is acquired through the application of no method whatsoever. Perhaps coming to know the color of an apple by looking at it counts as such. Part of the issue here concerns what we are going to count as a method, but I think this issue need not detain us. If we define decidability in such a way that knowledge of p does not yield the implication that p has been decided, the concept of decidability can be of no interest whatsoever to questions in the philosophy of logic, no more than it would be if the inference from having a proof of p to p's having been decided were undermined. For without such a connection, undecidables could be known to be true, and perhaps even proved to be true, leaving it utterly mysterious why anyone would think that issues regarding undecidability could have any effect on what logical principles should be endorsed. The best way to proceed, then, is to embrace a generous enough characterization of the notion of a method that all knowledge counts as an application of some method or other, leaving the inference from knowability to decidability intact.

The argument against treating decidability as a more restrictive concept than knowability can be put in another, more general way. First, if truth and knowability are intimately linked, it is not clear why the failure of a stronger-than-knowability concept could motivate any hesitance about either (LEM) or the associated semantic principle of bivalence. Consider an analogous attempt. Epistemologists over the last 100 years or so have consistently rejected the idea that all sentences are infallibly knowable. Being infallibly knowable entails being decidable, but the denial of infallible knowability would have no consequence whatsoever for either (LEM) or bivalence. The reason it could have no such consequence is because it lacks the requisite connection to the theory of truth. What makes decidability appear to be relevant to (LEM)

and bivalence is that it has been thought to be truth-related, that it has been thought to inform regarding what a proper theory of truth should look like. So the anti-realist needs some understanding of decidability so that the lack of some general, effective method for deciding an issue bears some connection to truth, unlike the concept of infallible knowability. When decidability is taken to be logically coextensive with knowability and truth is taken to be logically tied to knowability, decidability may be taken to have the required connection to truth. But if all we know is that decidability is strictly stronger than knowability, how can concerns about decidability have any implications for (LEM) or bivalence? More to the point, how could it have such implications when the concept of infallible knowability doesn't? Clearly, the implications would not be capturable in any formalization at all, and so merely pointing out that decidability, properly conceived, implies knowability but is not logically equivalent to it will be of no help whatsoever to an anti-realist who wishes to motivate intuitionistic logic by appeal to decidability.

Given this conclusion, intuitionists will be hard-pressed to endorse the knowability of all truth and yet use reservations about decidability to motivate a revision of classical logic. These considerations suggest strongly that decidability cannot be used successfully by anti-realists to motivate intuitionism or any other logic, so long as anti-realism continues to be characterized by a commitment to the knowability principle.

Still, other arguments may be forthcoming, and even if all the arguments that might be given on behalf of intuitionism fail, that does not show that the position is false. To give the position a full hearing, then, we ought to investigate whether adopting intuitionistic logic can rescue anti-realism from the paradox by providing an acceptable solution to it.

CAN INTUITIONISM SAVE ANTI-REALISM?

Timothy Williamson, among others, has argued that the lesson of the paradox is that the knowability principle is false without endorsing a revision of classical logic.[31] For, according to the classicist, to deny that some truths are not known is to affirm silly verificationism, the view that all truths are known. Such an equivalence rests, however, on the

[31] Timothy Williamson, "On Intuitionistic Modal Epistemic Logic", *Journal of Philosophical Logic* 21 (1992), pp. 63–89.

interdefinability of the quantifiers, something intuitionists deny. Given an intuitionistic treatment of negation, discharging the assumption that some truths are not known yields only

$$(\mathrm{IC})\ \mathrm{p} \rightarrow \sim\sim \mathrm{Kp},$$

or

$$(\mathrm{IC}')\ \sim \mathrm{Kp} \rightarrow \sim \mathrm{p}.$$

Williamson hold that whereas Fitch's conclusion is obviously silly, neither of these conclusions has been shown to be obviously silly.[32] Thus, the intuitionist may avoid having to abandon an anti-realist conception of truth in the face of the paradox by affirming instead the idea that if a claim is true, it is false that it is not known.

Given the multiple negations in each of these sentences, they are harder to process semantically than is the claim that all truths are known, but once the processing has been completed it is hard to see how they are more acceptable than the omniscience claim. After carefully considering what each of the above claims say—that nothing true is unknown and that nothing unknown is true—they both sound obviously false, as obviously false as claiming that all truths are known. So how can the intuitionist claim to have made progress by replacing the omniscience claim with either of them?

The standard maneuver at this point is to appeal to what the connectives mean for the intuitionist that is different from what they mean for the classicist. For example, it is sometimes said that the negation symbol means for the intuitionist "there is a disproof of", so that $\sim p$ means, for the intuitionist, "there is a proof that p is absurd", and that 'truth' for intuitionists means the same thing as 'provability'.

Consider first Timothy Williamson's attempt to show that the knowability paradox does not refute anti-realism. He does so in two ways, one of which involves questioning whether it is as "evidently absurd" to claim that nothing unknown is true as it is to claim that all truths are known. Though Williamson makes no explicit appeal to the idea that negation means for the intuitionist something different from what it means for the classicist, it is tempting to use such an appeal to provide an argument for Williamson's claim here. Williamson's other defense focuses on the role of the concept of implication in the paradox, and here Williamson is much more forthcoming, explicitly endorsing

[32] Timothy Williamson, "Intuitionism Disproved?" *Analysis* (1982), pp. 203–207.

the semantic maneuver above in describing constructivist treatments of the conditional. He says,

> However, a subtly different form of constructivist semantics is available. For example, a semantic treatment of the conditional in intuitionistic mathematics is often stated roughly as follows: a proof of $P \rightarrow Q$ is a function (in the sense of an *operation*) that evidently takes any proof of P to a proof of Q.[33]

Williamson's idea is that, because we represent the claim that nothing unknown is true using '$\rightarrow$', the intuitionistic might appeal to a constructivist semantical treatment of this connective. By doing so, the intuitionist can defend the claim that, though it is evidently absurd to hold that all truths are known, it is not evidently absurd to hold that nothing unknown is true. Williamson draws precisely this conclusion:

> I have argued that since anti-realists want to assert that all truths are knowable while denying that all truths are known, they have a *reason* to adopt constructivist semantics . . . No argument along the lines of Fitch's has yet shown it [intuitionistic anti-realism] to be incapable of acknowledging the truism that not all truths are known.[34]

So Williamson's writings give two ways in which it might be thought that intuitionistic anti-realism can endorse the claim that nothing unknown is true while denying that all truths are known. Both ways involve intuitionistic readings of key logical connectives, either negation or implication.

Neil Tennant does something similar. He says,

> For the anti-realist, the role of logical inference is to preserve warranted assertibility . . . First, consider 'ϕ; *ergo* $\sim\sim K\phi$'. Suppose there is warrant to assert ϕ. Then it would be absurd to suppose that a contradiction could be derived from the assumption that ϕ were known. So anti-realism seems to be unruffled.
>
> Secondly, consider '$\sim K\phi$; *ergo*, $\sim\phi$'. This too is in order. Remember that the anti-realist uses a strong interpretaton of negation. $\sim K\phi$ means that the assumption that ϕ is known (by someone, at some time) leads to absurdity. Surely this could only be so because ϕ itself leads to absurdity?[35]

Tennant's approach is similar to Williamson's. In both cases, to see if the knowability paradox refutes intuitionistic anti-realism, we "get inside" the view to see what the claims mean to those who hold the view, and we

[33] Timothy Williamson, "Knowability and Constructivism", *Philosophical Quarterly* 38 (1988), p. 429.
[34] Ibid., p. 432.
[35] Neil Tennant, *The Taming of the True*, p. 262.

identify the semantic meaning of the ordinary expressions such as "follows from" and "it is not the case that" with the proposal put forward by those who hold the view in question. So, for Tennant, when we consider an argument with ϕ as a premise and $\sim\sim K\,\phi$ as a conclusion, we don't interpret the premise as simply asserting the truth of ϕ. Instead, we see what role logical inference is supposed to play in the anti-realist conception of things, and then we treat the premise as saying that there is warrant to assert ϕ. Moreover, we do the same with the conclusion of the argument, and note that there is nothing untoward about the argument form as rephrased.

I want to argue that this move, which I will term the "semantic revision" response to the paradox, cannot succeed. First, I will argue that it threatens to undermine the possibility of a substantive philosophy of logic, and second, I will argue that it misdiagnoses the paradoxicality that arises from Fitch's proof.

On the issue of the threat to the possibility of a substantive philosophy of logic, it is worth noting Quine's view of intuitionism. He says,

> The intuitionist should not be viewed as controverting us as to the true laws of certain fixed logical operations, namely, negation and alternation. He should be viewed rather as opposing our negation and alternation as unscientific ideas, and propounding certain other ideas, somewhat analogous, of his own.[36]

According to Quine, intuitionists and classicists do not disagree about negation. Instead, the intuitionist is proposing that we speak a different language, one in which 'not' has a different meaning than it does in our language.[37] For example, we might say that when an intuitionist writes '$\sim$p' he or she means to assert "there is a disproof of p".

There is a line of argument that can be used to arrive at this conclusion. The metalinguistic project of describing the logical behavior of the connectives in an object language must have recourse to its own logical connectives, and a major concern is that the logic of the metalinguistic connectives—either classical or intuitionistic—will force a certain treatment of the connectives in the object language. In *The Logical Basis of Metaphysics*, Michael Dummett laments the "pernicious" tendency to insist that the metalanguage should obey the

[36] W. V. O. Quine, *Philosophy of Logic* (Englewood Cliffs, N. J., 1970), p. 87.

[37] Quine, of course, could not put the point in this way, given his antipathy toward meanings. I leave it to the Quineans to formulate the point without appeal to such.

same logical laws as the object language.[38] In response, Neil Tennant claims:

> What the reformist [non-classicist] has to watch out for . . . is any argument advanced by the classicist that invoked only principles that the reformist himself would accept . . . If the reformist takes up the gauntlet to provide a philosophically stable account of his logic, then by the same token the classicist may only seek to persuade him of the soundness of the disputed classical principles by providing an argument based somehow only on principles which the reformist can accept. If the latter turns out to be impossible (as it would if the reformist were right!), then so much the worse for polite attempts to accommodate the classical disputant. There just will not be any neutral point from which to adjudicate the dispute. And why not simply accept that?[39]

There are two issues here. One is whether there is always and everywhere neutral ground from which to address philosophical disputes, and Tennant is correct that there is no guarantee of such. The second point is whether disputes of this sort are actually disputes at all, as Tennant maintains. If we take the logical theory presented to be constitutive and exhaustive of the meaning of the connectives, the result is not that we have a dispute for which no neutral ground of adjudication can be found. When we take the logical theory in this way, we get Quine's result, the result that there is no dispute at all (not just an irresolvable one). The classicist and the intuitionist are simply speaking different languages. Dummett's "perniciousness" concern regarding the requirement of sameness of logic for both object language and metalanguage then becomes severe, since the requirement commits one to Quine's viewpoint once one takes the logic in question to be constitutive and exhaustive of meaning. Then one's only hope for a substantive philosophy of logic will be to find metalanguage that is meaning-compatible with the alternative logics, and there may be no such possibility.

It is easy, and, I want to suggest, appropriate, to be sympathetic with Dummett on this point. A truly substantive philosophy of logic ought to be possible, and for it to be possible in any robust way, theorists who disagree about natural deduction rules for certain connectives and about the proper semantical treatment of those connectives ought, at least sometimes, to be disagreeing with each other. Here, as elsewhere in philosophy, there is always the possibility that a dispute is merely verbal, but we should want to avoid the conclusion that all such disputes are

[38] Michael Dummett, *The Logical Basis of Metaphysics* (Cambridge, Mass., 1991), pp. 54–55.

[39] Neil Tennant, *The Taming of the True*, p. 307.

merely verbal. For such a substantive philosophy of logic to be possible, it will have to be possible for disputants to reason about the correct logic in the neutral medium of ordinary language, assessing which formal systems best capture natural language reasoning, both in general and regarding specific domains of discourse.

Here Tennant is right, however, for there is no guarantee that the medium of ordinary language is neutral at all. But if it isn't neutral, then there is no debate at all, but only different Quinean proposals regarding which language would be best to speak.

The point to note here is that if we adopt the semantic revision approach, we end up Quineans on this score. On the semantic revision approach, what various symbols mean is to be interpreted solely in terms of the logical theory being proposed. When the same symbols appear in different proposals, they simply have different meanings, which is Quine's point. The problem with this approach is the one Dummett shows concern over: the possibility of a substantive philosophy of logic gets lost in the process.

So before endorsing the semantic revision approach to the paradox, defenders of this approach need first to explain how their points are compatible with a substantive philosophy of logic. To point out merely that negation, or implication, means something different to the intuitionist than to the classicist is not by itself a suitable reply to the apparent embarrassment to anti-realists by Fitch's proof.

The second point I wish to argue is that the semantic revision proposal misdiagnoses the paradox. Recall that there is nothing classical involved in inferring a contradiction from the dual assumptions that every truth is knowable and that some truths are not known. Intuitionists disagree only about how to discharge the second assumption, given an inviolable commitment to the first. Classicists discharge by concluding that every truth is known, intuitionists by concluding that nothing unknown is true. It is here that the semantic revision proposal must intervene, if it is to save anti-realism, and it does so by claiming something about the meaning of the discharge step. It claims, that is, that if we adopt classical interpretations of the logical components of the claim that nothing unknown is true, then the claim is obviously false; but if we adopt intuitionistic interpretations of these same components the claim is not obviously false. The final step of the semantic revision response is crucial: if the discharge step is not itself obviously false, Fitch's proof displays no special problem for anti-realism and the knowability paradox is dissolved.

It is here that the fully general characterization of the paradox I have argued for and which guides this work comes into play. The heart of the paradox of knowability is simply not a matter of the truth or falsity, obvious or not, of the claim that all truths are known or that nothing unknown is true. The fully general characterization of the paradox involves a logical equivalence that implies a lost logical distinction between actuality and possibility. The standard presentation I have used throughout is that the equivalence is between the claims that all truths are known and all truths are knowable. Should we adopt intuitionistic strictures on Fitch's proof, we will need to use different claims in the statement of equivalence, the claims that nothing unknown is true and nothing unknowable is true. Regardless of which claims are used in the statement of logical equivalence, however, we still have a lost logical distinction between actuality and possibility. In the former case it is a lost distinction between universally known truth and universally knowable truth; in the latter case, between universally unknown truth and universally unknowable truth.

The point of this reminder of the most general paradoxicality triggered by Fitch's proof is to make clear how the semantic revision move adopts a too local picture of the paradox. It views the paradox as dissolved should we find that the legitimate conclusion to draw from the contradiction Fitch elicits is something not nearly as unpalatable as we first thought. The point of a more general understanding of what is paradoxical here is that the question of the acceptability of the claim that nothing unknown is true is simply not relevant to the fully general understanding of the paradox. What is crucial is not the question of the truth or falsity of the claim that nothing unknown is true, but rather the modal status of this claim. It is, by all appearances, just another contingent truth if it is true at all, and it is the juxtaposition of disparate modalities that is the heart of the knowability paradox.

CONCLUSION

Two conclusions can be drawn from our discussion. The first is that it is very hard to find a motivation to abandon classical logic based on issues concerning decidability once one has endorsed the claim that all truths are knowable. In a way, this point confirms the attempts by anti-realists canvassed in the last chapter to find some way to endorse the idea that truth is epistemic without endorsing the universal knowability claim,

but since that project is ancillary to the issues involved in the knowability paradox itself, I will not pursue that point any further. The second and more important point is that the attempt to save anti-realism by a change in logic is not promising. Given intuitionistic principles, one can derive that nothing unknown is true, but one cannot defend the acceptability of this claim by insisting that we adopt some intuitionistic meaning postulate about what negation or implication means. First, too much reliance on semantic revision undermines the possibility of a substantive philosophy of logic, but the more important point is that the semantic revision proposal misdiagnoses the heart of the paradox. It treats the paradox as fundamentally about whether it is obviously false that nothing unknown is true, when a proper appreciation of the heart of the paradox recognizes that the issue is really not about truth at all but rather about modality. The proper conclusion to draw is that adopting intuitionistic principles of reasoning will not help avoid the paradox. It merely changes what is paradoxical from a lost distinction between known truth and knowable truth to a lost distinction between unknown truth and unknowable truth. In both cases paradox remains, so we must look elsewhere for a solution to the paradox.

6

Semantical Moves

To summarize where we have been to this point, a bit of historical perspective on the paradox will be helpful. Responses to the paradox fall roughly into two time periods. The earlier time period saw attempts to rescue anti-realism through appeal to intuitionistic logic, and the more recent time period has seen attempts to rescue anti-realism through appeal to various restrictions on the claim that all truths are knowable. This division is only a rough generalization, for there are counter-examples in the literature to the claim as it stands that are easy to find. It is, however, roughly true, and in any case, the important point is the two different approaches to the paradoxes, not the time periods. As we have seen, the first approach of trying to rescue anti-realism through appeal to intuitionistic logic is guilty of the ploy of asking that we adopt new understandings of key terms in the paradox in order to show that nothing paradoxical has been affirmed. Hidden in this request is the idea that if intuitionism is correct, then the paradox need not trouble us, but that conclusion is mistaken. The paradox of knowability is not centrally about whether it is implausible to hold that all truths are known or that nothing unknown is true. It is, rather, about a lost logical distinction between actuality and possibility, and no reinterpretation of negation or implication or quantification tells us anything at all to assuage our perplexity at this loss.

The second approach, involving restrictions on the knowability claim, misses the point of the paradox as well, but in a different way. Conceived historically, this approach makes quite a bit of sense, for Fitch's proof has been seen to be a threat to anti-realist conceptions of truth, and early discussion by Hart, Mackie, and others focused on the threat to such conceptions, especially in the context of verificationist theories. This threat exists, but the fundamental paradoxicality revealed by that proof is not that anti-realism may be false. The fundamental paradoxicality involves, to repeat, the loss of an obvious distinction. It involves

collapsing the distinction between universally known and knowable truth, and even if some restriction strategy were successful in rescuing anti-realism from refutation by Fitch's proof, the fundamental paradox remains. No matter what one thinks of the nature of truth, one ought to hold that there is a distinction between universally known and knowable truth (or between universally unknown and unknowable truth).

Seen from this light, nearly the entire corpus of literature on the paradox is guilty of one gigantic *non sequiter*; in fact, if my arguments have been even marginally successful, the reader should have been chafing at each new proposal as at another new portrait in the halls of the philosophically inapposite. My hope is that by this point, the reader has felt an inexorable push back to the original, intuitive response to the paradox, according to which there must be *some* logical mistake in it. The problem is to identify some plausible approach to the logic of the paradox that allows us to maintain that it involves such an error. The approach I will develop and defend here attempts to find such a mistake, arguing that the mistake is in the general category of failure of substitutivity within intensional contexts. It is an approach I have defended elsewhere,[1] and which Timothy Williamson has criticized.[2] Williamson's remarks show the need for both an elaboration and motivation for the view as well as a defense of the proposal against objections to it. I begin with the elaboration and motivation for the view, after which I will turn to Williamson's concerns, of which there are three. We will find the first two easiest to address, but will spend considerable time on the third problem, after which I will attempt to give a philosophical defense of the proposal outlined next.

A NEW APPROACH: FAILURE OF SUBSTITUTIVITY INTO INTENSIONAL CONTEXTS

Substituting into intensional contexts is logically precarious, and examples of apparently fallacious arguments that do so abound. Here are two well-known examples:

The number of planets = 9; 9 is necessarily greater than 7; so the number of planets is necessarily greater than 7.

[1] See my "The Knowability Paradox and the Prospects for Anti-Realism."
[2] See *Knowledge and its Limits*, chapter 12.

Lois Lane is in trouble and wishes Superman would arrive to rescue her; Superman is Clark Kent; so Lois Lane wishes Clark Kent would arrive to rescue her.[3]

In the knowability paradox, substitutions occur in intensional contexts when we take an instance of an unknown truth and substitute it into the knowability claim, thereby deriving the claim that the unknown truth is knowable. It is this step in the proof that contains a mistake, a mistake of illicitly substituting into an intensional context, or so I shall argue.

The approach I take requires a distinction between sentences in a language and propositions, and it requires a distinction between sentences having the same meaning and sentences expressing the same proposition. Given these distinctions, we can divide sentences in a language into indexical and non-indexical sentences. An indexical sentence is one that can express different propositions in different contexts, and an indexical term is one that facilitates reference in a given context by being the kind of term that, in virtue of its linguistic meaning and function, allows the term to refer to different items in different contexts. The usual examples of such sentences contain pronouns: 'I am here now', 'It is raining', 'He is tired', etc. When two persons both assert the sentence 'I am here now', their assertions have the same meaning, on this proposal, but they express different propositions. For each person expresses a proposition involving himself or herself,[4] and no two propositions are identical if they involve different components.

This apparatus can be analogized to provide a response to the knowability paradox. One can treat quantified sentences, I claim, as modally indexical. That is, the same quantified sentence can be held to express different propositions in different possible worlds. A similar claim regarding extensional contexts is relatively common. Each semester I say to my logic class, "Everyone must take the final; no one is exempt." What I say is strictly false, unless we take the quantifier in question to be a restricted one. If we take it to be a restricted quantifier, then the quantifier behaves in an indexical fashion, expressing different

[3] Some theorists try to explain away this latter example rather than use it to place restrictions on substitutivity into modal contexts. To do so, one will have to adopt a triadic conception of intensional attitudes to explain the difference in the cognitive significance of "I want Superman" from "I want Clark Kent" for Lois. For a defense of this approach, see Nathan Salmon, *Frege's Puzzle* (Cambridge, Mass., 1986).

[4] Perhaps by being a Russellian proposition, so that the proposition has the individual in question as a constituent. The Russellian option is not the only one available, however. If each person has an essence, then the proposition can be about that person in the relevant sense by having that person's essence as a constituent.

propositions in different contexts. On the restricted interpretation of quantifiers, this statement expresses a different proposition each semester I make it, because the quantifier is implicitly restricted to a domain including all and only students in my logic class that semester. To treat quantified sentences as modally indexical does not require the stronger view that all quantification in extensional contexts is restricted quantification, however. The point is simply that the idea that quantification is modally indexical can be understood by analogy with such views.

On the indexical approach to quantification, the sentence 'All humans are mortal' expresses a proposition involving everything in the actual world, to the effect that each such thing has the conditional property of being mortal if human. The proposition expressed, on this view, involves the actual domain, in much the same way that the proposition expressed when I utter, 'I am tired' involves me. If the very same sentence with the very same meaning is asserted in a different possible world with a different domain, a different proposition is expressed. The proposition expressed in such a world by that sentence would be a proposition to the effect that each of the members of the domain of that world has a certain conditional property.

Such an account of quantification has the consequence that quantified sentences are modally indexical with regard to the proposition they express. This theory does not destroy the distinction between indexical and non-indexical sentences, for any sentence containing just a logical constant and a predicate would still be a non-indexical sentence.[5] Moreover, any sentence having as constituents only logical constants, predicates, and the usual connectives would be non-indexical as well, as would quantified sentences having restricted quantifiers ranging over necessary entities such as those of mathematics. In addition, if we distinguish *de dicto* from *de re* readings of modal claims solely in virtue of the scope of the modal operator, this account will not automatically bar inferences involving *de re* modal operators, such as occurs in the following argument:

> Everyone with brown hair might have had black hair; Michael has brown hair; hence Michael might have had black hair.

[5] They would, presuming that we specify that the domain includes properties that are the semantic value of predicates, rather than construing this value to be the extension of the predicate. Otherwise, the identity conditions on propositional analogues of predicates would be purely extensional, yielding the result that predicates are modally indexical as well. There are good reasons to avoid such an account that are well known, and if we

When the modal operator is inside the scope of the quantifier, the interpretation of the quantifier occurs first, leaving no grounds for concern that the interpretation of the possibility operator might undermine the validity of this argument.

This point should not be interpreted to mean that every argument in which a modal operator occurs on the consequent of some conditional is valid. My point is that it is an open question in such cases whether the approach I recommend invalidates the argument. For example, if one quantifies over propositions and attributes a modal property to them, the approach I take here will disallow inferences involving them unless special care is taken to insure that the propositions over which one quantifies are ones that exist in all worlds. In cases such as the example above, no such problematic quantification over propositions is involved, and hence the neo-Russellian view of quantification fails to have any implications at all, leaving us free to respect the obvious validity of the argument.

The costs of the view ought to be noted. First, if quantified sentences are indexical, the Fregean hope for a logically perfect language must be given up when intensional operators are added to standard first-order theory. Second, and more important, the Rule of Necessitation will have to be restricted. If quantified sentences express different propositions in different worlds, then the necessity of that proposition follows no more from its provability in any context than does the necessity of the claim that I exist from the fact that 'I exist' is true in every assertion context.

This cost may seem high, but one way to motivate allowing it is to tell a story about necessitation that makes sense of the restrictions in question. The example above illustrates how such a story might go. Provability is more akin to analyticity than it is to necessity. To show that a formula is provable is to show that it can be proved in any context whatsoever. But from that fact it does not follow that the proposition expressed by the formula is a necessary truth, in precisely the same way that noting that 'I am here now' cannot be falsely uttered fails to show that such a sentence expresses a necessary truth when I utter it. Once one sees cracks in the cement that ties analyticity and necessity together, the indexical theory of quantifiers serves merely to show how large the crack is. Roughly, the rule of necessitation can be used only on formulas that express the same proposition in any context, on formulas that are rigid in

accept an intensional account of the semantic value of predicates, then the propositional ingredient expressed by a predicate does not change from world to world.

the sense that some take proper names to be rigid. Such a restriction will allow the rule of necessitation to be applied when dealing with mathematical concerns. In such cases, we can use quantifiers to range only over the entities needed for such discussions and it is plausible to assume that the range of such quantifiers will not vary from world to world—that is, to hold that the objects of mathematical discourse are necessary.

It is also worth noting that no similar restriction is needed on the elimination rule for the necessity operator

$$\Box p \rightarrow p,$$

or its corollary introduction rule for possibility

$$p \rightarrow \Diamond p.$$

In particular, the latter principle is acceptable because, in semantic terms, the actual world is accessible from itself and so the consequent will be true in exactly the same world as the antecedent, regardless of any details regarding the syntactic elements of 'p'.

Another logical difference resulting from this approach is that some inferences will be licensed on the neo-Russellian theory that are not licensed on the standard theory. For example, consider the following three claims:

1. $\Diamond \forall x(Sx \rightarrow Fx)$
2. $\forall x \Diamond (Sx \rightarrow Fx)$
3. $\forall x(Sx \rightarrow \Diamond Fx)$.

Each of these three formulas disambiguates the ordinary English sentence 'The S's might be F's'. Consider, then, the argument from each of these claims together with

4. Sa

to the conclusion

5. $\Diamond Fa$.

On the indexical theory of quantification, each argument is valid, but on the standard interpretation, only the latter two are valid (since a world with no S's would make formula 1 true, but leave open the possibility that 5 is false).

Though this difference may be viewed as a liability for the neo-Russellian view, I do not think that is the appropriate view to take. If I am a North Dakotan, and it is possible that any North Dakotan is

beautiful, it is slightly jarring to find out that it is impossible for me to be beautiful. How could that be? The standard explanation requires us to talk about the possibility that I'm not a North Dakotan while, in the same possibility, any North Dakotan to be found is beautiful. If there are no theoretical costs to letting this explanation go through, we might allow it; but we needn't act as if it were intuitively obvious.

In the context of the particular substitution involving possible knowledge, the intuitive response goes the other way, however. Suppose it is true that all North Dakotans are Scandinavian, but that no one knows this fact. But suppose there is a world in which there are no North Dakotans, and this fact is known. By the principle of closure of knowledge, one who knows that there are no North Dakotans could come to know what would be expressed in that world by the quantified sentence 'every North Dakotan is Scandinavian'. Suppose such a deduction is completed; would we say that this achievement constituted a truth-maker for the claim that it is knowable that every North Dakotan is Scandinavian? I suspect that most will incline against saying that. Perhaps we could put the contrary point as follows: the truth-maker for the claim that the Scandinavian ancestry of North Dakotans is knowable ought to involve a possibility in which the ancestry of North Dakotans came to the attention of some possible cognizer. To get that result, however, one must reject the Fregean picture of quantification inside modal contexts, at least the modal context that mixes possibility and knowledge operators. There are cases that seem to cut the other way, however. Jason Stanley and Zoltán Szabó offer the following:

> Suppose that John has a strange habit of buying exactly 70 bottles every time he goes to a supermarket. Suppose that John visits a supermarket that has exactly 70 bottles on the shelf, and purchases every bottle. Someone could then truly utter the sentence:
>
> > If there were a few more bottles on the shelf, John would not have purchased every bottle. (37)
>
> However, if we assign to the contextual variable associated with 'every bottle' the set of bottles in the supermarket in the context of utterance of John's sentence, given the standard semantics for counterfactuals, (37) could not be truly uttered. To capture the reading of (37) on which it is true, one must treat the entity assigned to the contextual variable as a function from worlds and times to, say, the sets of bottles in the relevant supermarket at those worlds and times.[6]

[6] Jason Stanley and Zoltán Szabó, "On Quantifier Domain Restriction", *Mind and Language* 15 (2000), p. 252.

Stanley and Szabó use (37) to argue that the semantic value of a quantifier in a modal context must be a function from worlds and times to objects, rather than a set of objects. For if the semantic value of the quantifier in the consequent of (37) were the actual set of bottles on the shelf, then (37) would not be true. Since it is intuitively true, the neo-Russellian theory of quantification must be mistaken.

This argument is hasty, however, for there is another possibility to consider other than that (37) is true. That possibility is that (37) is false and that our sense that it is true is due to pragmatic factors that allow us to see clearly which truth is being pointed to, which is:

> If there were a few more bottles on the shelf, the sentence 'John purchased every bottle' and its translational equivalents would not have expressed a truth. (37)*

This use of the idea of translational equivalents will raise a red flag with those who know the history of appeals to such in semantic contexts. In general, it is a mistake to appeal to sentences and their translations when explaining the semantic content of a sentence. The most obvious reason for this fact is that if the sentence contains no reference to language itself, then the truth expressed by the sentence does not imply the existence of language, whereas any sentence that refers to sentences and translational equivalents does imply the existence of language. Hence, the meaning of the two sentences could not be the same.

It is important to notice that this problem does not affect the point here. (37)* is not being offered as meaning-equivalent to (37). It is not equivalent in this way. The relationship between the two is quite different. The explanation offered above claims that (37) is, strictly speaking, false. (37)* is, however, a closely related truth, and or ordinary communication practices often allow us to communicate truths by asserting claims that are strictly false. Hence the worry about appeal to translational equivalents in semantic contexts is irrelevant to the point being made here.

This treatment of (37) is akin to what we should say about the following kind of case. Matt is well over six feet tall, and I am not six feet tall. Given this information, consider the following counterfactual:

If the present context is such that 'I' refers to Matt, then I would be at least six feet tall. This counterfactual is false, but it is not, let us say, ludicrously false. If anyone other than Matt were to utter this sentence, we'd know pretty well what they were trying to say. They would be

trying to say that if 'I' refers to Matt in the present context, then the sentence 'I am six feet tall' would express a truth.

In the Stanley and Szabó example, we have the same kind of hesitation in ascribing truth to the statement in question, though perhaps not as much hesitation as in the Matt case. We can get ourselves in the frame of mind to reject (37) by reasoning as follows:

> If there were more than 10 bottles on the shelf, John still would have bought every bottle on the shelf, because what John did was to buy every bottle on the shelf and a world where he does the same thing is closer to the actual world than one in which he does something different. So (37) is false.

My point here is not to endorse this reasoning, but to display its plausibility so that it is clear that whether one's theory of quantification treats (37) as true or treats it as false, one will have to explain away the alternative intuition.

I propose, then, that the discussion of the indexical theory to this point provides at least a hint of respectability for the indexical theory; enough respectability, I suggest, to see what comes of the idea—whether it can withstand scrutiny and what are its implications for the knowability paradox. On the latter issue, the implications are straightforward: the paradox fails on this understanding of the quantifiers. The mistake in the proof occurs with the substitution of an instance of the second assumption into the first assumption as one of the knowable truths. Since propositions are the objects of knowledge, such a substitution is legitimate only if the formula expresses the same proposition in the substitutional context that it expresses in the original context. In the present case, the substitutional context is partially a modal one, for the consequent of the bound conditional in the first assumption is governed by a possibility operator. So for the substitution to be legitimate, the formula would have to be modally non-indexical. Otherwise the unknown proposition expressed by that formula in the actual world may not be the expressed value of that formula in the modal context in question. Since the substituted formula is a quantified one and quantified sentences are generally modally indexical, the argument fails because of an illegitimate substitution in a modal context. As a result, one is not forced to deny the logical distinction between universally known and universally knowable truth.

The appeal to the indexical view does not provide a complete solution to the paradox yet, however. If ordinary quantifiers are indexical, why not just introduce new ones that are stipulated to be non-indexical? In cases in which ordinary quantifiers seem to be restricted, such as the case above concerning my remark to logic classes that everyone must take

the final, we can easily stipulate that our formal language contain only unrestricted quantifiers. Doing so will force us to represent the implicit restrictions in natural language as part of an expression bound by the unrestricted quantifiers, and no problems arise from treating such implicit restrictions in this way. So, why not require the same in the case of the knowability paradox?

In the case of my remark to my logic class, representing such a remark with unrestricted quantifiers merely requires specifying in the antecedent of a conditional what part of the entire domain of the universe is in question. That is, unrestricted quantifiers range over the entire domain of discourse, with restrictions introduced within the expressions governed by such quantifiers. If we want to make an analogous maneuver when we are dealing with modal contexts such as that involved in the knowability paradox, we will have to introduce modally unrestricted quantifiers, i.e., quantifiers that range over the union of the domains of every possible world, that is, over the entire domain of possibilia. Anything less will still be susceptible to indexicality. In order to consider the above suggestion, then, we need to consider how to represent the two assumptions of the knowability paradox in terms of quantified formulas using such modally unrestricted quantifiers.

Using italicized versions of the usual quantifiers to be our modally unrestricted quantifiers, we can represent the claim that all truths are knowable as:

$$\forall p(p \rightarrow \exists x \exists t Kxpt), \tag{1}$$

which says that any truth is known at some possible time by some possible being. The second claim cannot be represented with the same structure as before, substituting italicized quantifiers for unitalicized, that is, as "$\exists p(p \,\&\, \sim \exists x\, \exists t K x\, Tpt)$", for that says that there is a truth which is not known by any possible being at any possible time. Such a claim is obviously inconsistent with (1); it amounts to nothing more than a denial of the first claim, and hence must be an improper representation of the claim that some truths are not known. To remedy the problem, we must introduce some restrictions into the expression which our unrestricted quantifiers govern. We can do so using an actuality operator:

$$\exists p(p \,\&\, \sim \exists x \exists t @Kxpt), \tag{2}$$

where '@Kxpt' is to be read as "the knowledge of p by x at t is actual", thereby requiring that the possible x be an actual x and the possible t an actual t and for the knowing itself to be actual as well.

There are two different semantical treatments one might give of this actuality operator. One way is to treat it as a rigid designator of the actual world, and the other is to treat it as picking out the actual world indexically. In the second case, sentence (2) would express a different proposition in each different possible world. In the first case, it would allow a sentence to express the same proposition in each world (as long as we hold other dimensions of interpretation fixed). Noting this difference shows that the only hope for resolving the paradox resides in interpreting the operator in the first way. For if (2) is an indexical sentence, the same substitution problems encountered earlier for indexical sentences will plague any argument employing this formula as well. Hence, to avoid the problem with the original proof, the only way to interpret the operator is on the model of a rigid designator.

On this interpretation, however, the results of the proof strategy of substituting (2) as the value for a knowable truth in (1) are no longer paradoxical. All that follows from the fact that some instance of (2) is true is that there is a world in which the following is true:

$$\exists x \exists t K(p \,\&\, \sim \exists x \exists t(@Kxpt). \qquad (3)$$

(3) only says that some possible being knows both that p is true and that no being in the actual world knows that p is true. Nothing paradoxical follows from such a claim, and the lack of paradoxicality is just like that found in Edgington's approach to the paradox.[7]

Note that I am not claiming here that (3) is true, or that sense can be made of the idea of knowing from the vantage point of one world what is going on in another world. My only claim is that no paradox, no contradiction, is derivable from (3), and that claim is the positive contribution I find in Edgington's approach to the paradox, even if her solution founders on other issues.

So the indexical theory of quantifiers undermines Fitch's proof and hence blocks the formulation of the knowability paradox. If quantifiers are indexical, the proof relies on an illegitimate substitution into an intensional context. If this problem is repaired by introducing non-indexical quantifiers, such quantifiers must range over possibilia. If they do, the representations of the two assumptions must be altered from those at the beginning of this paper, and given the appropriate alterations, no paradoxical result follows.

[7] Dorothy Edgington, "The Paradox of Knowability", *Mind* 94 (1985), pp. 557–568.

THE PROSPECTS FOR ANTI-REALISM

This approach to the paradox leads directly to the question of what formula might be used in the modal context to express the appropriate propositional content in question, the proposition that specifies an unknown truth. The truth of anti-realism turns on this issue, for if the proposition is knowable, then there must be a possible world where that proposition can be known. Since the proposition in question may not be expressed by the sentence that expresses it in the actual world, it would seem that some other sentence must be found to do the job. Just what sentence can do that, however?

I want to explore this issue a bit below, but before doing so I want to make clear that this issue does not affect the adequacy of the indexical response to the knowability paradox and, as such, the question for this section is a detour from the main line of argument of this chapter, which is to explain the indexical view of quantification, defend it against objections, and provide a positive defense of the proposal. The paradox itself concerns the logical distinction between two concepts, the concept of a known truth and the concept of a knowable truth. The solution I propose to the paradox raises a further question, the question of whether the universal knowability assumption is, or can be, true. If it cannot, anti-realism is in jeopardy, but the solution to the paradox stands. For the solution only needs to demonstrate that the logical distinction between known and knowable truth does not collapse on the basis of Fitch's proof. Such a solution may not rescue anti-realism from the paradox, if it preserves the logical distinction in a way that leaves the anti-realist with no resources for explaining how an unknown truth can be known. The resolution of the paradox is one thing, however, and the sustainability of anti-realism in the face of the paradox is another.

Nonetheless, since the knowability paradox raises the specter that anti-realism must be abandoned, it would be nice if a solution to the paradox left the prospects for anti-realism an open question. The solution I propose, however, seems not to do so, in virtue of the difficulty of being able to express propositions about the actual domain in other modal contexts, and if such a proposition cannot be expressed in a different modal context, there will be no such context available for the anti-realist to point to when explaining how an unknown truth is capable of being known to be an unknown truth. I will argue that the

appearances here are misleading, that the solution I propose leaves anti-realists some room to maneuver in defending their view of truth.

The question we face is how to express in other possible worlds propositions expressed in the actual world by modally indexical sentences. We can begin our search for an answer to this question by first noting how ordinary indexical sentences work. When Hume say, "I am tired", some function from interpreted sentence and actual first-person context yields expressed proposition. Heimsohn, taking Hume at his word, believes the very same proposition that Hume expresses.[8] He need not do so by translating Hume's sentence into the third person "He is tired", though of course he could. Instead, because Heimsohn knows the language, he knows that Hume is making a self-reflexive utterance. Such knowledge pins down in a suitable fashion the domain on which to interpret the utterance (i.e., the person producing the utterance), and Heimsohn grasps the proposition Hume expresses in virtue of the same function on interpreted sentence and assumed context that yields the propositional content of Hume's utterance.

Similar remarks can be made about model theory. When we do model theory, we take a quantified sentence and interpret it with respect to a given domain. The sentence (and its linguistic meaning) stays the same from model to model, but the truth-value varies. We grasp the proposition (or formal surrogate of such) by knowing the meaning of the sentence and the assumed context of use, including the domain in question. So, grasping propositional content is a function of sentence, meaning, and assumed context. We do not interpret by finding some sentence that expresses the same proposition in every context nor do we interpret by imagining a collection of sentences each of which in its own context expresses one and same proposition. We don't have to purge language of indexicality to understand it, and we don't have to envision or entertain other sentences that would express the same proposition in other contexts. To know the meaning of an indexical sentence doesn't require adverting to other sentences in this fashion.

The model theory example is a formal analogue of what the anti-realist needs to say about expressing propositions in other worlds that are expressed by modally indexical sentences in the actual world. We grasp such propositions when we have a grip on the relevant sentence

[8] These remarks presuppose a triadic conception of intensional attitudes. For more on such issues, see Nathan Salmon, *Frege's Puzzle* (Oxford, 1989).

and its meaning, and can grasp the relevant context in which to interpret the sentence. So, if we are imagining possible worlds with individuals in them attempting to grasp what is expressed by modally indexical sentences in our world, such residents could know the propositions expressed by quantified sentences in our world as long as they can specify the domain of our world. If they can do that, they are doing something analogous to what we do more formally when we calculate the truth-value of a quantified sentence on a model.

One might worry that there is no way to refer to the domain of another possible world. This worry may be correct. Even so, one can still fix the domain indirectly *via* description. Imagine a world just like the actual one, except that it contains one extremely gifted cognizer, one that can hold before its mind all the individuals in that world. Such an individual could then conceive of the actual domain: just imagine the domain for its own world minus itself.

Given the capacity to grasp the domain, knowing that the truth-value of the proposition expressed relative to a particular domain by the sentence 'the truth that q is not known' is relatively trivial. The gifted cognizer, for example, could be aware that it was the most gifted cognizer in its world. Such a cognizer could even know that some of its knowledge was necessarily unique among the cognizers of its world. It could thus know that some proposition q was true and not known by anyone in its world but itself. Such a being could thereby know that q was true but unknown in the actual world, for by hypothesis the only difference between the two domains is the existence of the gifted cognizer.

What if there are multiple worlds of the sort in question, that is, multiple worlds in which q is not known and the world has precisely the same elements in the domain as does the actual world? The anti-realist needs some reason here for thinking that the knowledge possessed by the possible gifted cognizer is sufficient for knowing that the claim in question is an unknown truth in the actual world. This problem is akin to the problem for Edgington's approach to the theory, for on her view, denizens of other worlds help to rescue anti-realism by knowing what is true in the actual world, and thus need to be able to grasp propositions about the actual world.[9] The problem for Edgington's proposal seems insuperable, as we saw in Chapter 3, but the problem here leaves a bit more hope for

[9] See Edgington's "The Paradox of Knowability", and the critical discussion of it by Timothy Williamson in "On the Paradox of Knowability."

the anti-realist. What is needed in the present context is only some way of expressing the very same proposition from the vantage point of a different possible world. The anti-realist need not find a way for some denizen of another world to be thinking about the actual world and what is true in it. As such, it doesn't matter whether there is only one world, the actual world, in which q is true or whether there are several such worlds. The existence of multiple worlds threatens Edgington's requirement that the denizen of another world know what is true in the actual world. But it does not threaten the possibility of such a denizen grasping the content of a proposition that, as a matter of fact, is expressed in the actual world by the claim that a certain proposition is an unknown truth.

There are complexities here that I have not noted, and I wish to make it clear that I make no pretense of having provided a full response on behalf of anti-realism to these problems. All I have shown is that the particular problem, as presented, is amenable to a solution for anti-realism. The defense leaves open the possibility that there are other problems, or more sophisticated versions of the present problem, that cannot be answered in this way. If so, anti-realism will be indefensible.

The central point to keep in mind, however, is that the defensibility of anti-realism from this problem and related ones presents no challenge to the indexical response to the knowability paradox. On that solution, we treat quantifiers as modally indexical. Such a move implies that the substitution into a modal context in the paradox is illegitimate because the second conjunct of the substitution is a quantified formula. When such a substitution occurs involving indexical sentences, the substitution is legitimate only if there is some guarantee that the domain of the modal context into which the substitution is made is the same as the actual domain. As a result, anti-realists face a challenge from the paradox even granting the success of the indexicality response to it, for this response to the paradox raises the specter that some propositions may be expressible only from a world with the same domain as the actual world.

I have suggested why I doubt that anti-realism is doomed by such considerations. If there are propositions expressible only from within certain possible contexts, anti-realists will begin to squirm. The point I want to emphasize here, though, is that the demise for anti-realism such truths might engender does not reach to the indexical theory itself. Such a result would be a surprising development in our understanding of the paradox, but the possibility exists that anti-realists can take no comfort in the present solution to the paradox.

DETAILS OF AND QUESTIONS ABOUT THE THEORY

We return, then, to the main purpose of the chapter, which is the clarification and defense of the indexical theory of quantification. On the theory in question, we divide sentences in a language into two groups, in terms of the relationship between sentence and proposition expressed. Some sentences, such as those involving indexicals and demonstratives, require a specification of context in order to determine which proposition the sentence expresses. Other sentences are context independent. If there were a separate and unique name for every item named in a language, sentences involving names would be an example of the latter. Similarly, if the language contains no ambiguities, then sentences concerning relationships between properties typically will be context independent, e.g., red is a color. Context independence thus often requires a bit of idealization regarding natural languages, but no matter what amount of idealization is involved, sentences involving indexicals and demonstratives will never come out context independent. So even if the distinction is not a clean one applied to natural languages, there is a distinction between sentences which are context independent in suitable idealizations of the language and sentences which will remain context dependent under such idealization.

One standard type of flaw in inferences involving intensional contexts is failure of substitutivity, and in the knowability paradox, a substitution must occur in such a context. In order to generate the paradox, we take a particular unknown truth p, and use $p\&\sim Kp$ as the value for p in $p \rightarrow \Diamond Kp$. The phenomenon of context dependence arises in this substitution in the following way. Since we know that a sentence in a language can require specification of a context in order to determine proposition expressed, the question needs to be asked whether a sentence in one possible world expresses the same proposition in every other world (holding meaning fixed, of course). By analogy with the way in which indexicals generate context dependence in extensional contexts, I term this possibility "modal indexicality". The terminology is not important, however, and it is analogical at best; what matters is the phenomenon itself. If we cannot guarantee that the sentence 'p&~Kp' expresses the proposition $p\&\sim Kp$ in every possible world (as Kripkeites might put it, "is (analogous to) a rigid designator"), then the

substitution should be rejected (for roughly the same kinds of reasons that prohibit substituting '9' for 'the number of planets' in intensional contexts).[10] (I note also that this condition for failure of substitutivity is sufficient but perhaps not necessary: 'Clark Kent' and 'Superman' co-designate and, in a suitably idealized language, rigidly designate, but there are concerns about intersubstitutivity of these in intensional contexts nonetheless.)

Are there any grounds available for supposing that 'p&∼Kp' might be modally indexical? Yes there are, since the K-operator includes embedded quantifiers (it is to be read "it is known by someone at some time that"), and, as I have been arguing, there is a theory of propositions available that treats quantifiers in a way that makes sentences involving them modally indexical.

One theory of propositions, the standard one, models the expression relationship between sentences and propositions on the rigid designation idea. We might call this theory "the Fregean theory". For a simple quantified sentence, '∀xFx', the Fregean theory claims that the proposition expressed involves, as constituents, a first-order property *being F* and the particular second-order property expressed by the universal quantifier, perhaps *having no counterinstances*. Such a viewpoint makes the relationship between sentence and proposition very much like involved in the use of a sentence expressing a relationship between properties, such as "red is a color". Since the latter is a paradigm example of a context independent sentence, the Fregean theory can be taken, with good reason, to hold that quantifiers introduce no element of modal indexicality. The key element of the Fregean theory is to identify the constituents of proposition expressed with properties, with the background assumption that the same properties exist in all worlds, thereby rendering context-independent any sentence that expresses a proposition constituted only by properties.

It is well-known that this Fregean theory has difficulty with other types of sentences. For example, it has difficulty with sentences involving names, since it takes quite a bit of creative theorizing to find a property to be the semantic value of a name.[11] A simpler theory is that a name

[10] My use of the examples involving the number nine as the number of planets assumes that there really are nine planets. As I understand recent science, that assumption is no longer thought to be true, but I will leave the examples resting on this mistaken assumption because of the history of such examples in the literature and the fact that it is not yet common knowledge that there aren't nine planets.

[11] For an example of the kind of gyrations needed to find such a property, see Alvin Plantinga, "The Boethian Compromise", *American Philosophical Quarterly* 12 (1978), pp. 429–38.

has what it names as its semantic value and hence that the proposition expressed by a sentence involving that name will have the individual named as a constituent. We might call this theory "the Russellian theory". Such a Russellian theory does not need to despair of the context-independence of sentences involving names, however, as long as we are dealing with a suitably idealized language in which all names are unique, so context independence is an issue orthogonal to whether one adopts a Russellian or Fregean approach to names. When we consider indexicals and demonstratives, however, things are different. The same reasons for being a Russellian about names give one reason to be a Russellian about indexicals and demonstratives, holding that the thing referred to is a constituent of the proposition expressed.[12] Moreover, since sentences involving indexicals and demonstratives refer in a way that is context-dependent, this Russellian account of the expression relation between sentences and propositions renders indexical sentences inappropriate for substitution into intensional contexts.

A neo-Russellian approach to quantification can be developed modeled on this Russellian treatment of indexicals and demonstratives. On the Fregean picture, as I use the term here, semantic values are always and everywhere properties, which are constituents of the proposition expressed. The distinctive feature of a Russellian approach is that sometimes linguistic elements do not express properties, but pick out elements of the domain which are themselves constituents of the proposition expressed. Applying this Russellian idea to quantifiers, the theory of propositions that results places the domain of quantification into the proposition expressed. Thus, the sentence '$\forall$xFx' expresses, on this neo-Russellian approach, a proposition that has as constituents *being F* and the domain of quantification (if one is a fan of the arguments for restricted quantification, then the propositional constituent will be the relevant subset of the entire domain; if one opts for unrestricted quantifiers, then the entire domain will be a component of the proposition expressed).

This neo-Russellian theory challenges the legitimacy of substituting quantified formulas into intensional contexts, for there is no guarantee that the same proposition will be expressed in the new context. That is so because *x* and *y* are the same proposition only if they have the same

[12] Lest I be misunderstood in my use of the term 'Russellian', I note here that my use of the term diverges from Russell's views. Russell thought that only the demonstratives 'this' and 'that' were logically proper names, and that ordinary names such as 'Joe' or 'Frank' were not logically proper names (because only things with which we are directly acquainted can be named). See *The Problems of Philosophy* (Oxford, 1912).

constituents, and on the neo-Russellian view of quantifiers, quantified sentences may express propositions with different constituents in different possible worlds. In virtue of the possibility of such a neo-Russellian view, the knowability paradox rests on an unsuccessful derivation of Fitch's result unless and until reasons can be given for rejecting the neo-Russellian view.

As already noted, there is one caveat to this claim, that a defender of the paradox might introduce unrestricted quantifiers ranging over the union of the domains of all possible worlds, i.e., ranging over possibilia. Using these fully general quantifiers, one might try to resurrect the paradox on grounds that propositions expressed by sentences have such quantifiers in them will not vary from world to world, even on the neo-Russellian view (since all possibilia would be in every such proposition).

My response to this attempt is to argue that the distinction between strong and weak verificationism collapses on this view. Strong verificationism is the view that all truths are known, ordinarily formulated as

$$\text{SVER } \forall p(p \rightarrow \exists s \exists t K s p t).$$

Weak verificationism is the view that all truths are knowable, formulable as

$$\text{WVER } \forall p(p \rightarrow \Diamond \exists s \exists t K s p t).$$

Since these quantifiers are ordinary ones, they need to be replaced by fully general ones ranging over possibilia if the paradox is to be resurrected by such devices, and I indicate the range of such quantifiers by italicizing them:

> *SVER* $\forall p(p \rightarrow \exists s \exists t K s p t)$(Every truth is known by some possible being at some possible time);
>
> *WVER* $\forall p(p \rightarrow \Diamond \exists s \exists t K s p t)$(Every truth could be known by some possible being at some possible time).

The standard paradox says that WVER is problematic because it implies SVER. The resurrected paradox will have to claim that *WVER* is problematic because it implies *SVER*. I agree that the latter implication holds, but deny that it is problematic. It is not problematic because the two say, in essence, the same thing. *SVER* says that, once you find a truth, you can find some possible being who knows it at some possible time. So, the semantical treatment of the proposition sends you out into the array of possible worlds, once you've latched onto a truth, and tells you that you'll be able to find some world in which the claim is known. *WVER* merely prefaces this latter with another possibility operator, and

given idempotence for possibility (i.e., $\Diamond\Diamond p \dashv\vdash \Diamond p$), the two claims are such that no logical distinction exists between them. Another way to put this point is that *SVER* says the same thing as WVER, that all truths are knowable, while *WVER* only adds a double modal, saying that all truths could be knowable. Double modals are a fascinating linguistic phenomenon (especially to those of us intimately acquainted with dialects in the southern part of the United States), but the logic of this one is fairly straightforward: we might as well identify the two italicized readings by adopting idempotence (in a Southern drawl, say, "We might could identify them."). So the implication loses its problematic character when we move to quantifying over possibilia.

The introduction of such quantifiers becomes an issue only if the neo-Russellian account is not found wanting in other respects, and Timothy Williamson attempts to show that it is.[13] He begins with what I will describe as border-skirmishing, but in the end his objections can be cast in terms of a dilemma for the view. Before examining the dilemma, however, addressing the border-skirmishing will help us to see better the nature of the neo-Russellian proposal and what kinds of objections may and may not be lodged against it.

The border-skirmishing involves Williamson taking issue with my use of the concept of indexicality in describing the neo-Russellian view. He claims that the proper concept is that of rigid designation, and that since I describe the view in terms of "a kind of indexicality", I am confusing the issue. This point is a red herring. I do describe the view using the term "modally indexical", but to infer from that terminology that I think of the view in terms of a kind of indexicality is akin to accusing someone who speaks of former Senators as positing a new kind of Senator. Moreover, the relationship between sentences and propositions is not one of designation, as Williamson's preferred language of rigid designation would have it: sentences don't *refer to* or *designate* anything. So insisting that rigid designation is the proper concept by which to explain the neo-Russellian view is no advance over the language of indexicality should one find that language bothersome. The bottom line on this issue is that both the language of modal indexicality and rigid designation are analogical uses of language in this context, and the heart of the issue does not turn on what analogies are pressed into service to

[13] Timothy Williamson, *Knowledge and its Limits* (Oxford, 2000), chapter 12, section 3, pp. 285–289. Williamson's discussion is aimed at my earlier presentation of the present solution to the paradox in "The Knowability Paradox and the Prospects for Anti-Realism", *Noûs* 29 (1996), pp. 481–500.

convey the central ideas of the theory. The point of the analogies is to get the reader to see the true nature of the theory, and any objection that will carry any weight will be one against the heart of the theory and not the analogies used to get a reader to understand the theory.

After objecting to the language of indexicality, Williamson turns to a different issue, expressing puzzlement at the view. He says,

> We do not expect variation in the extension of 'dog', as uttered in a fixed context, with respect to different circumstances of evaluation to constitute variation in the proposition expressed by the sentence 'Fido is a dog'; why should it constitute variation in the proposition expressed by the sentence 'Some dogs bark'?[14]

Here Williamson contrasts the context-independence of 'Fido is a dog' with the purported context-dependence of 'Some dogs bark' involved in the view developed here, and is perplexed. The answer to his perplexity can perhaps be given by an analogy. Suppose he had said,

> We do not expect variation in the extension of 'dog', as uttered in a fixed context, with respect to different circumstances of evaluation to constitute variation in the proposition expressed by the sentence 'Fido is a dog'; why should it constitute variation in the proposition expressed by the sentence 'That is a dog'?

This analogy makes it clear that Williamson's question only makes sense when one has not yet appreciated the details of the theory in question. On fully Fregean assumptions, according to which properties are the semantic values expressed by linguistic items, any kind of Russellian theory is puzzling. So a thoroughly Fregean perspective might query why 'Fido is a dog' receives a different semantical treatment than does 'Red is a color' (on a standard Russellian view that makes the dog Fido a constituent of the proposition expressed). The answer to such a question would involve explaining the Russellian theory anew, hoping to shake loose some of the bindings of the thoroughly Fregean perspective. Furthermore, even to a theory that allows mild Russellian encroachment to allow sentences with names to express propositions containing the named individual, one might be puzzled by the suggestion that context-dependent demonstratives can function like names; why, that is, 'Fido is a dog' would receive context-independent treatment in an idealized language but 'That is a dog' would not. The answer to the query is to explain the second kind of Russellian view, on which proposition expressed can contain members of the domain in two different ways, depending on whether the relation of expression between a sentence and

[14] Williamson, *Knowledge and its Limits*, p. 288.

a proposition is a two-place relation or a three-place one, with the third place involving context.

A similar approach should be taken to Williamson's question. He apparently doesn't see the aspect of the neo-Russellian view that results in quantified sentences being treated differently than unquantified ones. The answer is to explain again the details of the neo-Russellian view, on which elements of the domain are sometimes propositional constituents, and are so when the sentence that expresses the proposition in question is a quantified ones, hoping the audience will see the difference.

That he doesn't see the difference is confirmed by the next paragraph in which he voices the perspective of the more standard Fregean view on the semantic import of quantification. He says,

> We have no grounds to suppose that a sentence, as uttered in a fixed context, can designate different propositions with respect to different circumstances of evaluation. The idea that it can seems to confuse expression with evaluation. What a sentences expresses is conceptually prior to the procedure of evaluation, and not relative to a circumstance of evaluation.[15]

The idea here is nicely Fregean: there is a difference between the context of evaluation (in which the domain in question plays a crucial role in determining truth-value) and the theory of expression, which is conceptually prior to that of evaluation. No Russellian view will allow this point, however. Elements of the domain encroach, on any Russellian view, into the theory of expression, and are not limited to the context of evaluation. It is true that nothing about indexicality or rigidity of designation gives us a reason to adopt the neo-Russellian view of quantified sentences, but that is beside the point (though we will return later to the question of whether there are adequate grounds for endorsing this neo-Russellian view of quantification). Russellianism of any sort provides a model for such a view, and any version of Russellianism will reject a characterization on which the semantic evaluation of a proposition is wholly distinct from the expression relation between sentence and proposition.

It may be useful to put this point in terms of the standard approach to indexicals in the terminology deriving from David Kaplan's work on demonstratives and indexicals.[16] We distinguish the theory of expression, which involves a function from sentences to propositions, from the theory of evaluation, which is a function from propositions to

[15] Ibid., p. 288.
[16] See, e.g., "Dthat", *Syntax and Semantics* 9 (1978), pp. 221–243.

truth-values. The disagreement between the Fregean and neo-Russellian views is a disagreement within the theory of expression. The usual Fregean treatment of quantification takes its role as that of generating a proposition that has a second-order property as a constituent, so that the existential quantifier in the sentence '$\exists x \phi x$' contributes to the proposition expressed the second-order property that the property expressed by 'ϕ' has at least one instance. The neo-Russellian view of quantification claims, instead, that the quantifier contributes to the proposition expressed the domain of quantification in question. Quantifiers do not express properties, according to the neo-Russellian view of quantification, but refer to domains, thereby generating in the theory of expression a constraint on the function from sentences to propositions such that domains are constituents of propositions.

So much for the border-skirmishing, since some of what Williamson claims involves an apparent failure to appreciate where the true conflict between the Fregean and neo-Russellian views lies. Williamson has objections to the view, however, which do not fail to appreciate the nature of the proposal. He holds that the theory has two problems. One problem threatens the neo-Russellian view itself, and the other problem arises when we supplement our language with a possibilist liberalization of the quantifiers. Regarding the latter problem, Williamson argues that the possibilist liberalization leaves the inference from *WVER* to *SVER* problematic, and regarding the former problem, he argues that the neo-Russellian view makes the contingency of certain claims hard or impossible to express.

Since I think the real issue here is the neo-Russellian theory itself, I want to consider first Williamson's claim that the possibilist liberalization of the quantifiers leaves the paradox intact in order to focus more on what I consider to be the deeper issue. The inference involved in Williamson's complaint about the possibilitst liberalization of the quantifiers is from

> *WVER* $\forall p(p \rightarrow \Diamond \exists s \exists t Kspt)$(Every truth could be known by some possible being at some possible time)
>
> *SVER* $\forall p(p \rightarrow \exists s \exists t Kspt)$(Every truth is known by some possible being at some possible time).

My claim was that the inference is sound, but unproblematic because the two claims are not logically distinct in virtue of the idempotence of the possibility operator. Williamson argues otherwise:

[Kvanvig] supposes that *SVER*, read as 'Every proposition, if true, is known at some possible time by some possible being', says no more than *WVER*, read as

'Every proposition, if true, could be known at some time by some being.' But that is a mistake. Although 'time' and 'being' occur within the scope of 'possible' in the reading of *SVER*, 'known' does not.[17]

If we attend to the parenthetical renditions of the quantified formulas above, Williamson is right: in *SVER* 'known' occurs before any concept of possibility, and the same is not true for *WVER*. This surface grammar point is not telling, however. The reason this order appears as he claims is that the parenthetical rendition is put in the passive voice. We could have put it in the active voice, in which case, the order would not be as claimed (*SVER* could be read as "Every true proposition is such that some possible being at some possible time knows it to be true"). Moreover, it is not the paraphrases into English that carry the weight, it is the formalism itself that matters. And given the possibilist liberalization of the quantifiers, *SVER* involves a concept of possibility preceding the knowledge operator, since the concept of possibility is contained within the quantifiers themselves.

The mistake here is similar to a mistaken criticism of fixed-domain quantification in modal contexts, where quantifiers are assumed to range over possibilia (i.e., my italicized quantifiers above). On this interpretation, we get the following provable result:

$$\vdash \forall y \exists x(x = y),$$

which, if we assume that $\exists x(x = y)$ reports that y exists, says that everything exists necessarily—surely an untoward consequence!

The mistake being made here is to confuse what the formal sentence expresses with what the informal sentence 'everything exists necessarily' says. The formal sentence has a quantifier that ranges over possibilia, and so what it says, more precisely, is that every *possible* object is found in the domain of all possible objects, and that claim is certainly innocuous.[18]

Williamson goes on to make a further point. He says,

For *SVER* to be true in the actual world, every truth in the actual world must be known *in the actual world* at some possible time by some possible being. Since actual knowing happens at an actual time by an actual subject, the possibilist

[17] Williamson, *Knowledge and its Limits*, p. 288. I have altered the quoted passage by italicizing 'SVER' and 'WVER' to avoid confusion as to which principle Williamson is discussing.

[18] Compare James Garson, "Modal Logic", *The Stanford Encyclopedia of Philosophy (Winter 2001 Edition)*, Edward N. Zalta (ed.) <http://plato.stanford.edu/archives/win2001/entries/logic-modal/>.

liberalization of the quantifiers makes no substantive difference to what the actual truth of *SVER* requires. Thus *SVER* is still absurd . . .[19]

This reading of *SVER* is mistaken. The semantics for the consequent of the conditional send one out into the array of possible worlds right from the first quantifier—that is what is required by the fact that the quantifier ranges over *possibilia*. If we were to reject the indexical account of quantification, then the consequent of *SVER* could equally be expressed by:

$$\Diamond\exists s\exists tKspt.$$

This claim surely does not report what Williamson claims is true of the consequent of *SVER*, so the proper conclusion to draw is that the possibilist liberalization of the quantifiers makes every difference in the world.

Williamson's language is instructive here, I think, for notice that Williamson's gloss on the content of *SVER* adverts to the concept of actuality ("every truth in the actual world must be known *in the actual world*"), whereas *SVER* has no actuality operator in it. If this gloss were correct, the consequent of *SVER* should be:

$$\exists s\exists t@Kspt.$$

What this formula says is said, on the ordinary understanding of quantification in place of the indexical view, by:

$$\Diamond\exists s\exists t@Kspt.$$

The point to note, however, is that *SVER* has no actuality operator in it, and so Williamson's gloss cannot be correct.

This point is so obvious, it should lead us to wonder if there is some other point in the neighborhood of Williamson's objection that is worth exploring. If we attend to the language of actuality operators in the above quote, I think we can find such a point. Suppose we consider a (simplified) version of strong verificationism formulated using actuality operators:

$$SVER@\ @p \rightarrow @Kp,$$

to be contrasted with the standard simplified version of the view:

$$SVER\ p \rightarrow Kp.$$

This difference is interesting because there is a methodological error that can be, and has been, made in considering the knowability paradox.

[19] Williamson, *Knowledge and its Limits*, p. 289.

The mistake is to think the paradox can be avoided by substituting a different claim for the claim that all truths are knowable, and argue that the new claim doesn't entail *SVER* by the usual steps involved in the knowability paradox. One reason this approach is methodologically unsound is that the new claim may entail that all truths are knowable,[20] so this approach must be accompanied, at the very least, by a sound argument that the new claim doesn't entail the claim that all truths are knowable.[21]

A similar issue can arise with the formulation of the derived formula. The paradox is ordinarily presented to show that *SVER* follows from *WVER*. The possibilist liberalization of the quantifiers eliminates this concern, but eliminating that concern might leave other concerns. For example, suppose that *WVER* entails *SVER*@, and *SVER*@ cannot be argued to be equivalent to *WVER* by the innocuous logical principle of idempotence. Instead, *SVER*@ requires actual knowledge of actual truth, something much stronger than the possible knowledge recorded in *WVER*. Such a result would be every bit as paradoxical as the original derivation of Fitch's result, for *SVER*@ is obviously distinct logically from *WVER*. So even if Williamson's precise point cannot be sustained, this closely related point may still signal a concern that we must address.

The problem is only an apparent one, however. Once we move from simplified versions of *SVER* and *SVER*@ to one with the quantifiers explicit, we find that there are two different reads of the claim:

READING 1 $@p \rightarrow @\exists s \exists t Kspt$
READING 2 $@p \rightarrow \exists s \exists t @Kspt$

The first is provable in the same way that $p \rightarrow Kp$ is provable:

1. $@p \ \& \sim @Kp$	Assumption
2. $p \ \& \sim Kp$	actuality rule $p \dashv\vdash @p$
3. $\Diamond K(p \ \& \sim Kp)$	2, and the knowability claim
4. $\Diamond(Kp \ \& \ K\sim Kp)$	3, K-Dist
5. $\Diamond(Kp \ \& \sim Kp)$	4, KIT
6. $Kp \ \& \sim Kp$	5, by prior derivation
7. So, $@p \rightarrow @Kp$	6, RAA (1)

[20] I believe this mistake is made in C. Cozzo, "What We Can Learn from the Paradox of Knowability", *Topoi* 13, (1994), pp. 71–78.

[21] In fairness to Cozzo, he does give an argument why his new anti-realist commitment does not entail the knowability of all truth. His discussion of the point is vitiated, however, by the fallacy of confusing the necessity of the consequence with the necessity of the consequent. Since these issues are ancillary to our present discussion, I will not pursue the point here.

This proof relies at step 2 on the rule

$$p \dashv\vdash @p,$$

the standard rule for the actuality operator. Reliance on this rule is troubling, since, as we have already noted, any formula governed by the actuality operator is necessarily true if true at all. If the governed formula is contingent, then the rule in question is not necessarily truth-preserving, since there will be possible circumstances in which one side is true and the other false.

What justifies this rule, if anything does, is that it is *fixity-preserving*. Suppose we understand fixity ($\mathcal{F}$), following Edgington,[22] roughly so that '$\mathcal{F}$p' is true if and only if, whichever possible situation is designated as actual, 'p' is true. The above rule for the actuality operator is then fixity-preserving (because '$\mathcal{F}(p \leftrightarrow @p)$' is valid for such a logic),[23] though not necessarily truth-preserving.

It is interesting to note that the operator $\mathcal{F}$ is redundant for any formula not involving the operator @, and any fixity-preserving argument involving fixity operators but no actuality operator will also be necessarily truth-preserving if the related argument minus the fixity operators is necessarily truth-preserving. This fact might lead us to consider revising our notion of a formally adequate proof in terms of the concept of what is fixity-preserving rather than in terms of necessary truth-preservation, since the only time a difference will show up is when we are dealing with the logic of actuality.

More discussion is needed to make this change palatable, but we do not need to pursue that line of thought here in order to dispense with concern over Reading 1 of *SVER*@. If we attend to the modal elements involved in the domain of quantification, this reading involves no conflation of the concepts of knowable and known truth that is at the heart of the paradox of knowability. Reading 1 says that if 'p' is actually true, then it is also actually true that some possible being at some possible time knows that 'p' is true. In slogan form, such a claim says no more than that actual truth is actual knowable truth, a claim that fails to involve the core difficulty of the paradox concerning the loss of a distinction between universally known and universally knowable truth.

[22] Dorothy Edgington, "The Paradox of Knowability", pp. 567–568. Edgington borrows this notion from Martin Davies and Lloyd Humberstone, "Two Notions of Necessity", *Philosophical Studies* (1980).

[23] See Davies and Humberstone, "Two Notions of Necessity."

The second reading has more bite, however, for no matter what possible world we look to for the possibilia to instantiate the quantifiers in the consequent, the embedded actuality operator bumps us back to the actual world for knowledge by the possibilia in question of the truth of the proposition p. The problem for this reading is thus not that it is innocuous. Instead, the problem is that it doesn't follow from the premise that all truths are knowable. Crucial to the proof is a step that relies on the fact that knowledge implies truth, with a contradiction resulting from the application of that rule. In order for the proof to proceed in this way, we will, at some point, have to have Kp and K∼Kp. But if we make the standard assumption to prove *SVER*, we'll only get

$$\Diamond(K@p \,\&\, K \sim @Kp),$$

to which an application of *knowledge implies truth* yields

$$\Diamond(K@p \,\&\, \sim @Kp),$$

which is not a possible contradiction. Even if we derive ∼Kp from ∼@Kp, yielding

$$\Diamond(K@p \,\&\, \sim Kp)$$

we would need to derive Kp from K@p to give us the possible contradiction:

$$\Diamond(Kp \,\&\, \sim Kp).$$

This latter inference, involving an application of $p \dashv\vdash @p$ inside intensional contexts, leads to absurdity. To see the absurdity, one need only note that @p is necessarily true if true, and hence will be true even in worlds where p is false. So in worlds where p is false and yet K@p is true, the inference pattern in question would allow us to infer Kp even though p is false. Hence, a derivation of the a paradoxical result can only be obtained in this fashion by forcing a counterexample to the KIT rule that is used at an earlier stage in the proof.

There is also another issue here as well. In giving the above proofs, I have left unstated the implicit quantification necessary over the propositions in questions. So, on both readings, the claims are implicitly quantified universal formulas. If these quantifiers were explicitly included, an additional problem arises, for on the neo-Russellian view, we would have no guarantee that the value for 'p' in the antecedent of each reading is the same as the value for 'p' in the consequent. Without such a guarantee, the proof of any version of *SVER* from *WVER* is suspect.

Regarding Williamson's second objection, then, two claims are in order. First, his explicit comments fail to show that *WVER* and *SVER* are not logically equivalent in an innocuous way. Second, his gloss of the content of *SVER* raises a further question of whether *SVER*@ is derivable from *WVER*, a consequence that is, on the face of it, troubling. The appearance of a problem disappears, however, when we see that there are two readings of *SVER*@, one of which may follow from *WVER* but is not troubling and another which is troubling but does not follow. Williamson's second objection therefore gives no reason to abandon the neo-Russellian approach to the paradox.

This result is welcome since it forces attention back onto the neo-Russellian approach itself, to weaknesses it might have that would lead us to abandon this approach in dealing with the paradox. Williamson's first objection is precisely of this sort, for he claims that the neo-Russellian approach has a difficult time expressing ideas of contingency. He first notes, regarding the example 'Some dogs bark', that "Variation in the extensions of constituent terms of a proposition without variation in the proposition itself is just what is needed for the proposition to have its truth-value contingently."

This point by itself is not telling. One does not need a sentence to express the same proposition in every context or world in order for the proposition expressed to be contingent, otherwise an utterance or inscription of 'I am tired' would express a necessary truth. So we should expect more precision from Williamson on this point, and he attempts such, using as an example the claim that the F's are exactly a1 . . . an. He proposes to use restricted quantifiers $\exists_F$ and $\forall_F$ to represent this claim:

$\exists_F x1, \ldots, \exists_F xn(x1 = a1 \& \ldots \& xn = an) \& \forall_F y(y = a1 \vee \cdots y = an)$
(a1 . . . an are F and every F is one of them).

Let us call this formula 'Fexact'.

He then says, "If the . . . constituents actually expressed . . . were tied to a1, . . ., an in the way proposed," then the above sentence "should be a necessary truth," or more accurately, should express a necessary truth.[24] If this point is correct, then, Williamson claims that the sentence

$$\Diamond \sim \text{Fexact}$$

will express a false proposition, contrary to the idea that what is an F and what isn't should be able to be contingent.

[24] Williamson, *Knowledge and its Limits*, p. 288.

The first step of this argument requires showing that '◇∼Fexact' will always express a false proposition, on the neo-Russellian theory of quantifiers. I think Williamson is thinking as follows. For '◇∼Fexact' to come out true, there must be a world in which '∼Fexact' is true. The proposition expressed by this sentence is one containing the property of being F, a1 . . . an, and the part of the domain expressed by the quantifiers (which is exactly a1 . . . an). If so, however, the possible world in question will have to be one in which a1 . . . an exist and in which they are all the F's. So there is no world in which '∼Fexact' expresses a truth, hence '◇∼Fexact' is always and everywhere false.

This step of the argument employs restricted quantification, and one might think that the force of the argument can be escaped by rejecting restricted quantification. I do not plan to take this route, however, for I think there is no guarantee that the same issue cannot recur with unrestricted quantification. For example, it ought to be contingent how many things there are, and claims such as this raise precisely the same problem that Williamson discusses in the context of restricted quantification. So I will not pursue this line of response to the first step of the argument.

The second step of this argument is found in the claim that if '◇∼Fexact' expresses a false proposition, that is incompatible with the contingency of what is an F and what isn't. It is this step that I will argue is mistaken.

Williamson is correct that it ought to be able to turn out contingent what F's there are, but we need not accept the idea that for any sentence expressing a contingent truth, its contingency can be expressed by prefixing to it '◇∼', yielding a further sentence that is true. Compare standard Russellianism on the utterance 'I am here now.' It expresses a truth, and a contingent one, in my present context. Suppose we propose the following requirement: for utterance *u* to express a contingent truth *p*, there must be another utterance ∼*u* containing *u* and which is such that ∼*u* used in another context expresses the proposition ∼*p*. That requirement can't be satisfied, because the utterance 'I am here now', used in any other context, would express a different proposition, and 'I am not here now' would not express in any other context the denial of the proposition I express by saying 'I am here now' in the present context. Similarly, to express the contingency of what F's there are, there can be no requirement that we use 'Fexact' to express that claim.

Consider again the Russellian theory regarding 'I am here now.' Let us term such sentences *perspectival*, to honor the fact that the proposition

expressed can only be expressed by that sentence in a very unique context, the one I occupy in the vanishing present. Once time passes, as it just did, I can no longer express the former proposition using the same sentence. But I can still ascribe contingency to it. Perhaps I name the former proposition *p*, and assert that *p* is possibly false. Or perhaps its contingency can be expressed by employing a definite description, e.g., "the proposition expressed by my use of the sentence 'I am here now' at spatiotemporal location (s,t) might have been false."

One might object here that these new sentences fail to express the contingency of the proposition expressed by an utterance of 'I am here now.' In both cases, what is claimed to be contingent, it might be argued, has a cognitive significance not possessed by my original utterance of 'I am here now', and hence could not express the contingency of the very proposition expressed by that utterance. For I can believe that I am here now and not believe that to which contingency is ascribed on these accounts, and one can believe that to which contingency is ascribed without believing that I am here now.

This response raises the important and difficult issue of the nature of intensional attitudes and the proper way of individuating them. In order to answer it acceptably, we will need to take a brief tour through the topic of attitudes, assertions, and the cognitive significance that attaches to them. For convenience, we can proceed in terms of the concept of belief, for this discussion will easily generalize to other attitudes and to assertions as well. In what follows, I will not present anything close to a complete account of the matter, but will sketch reasons for conceiving of the topic in a certain way. The point of the discussion is to indicate the direction I think a proper account of the matter will go, and then to show that, conceived in this way, the problem raised by Williamson can be answered.

PROPOSITIONAL CONTENT AND COGNITIVE SIGNIFICANCE

When a person has a belief, it may take one of three forms. First, the person may believe (*de dicto*) that a certain proposition is true. Second, a person believe (*de re*) of an object that it has a certain property. Third, a person may believe (*de se*) of himself that he, himself has a certain characteristic.

Various theories are generated by taking one type of belief as fundamental and attempting to explain the others in terms of it.

A propositional theory takes *de dicto* belief to be the primary form of belief and attempts to explain *de re* and *de se* belief as special cases of *de dicto* belief.[25] A first person theory takes *de se* belief as primary, and clarifies the others in terms of it.[26] Finally, a property theory takes *de re* belief to be fundamental and explains the others in terms of it.[27]

In order to address Williamson's objection, I will do two things. First, I will suggest why I find a propositional theory preferable to the alternatives. In doing so, I will suggest the kinds of arguments I find persuasive, though I will not have the space to develop the arguments in detail here. The point of the discussion will be, rather, to make plausible preferring a propositional approach, even if the grounds cited fall quite a bit short of being compelling or convincing. To do the latter would require a book-length treatment of its own, and such an extensive detour would be out of place here. After motivating attention on a propositional approach, I will distinguish two types of such an approach, a dyadic and a triadic version. I will argue that on either of these approaches to a propositional theory, there are suitable answers available to Williamson's objection concerning how, on the neo-Russellian approach to quantification, to express the contingency of certain claims. What follows will therefore be quite cursory and brief, with indications in the notes as to where fuller treatments of these issues can be found.

There are good reasons for thinking that neither of the second two theories can be as successful as a propositional theory of intensional attitudes. The property theory fails both in its explanation of *de dicto* and *de se* belief. Its explanation of *de dicto* belief must isolate some object of which some property is predicated when a person believes, for example, that Bush is President. No such explanation can be adequate, for a person can believe this proposition even if the believer is the only concrete object that exists. If the believer is the only thing there is, there are only two options open to the property theory. In believing that Bush is President, the person must either be predicating a property of himself or of some abstract or mathematical object. Such a person cannot be predicating a property of himself, for the content of the resulting belief could not be true unless the believer existed; yet, clearly, Bush could be President even if the believer did not exist.

[25] See, e.g., Roderick Chisholm, *Person and Object: A Metaphysical Study* (London, 1976). [26] See, e.g., Roderick Chisholm, *The First Person* (Minneapolis, 1981).

[27] See, e.g., S. E. Boer and William G. Lycan, "Who, Me?" *The Philosophical Review* 89 (1980), pp. 427–466; and B. L. Davidson, "Belief *De Re* and *De Se*", *The Australasian Journal of Philosophy* 63 (1985), pp. 389–406.

The second option, on which the believer predicates a property of some abstract object is equally problematic. This option adds more to the content of the belief than is actually there. When normal English speakers believe what is expressed by 'Bush is President', their thoughts involve no abstract object other than the property of being President. Hence, any attempt to make some further abstract object the object of a predicated property mistakenly posits more to the content of a *de dicto* belief than is actually there.

The property theory has equally severe problems with explaining *de se* belief. There is a difference between believing of a person who happens to be oneself that his pants are on fire; it is quite another thing to believe *de se* of oneself that one's own pants are on fire.

A property theorist might attempt to handle this difficulty by claiming that there is a special description which isolates *de se* from mere *de re* belief. For example, one might insist that one pick out the individual in question by the description 'person identical with me.' This, of course, will not do; not all persons who hold *de se* beliefs are speakers of English. Altering the requirements on the description so that the description is translationally equivalent with the description 'person identical with me' will not work either. The content of a *de se* belief that one's pants are on fire does not imply that there is any such thing as translational equivalence.

Alternatively, a property theorist may prefer to talk of properties rather than descriptions. Presumably, the property in question for *de se* belief would be the relation of identity: to believe *de se* that one's pants are on fire is to believe of oneself, under the relation of identity, that one's pants are on fire. In order for such a property to do the explanatory work required, it must be impossible for a person to believe something of himself under that property without having a *de se* belief. It is not obvious that there is any such property. Suppose Hume believes 'The author of that book is smart', when he is the author in question, though unaware of that fact. In such a case, it is initially plausible to suppose that Hume attributes to himself, under the relation of identity, the property of being smart; yet he does not believe *de se* that he is smart. Hence, the proposed relation of identity must be a very special sort of relation in order to do the explanatory work required of it. If it is, though, the property theory succeeds only by definitional fiat. If the property in question is special in the required way, then there are two fundamentally different kinds of *de re* belief: one which has a public content, in that it can be grasped by any believer whatsoever;

one which is private, in the sense that it includes a property graspable only by one person at most.

The first person theory fares no better than the property theory. On it, whenever a person has a belief, he is directly attributing a property to himself. This view gives an inadequate construal of *de dicto* belief. On the first person theory, when I believe that a certain horse is a sorrel, I am believing something that entails my own existence, for I am directly attributing some property to myself. This implication is mistaken; the content of my belief about this horse could be true even if I did not exist.

These considerations should lead us to reject both the property and first person theories and focus instead on developing a propositional theory. Such a theory is most relevant to the charge involved in Williamson's example, so we can make progress toward addressing that charge by considering how a propositional theory might be developed.

The Propositional Theory

The propositional theory is not without its own problems, however, but the special issue most relevant to the topic of cognitive significance is the propositional account of *de se* belief. Such beliefs present the same kind of threat posed above for the neo-Russellian view of quantification, especially once one notices that I can believe that I am tired without believing that J.K. is tired (amnesia might rob me of the latter belief, but it won't rob me of the former). How can a propositional theory get around this difficulty?

Let us suppose, then, that a propositional theory takes intensional attitudes to be a dyadic relation between a particular type of mental state such as belief and a propositional content. Such a dyadic propositional theory, in attempting to reduce *de se* to *de dicto* belief, may be either a single proposition theory or a multiple proposition theory.[28] A single proposition theory holds that, when Hume and Heimsohn each believe what is expressed for them by the sentence 'I am tired,' they are believing the same, single proposition. The multiple proposition theory holds that Hume and Heimsohn are each believing different propositions: Hume is believing one about Hume, and Heimsohn one about Heimsohn.

The single proposition theory is not the most attractive way to develop the propositional theory. One difficulty it faces concerns first

[28] For a characteristically illuminating discussion of these two approaches, see Ernest Sosa, "Consciousness of the Self and the Present", in J. E. Tomerlin (ed.), *Agent, Language and the Structure of the World* (Indianapolis, 1983), pp. 131–143.

person propositions such as *I am tired*. According to the single proposition theory, when Hume believes that proposition, its truth or falsity is to be evaluated at a certain perspective: the ordered pair constituted by Hume and the present moment. Not only is the truth of the proposition to be evaluated at this perspective, in order to make sense of Hume's belief, Hume himself must believe the proposition in question to be true of himself and the time in question. If he does not believe the proposition to be true of himself and the time in question, there is nothing in the account of the belief in question which isolates any person or time as the object of belief, for by hypothesis, the proposition is not about any particular person or time since it can be believed by any person at any time. Hence, in believing *I am tired*, Hume is really believing that this proposition is true of himself and the present moment; *de se* belief is thus a four-place relation on the single proposition theory between a person, a proposition, the person of whom the proposition is true, and the time in question.

Furthermore, the single proposition theory cannot construe all cases of ordinary *de dicto* belief as involving a four-place relation between a believer, proposition, object, and time. Consider believing that the tallest spy is a spy. This proposition can be believed even if there is no tallest spy, hence the object in question (the third relata above) cannot be any spy. Further, the object cannot be the believer himself, for the proposition *the tallest spy is a spy* does not entail the existence of the believer.

So some *de dicto* beliefs do not involve the four-place relation constitutive of *de se* belief on the single proposition theory. This fact counts against the single proposition theory, for we no longer have a unified account of belief on it. According to the single proposition theory, *de se* belief is a four-place relation; ordinary *de dicto* belief is not.

The multiple proposition theory, where belief is construed as a two-place relation between a believer and a proposition, suffers from the kind of problem Williamson raises against the neo-Russellian theory of quantification. The feature in question concerns the commitment of the multiple proposition theory to private propositions: propositions that can be accessed at most by one person. When Hume believes what is expressed by 'I am tired', he believes a proposition which no one else can believe (or take any other intensional attitude towards). The reason that the proposition expressed is private is that it motivates behavior in a way that belief in no other proposition motivates. For example, if Hume believes what is expressed by 'Hume is tired', that belief may not motivate him to lie down, for he may not realize that he is Hume. Furthermore, when Hume believes *de se* that he is tired, were that

proposition capable of being believed by Heimson as well, we would have no explanation in terms of belief why Hume and not Heimson stands up when a tired volunteer is asked to stand.

Several multiple proposition theorists have readily accepted the private proposition implication of their theory, and have held that it is an important corollary of their view.[29] It is instructive in our own context to note the implications of this attitude for Williamson's objection that the contingency of certain claims cannot be adequately expressed on the neo-Russellian account of quantification. If private propositions are unproblematic, it should equally be unproblematic for the neo-Russellian view of quantification to claim that there are propositions that are contingently true, but where the contingency in question cannot be expressed in the ordinary way by prefacing the sentence that expresses that proposition with the logical device 'it is possibly not true that'. When we talk of Hume's *de se* belief that he is tired, we have not expressed the proposition that is the content of Hume's belief; instead, we have only pointed to it. Just so, on the neo-Russellian view of quantification, the proposition expressed by a quantified sentence may be only contingently true even though the claim that this very proposition is contingent is one that cannot be expressed but can only be pointed to.

So one response to Williamson's complaint is to maintain that it imposes unrealistic demands on the theory of expression. The unrealistic demand is that, for every truth, there must be a linguistic vehicle that expresses it. In the case of *de se* beliefs, there is the perspective of the *de se* believer him/herself from which the perspectival proposition can be expressed, but this fact seems contingent at best. If we think of zombies, for whom self-consciousness is simply non-existent and perhaps not possible, the perspectival propositions distinctive of *de se* attitudes would be inexpressible (since only the zombie could express them). But even if inexpressible, the propositions would still exist on the multiple propositions version of a dyadic propositional approach. Furthermore, if this response is acceptable here, then there is no reason to balk if the neo-Russellian theory of quantification implies that certain truths cannot be expressed, either.

We will return to this response to Williamson's complaint below, but I want first to point out that placing too much weight on appeals to private propositions in the discussion of *de se* attitudes would be unwise. First, there is linguistic data that count against the view. When my

[29] See, e.g., Chisholm, *Person and Object*. It is interesting to note that only five years later, Chisholm had abandoned this view.

wife cheerily says, "I went to the dentist today", I may correctly, but grumpily, respond, "I know that." What do I know? The most obvious answer is that the content of my knowledge is the same as the content of her assertion, but on privacy views, that account is mistaken.

The multiple proposition theory must hold that, in cases like this, what is known is not the very same proposition which was asserted, but some other proposition very similar to the asserted proposition. It is difficult, though, to muster motivation for this response, other than the fact that the multiple propositions theory demands it.

A further problem is that if we consider what a proposition might be, all such views leave it mysterious and inexplicable how there could be such things as private propositions. A proposition might be quite subjective: it might be the cognitive content of an asserted sentence, or it might be the sort of thing that captures a person's subjective conception of a particular situation. No proposition on this construal will be private, for any subjective feature Hume might exemplify in believing *de se* that he is tired can be matched by Heimsohn in believing *de re* of Hume that he is tired. My point here is reminiscent of Hume's claim that when he looked inside of himself, he never found himself; he only found a succession of ideas. The relevant analogue here is that there is no feature of a person's subjective conception of himself that is necessarily unique: anyway each of us thinks of himself is a way that others seemingly can duplicate.

In a different vein, a proposition might be conceived to be something quite objective: the bearer of a truth-value, the sort of thing that corresponds to a fact, or perhaps the sort of thing that is true or false when asserted. Such a thing could only be private if there were private objects or facts as well: objects or facts with which only one person at most could be acquainted. Perhaps there are inner objects with such a property, though such a claim has been highly controversial in philosophy since the late Wittgenstein. Such a defense of the idea of privacy is not sufficient, however, for the kinds of private propositions that are required include private propositions about public objects such as Hume and his being tired. If propositions are objective, it is hard to see any justification from the nature of propositions for thinking that propositions about public objects can be private.

So it is hard to place much stock in a propositional theory that embraces the existence of private propositions. As a result, it is difficult to use this approach to motivate a response to Williamson's example that gladly accepts the consequence he notes, that there are contingent truths

whose contingency cannot be expressed (or cannot be expressed in the ordinary way). Perhaps, though, there is a different propositional approach that will provide a basis for a reply to Williamson's complaint.

An Alternative Multiple Proposition Theory

The versions of a propositional approach rejected above are all dyadic theories: they attempt to explain all belief in terms of a two-place relation between a believer and a proposition. An alternative is to accept a theory on which the propositional content of belief is not intended to account for the entire cognitive significance of the belief.[30] Such a theory is a triadic theory of belief, on which the terms of the belief relation are the believer, an objective proposition, and perhaps a subjectively conceived way of accessing the proposition in question.[31] The point of the third element is to explain how the belief with the content it has can play the causal role that it in fact plays.

On this theory, we assign a propositional object to belief in order to explain how some beliefs are true and others are false. As such, the propositional object of belief is objective, corresponding (or failing to correspond) to concrete states, events, and so on. We assign a means of access to the proposition to the belief relation in order to explain the role that a person's subjective conception of the situation plays in belief. Presumably, each proposition will have a distinctive range of subjective conceptions associated with it, and each subjective conception will have a distinctive range of propositions associated with it. It will not be the case that any means of access can take one to any proposition.

Are any propositions private on this triadic multiple proposition theory? No, for when Hume believes *de se* that he is tired, he believes the same proposition Heimsohn believes when Heimsohn believes *de re* of Hume that Hume is tired. Each of them accesses the proposition in a different manner in virtue of a different subjective conception of the proposition in question.

I do not maintain that such a triadic theory is trouble-free, but only wish to note its capacity for handling some problems raised by dyadic

[30] An excellent source regarding the need for a triadic theory is Nathan Salmon's *Frege's Puzzle* (Cambridge, Mass., 1986). See also John Perry, "The Problem of the Essential Indexical", *Noûs* 13 (1979), pp. 3–21.

[31] An alternative way of developing this viewpoint distinguishes two types of content for a belief, a wide content and a narrow content. One important view along these lines is developed by David Chalmers in "The Components of Content", *Philosophy of Mind: Classical and Contemporary Readings* (Oxford, 2002).

theories. In our context, we are interested in seeing what a discussion of cognitive significance can teach us about the power of Williamson's complaint that on the neo-Russellian view of quantification, the contingency of certain claims cannot be readily expressed. In reply, I suggested ways in which the contingency in question could be expressed, though not using the same language of quantification used to express the contingent truth itself. The question is whether the suggested ways of expressing the contingency in question are adequate—whether they really express the contingency of the precise proposition under discussion, and the challenge raised on behalf of Williamson's complaint was that cognitive significance was not preserved, and hence the devices I used for expressing contingency could not be adequate.

The point of the above discussion is to reveal some of the problems with the analogous issue regarding the contents of belief, and to show how Williamson's complaint faces an important dilemma in attempting to find a problem for the present view of quantification. One can think of the issue either in terms of a dyadic or triadic theory of intensional attitudes and assertions, and if we assume a dyadic theory, we will have to endorse the existence of *perspectival* propositions,[32] propositions that are limited in terms of accessibility to particular situations. The other option is to adopt a triadic theory.[33]

If one adopts a dyadic theory, one will have to rest content with the existence of perspectival propositions, and then Williamson's objection will disappear. For if there are perspectival propositions such as that expressed by 'I am here now', the proposition doesn't cease to be contingent simply because a change in context robs this sentence of the power to express the proposition it actually expresses and hence robs us of the capacity to express the contingency of this proposition in any other context by uttering 'it is possible that I am not here now.' It is true that the contingency of the proposition expressed by uttering 'I am here now' can be expressed by uttering in the same context 'it is possible that I am not here now', but that does not affect the point on behalf of the neo-Russellian view of quantification. The phrase 'it is possible that' forces a change of context when it prefaces a quantified sentence on the neo-Russellian view, but not otherwise, and the point of the analogy between the example involved in Williamson's complaint and the example

[32] Roderick Chisholm defended such a view in *Person and Object: A Metaphysical Study* (London, 1976).

[33] For a full discussion of these issues, see Nathan Salmon, *Frege's Puzzle* (Cambridge, Mass., 1986).

involving *de se* belief is that context shifts sometimes forces a change in cognitive significance. To point out that there are other ways in which the two examples are disanalogous would not restore the force of Williamson's complaint unless accompanied by an argument that every theory of expression must allow the phrase 'it is possible that' to preface any sentence without engendering a context shift. Such an argument would, of course, simply be a refutation of the neo-Russellian view of quantification, and hence would render Williamson's complaint irrelevant.

The point to note, then, is that in order to press Williamson's objection, one will have to insist that any theory that implies the existence of perspectival propositions is problematic. In that case, however, one will have to adopt a triadic theory, one result of which is that there are multiple ways to express the same propositional content. For example, the same propositional content can be expressed by Hume who says, "I am tired", and Heimsohn who says, "Hume is tired." The difference between the two assertions has to do with the subjective mode of access to the propositional information in question.

If we apply this analogy to the quantificational theory in question, the original response I gave to Williamson's complaint must be granted to be adequate. What I claimed is that even though the contingency of the proposition in question could not be expressed by prefacing it with '$\Diamond\sim$', the contingency could be expressed nonetheless by naming the proposition and predicating contingency of it via that name. The reply on behalf of Williamson's complaint was that such a maneuver does not preserve aspects of cognitive significance that an adequate theory of propositions must preserve. On a triadic theory, however, such a response confuses the second relatum of the belief relation with the third. The second relatum is the proposition in question, and the third relatum is some item whose explanatory role is to account for cognitive significance. Once we get clear on the nature of a triadic theory, it becomes equally clear how easy it is to answer Williamson's complaint. We can pick out the proposition in question by a description: the proposition that would be expressed in a world with such-and-such a domain by the use of sentence 's', having the same meaning that 's' has in our language (note that such a description requires a distinction between proposition expressed and meaning, but that distinction is already secured by the fact that 'I am tired' expresses a different proposition when two different people utter it). The description in question can be used to establish reference for a name—call it 'p'—and then the

contingency of the quantified claim can be expressed by asserting 'p might have been false.'[34]

The conclusion to draw, then, is that Williamson's discussion does not undermine the neo-Russellian theory. The most serious concern he raises is this last one about contingency, but pressing the objection requires placing more burden on the connection between the theory of propositions and the theory of cognitive significance than any adequate account can bear.

The fact that known objections to a view fail to undermine it provides little assurance, however, that the view is plausible. Hence, even with these responses in hand, the question remains whether the neo-Russellian view of quantification is defensible, and because this question is still an open one, it remains an open question whether the proof of contradiction central to the knowability paradox is valid.

A DEFENSE OF THE VIEW

To this point, I have argued for two claims. First, I have argued that the neo-Russellian view of quantification blocks the proofs from knowable truth to known truth (and from unknown truth to unknowable truth). Second, I have argued that the objections that have been raised in the literature to this approach are not damaging. These points leave unaddressed the question of why we should believe that this approach to the paradox is correct. To put the point more forcefully, we might ask whether there are any independent grounds for thinking that this account of quantification is the correct one, or whether this approach is simply another *ad hoc* response to the paradox.

A first point to note is that there is nothing within the subject matter of first-order logical theory that would call for a choice between the Fregean view of quantification (according to which quantifiers express properties of a certain sort) and this neo-Russellian view. In this respect, such logical theory is insulated from the philosophical issue of the nature of the proposition or thought expressed by a quantified sentence. So it would be a mistake to think of the present proposal as involving a substantive revision of the logical tradition derived from Frege. Instead,

[34] For more detailed discussion of these issues, see my *The Possibility of an All-Knowing God*, (London, 1986), especially ch. 2; and "The Haecceity Theory and Perspectival Limitation", *Australasian Journal of Philosophy* 67.3 (September 1989), pp. 295–305.

it is only a proposed revision about the relation of expression between formal sentences and the propositions or thoughts that they express.

Where the revision has some teeth to it is in the modal domain, and to see what kind of an answer can be given to the question of what justifies such a revised conception of quantification, it is useful to consider how features of quantified modal logic and the semantics for it are usually justified. Suppose we begin, then, with a very simple logic[35] consisting of the apparatus of first-order theory together with a necessity operator, an additional clause to the standard account of what counts as a formula, to the effect that if φ is a formula, then so it □φ, and a definition of the possibility operator ◇ as the dual of the necessity operator (i.e., ◇φ = df. ∼□∼φ). The simplest semantics for this language specifies a class of interpretations with two mutually exclusive non-empty domains (a domain of possible worlds and a domain of individuals), where each interpretation specifies for each *n*-place predicate '*P*' whether '*P*' applies to an *n*-numbered sequence of objects at a world or not. Given this apparatus, the semantics defines truth conditions for formulas of the language by clarifying recursively what it is for a formula to be true under a given interpretation and assignment function at a world. We first define a denotation function d relative to I and f so that this function on a term is whatever individual in the domain the interpretation assigns to that term, and for variable x, the denotation function relative to I and f is just f(x). With this apparatus, we can give the clauses for truth for open formulas and for the clauses of interest here, namely, those for universally quantified formulas and for modal formulas:

The open, atomic formula 'Fx' is true at w relative to I and f just in case I specifies that 'F' applies to whatever d(x) (relative to I and F) is at w.

A quantified formula '∀xFx' is true at w relative to interpretation I and assignment function f just in case I specifies that for all individuals a, the open formula 'Fx' is true under I and f[x,a], where f[x,a] is f if f(x) = a, and is just like f otherwise except that it assigns a to x instead of f(x).

The open modal formula '□Fx' is true relative to I and f at w just in case I specifies that for every possible world w*, 'Fx' is true relative to I and f at w*.

We can then define what it is for a formula to be true at an interpretation (for every assignment function f, the formula is true relative to I and f at the actual world), and what it is for a formula to be logically true (it is true relative to I for all interpretations I).

[35] Here I borrow from Christopher Menzel "Actualism", *The Stanford Encyclopedia of Philosophy (Spring 2003 Edition)*, Edward N. Zalta (ed.) <http://plato.stanford.edu/archives/spr2003/entries/actualism/>.

To complete this logic, we add to the the rules of classical first-order logic the following axioms plus the rule of necessitation:

$$\text{K axiom: } \Box(\varphi \rightarrow \phi) \rightarrow (\Box\varphi \rightarrow \Box\phi)$$
$$\text{T axiom: } \Box\varphi \rightarrow \varphi$$
$$\text{5 axiom : } \Diamond\varphi \rightarrow \Box\Diamond\varphi.$$

This system is problematic to actualists, those philosophers who maintain that everything that exists is actual, on a number of grounds. One example will suffice for present purposes. Suppose we understand the claim that something is contingent as $\Diamond\sim\exists y(y = x)$; then the above logic can be used to show that it is impossible for something to be contingent, i.e., $\sim\Diamond\exists x\Diamond\sim\exists y(y = x)$, since it is a theorem of this logic that $\Box\forall x\Box\exists y(y = x)$ (if we understand existence in terms of '$\exists y(y = x)$', then this theorem disturbingly says that it is necessary that everything exists necessarily).

Kripke's 1963 semantics for quantified modal logic appeals to those who view such results as problematic.[36] The trick is to replace in the above system the single domain of individuals with a function that assigns to each world its own domain of individuals. Domains are thereby permitted to vary from world to world, blocking the problematic results above as well as other problematic results.

The important question for us in our present context is to ask about the justification of such a change from the simple system above. The justification involves two ideas. First, the change blocks the problematic results noted above. Such blocking could be achieved in other, more problematic ways: we could, for example, restrict the rules so that they couldn't be used in the derivation of the precise results we view as problematic. Such a change would surely be *ad hoc*. So what distinguishes the above blocking from obviously *ad hoc* ones? This question brings us to the second idea in the justification of the change, for the change developed in the Kripke semantics presents a coherent picture in which quantified formulas in modal contexts are evaluated with respect to the objects that exist in the modal context in question, not with respect to objects that exist in other domains (as in the above simple quantified modal logic).

This approach provides a model for how to justify the neo-Russellian view of quantification. First, the change blocks some problematic results,

[36] Saul Kripke, "Semantical Considerations on Modal Logic", *Acta Philosophica Fennica* 16 (1963), pp. 83–94.

especially it blocks the deeply perplexing loss of a logical distinction between actuality and possibility that is the heart of the paradox we are investigating. Further, it does so without being *ad hoc*. When we consider a truth and ask whether it can be known, the first thing that happens semantically is that we are taken to some possible world (distinct from the actual world if the claim in question is not known). If we think about what needs to be known in that possible situation, the coherent and plausible answer is that what must be known is this: it is something about the actual world that must be known. It is this insight that prompts the details of Edgington's restriction strategy discussed earlier, and even if the strategy she follows proves to be unsuccessful, the motivation is nonetheless pristine. To make the point clear, suppose we have two worlds, one actual and one possible. In the first world are objects a, b, and c; in the second world are objects d, e, and f. All of these objects have the property P, but in the first world, no one knows that everything is P. In the second world, someone does know that everything is P. Does this detail show that what is unknown in the first world is in fact knowable? Edgington's insight is that it does not, because what is known in the second world doesn't appropriately characterize what is unknown in the first world. The neo-Russellian picture of quantification explains precisely the failure. The problem is that, though the sentences used in the two modal contexts are the same, they do not express the same proposition.

It is important to note the analogies between this defense of the neo-Russellian picture of quantification and the actualist defense of Kripke semantics. In both cases, what prompts the revision of a simple semantics is untoward consequences. In both cases, an *ad hoc* approach is jettisoned, according to which we restrict the rules of inference to block the unwanted consequences. Finally, in both cases, there is an intuitive picture in terms of possible worlds to appeal to in presenting and justifying the particular revision suggested. Since the acceptability of Kripke semantics over the simple semantics first outlined above is clear and persuasive, I submit that the same qualities inhere in the present defense of the neo-Russellian picture of quantification.

CONCLUSION

Thus, the neo-Russellian view of quantification has the same intuitive appeal that we find in the alteration of the simple semantics above, which led to problematic results. It is justified, therefore, in the same way

as that alteration. The end product is a view with the happy result that the logical distinction between known and knowable truth is preserved.

The view contains a surprise, however, for it provides little comfort to those who wish a solution to the paradox to provide a defense of the view that all truths are knowable. To return to the simple example above, there is no guarantee that there is another world containing only a, b, and c in which someone knows that all of these are P, nor that there is some other world with a different domain of objects in which the proposition in question is expressible or in which a surrogate for that proposition is expressible and knowledge of which would be sufficient for showing that the proposition in question is knowable. These are all details left to anti-realists to solve, leaving the somewhat surprising result that the knowability paradox can be solved but perhaps not in a way that can rescue anti-realism from its clutches. Perhaps the headline should read: "Freed from Paradox, Anti-Realists Still Tremble."

7

Conclusion

The direction of our discussion has shown a progressive narrowing of options for solving the knowability paradox. A standard first impression when confronting the paradox is to think that it involves some type of logical trick of the sort involved in sophistical proofs such as the proof by induction that all horses are the same color.[1] Such is not the case, however.

Those most motivated to find a way out of the paradox are those most committed to the first assumption of the paradox, the assumption that all truths are knowable. Once convinced that no logical trick is involved in a proof of Fitch's result, the reaction of those attracted to the knowability claim will be some motivation for questioning the wisdom of full commitment to this universal generalization. The result is then to adopt some restriction strategy, denying that all truths are knowable, substituting instead a claim of the form that all truths of a certain type are knowable. The restricted claim needs only an explanation of why it is all an anti-realist ever needed to endorse in order to maintain suitable philosophical distance from realism. Once such an explanation is developed, such theorists will experience the wonderful phenomenon of a sated philosophical appetite, and can turn to other philosophical issues and projects.

Such a motivation and response, as well as the strategy that results, may be perfectly rational, but it fails to appreciate the fundamental

[1] The proof proceeds on the size of sets, with sets having one horse in them as the base case (we assume for simplicity that each horse is uniform in color, i.e., no paints or appaloosas allowed), so the base case is trivial. Then show that if all sets of size n have only horses with the same color in them, then so do sets of size $n + 1$ (the argument for it being that if the $n + 1$ set has different colored horses in it, you could drop a same-colored horse from that set, and violated the assumption for a conditional proof that all sets of size n have only same-colored horses in them, e.g., if a set of 6 horses has 5 of the same color and one of a different color, then there is a set of 5 horses having 4 of the same color and one of a different color).

paradoxicality involved in a derivation of Fitch's result. The paradox, at bottom, concerns a lost logical distinction between possibility and actuality. The usual approach to the paradox treats the lost distinction in terms of collapsing the distinction between known and knowable truth, deriving this collapse from the truth of global anti-realism. There is something right about this approach and something wrong. What is correct about it is that the fundamental paradoxicality involves a lost logical distinction between possibility and actuality. What is wrong is that it takes an affirmation of global anti-realism to generate the loss. Global anti-realism entails that all and only knowable truths are known, if Fitch's proof is sound, but a related point requires no assumptions whatsoever. If Fitch's proof is sound, there is a lost logical distinction between possible universal knowledge and actual universal knowledge, for Fitch's proof purports to establish the harder direction of demonstrating that the proposition *All truths are knowable* is logically equivalent to the claim that *All truths are known*. The usual construal of the paradox in terms of an implication of global anti-realism is thus only one instance of the perplexity occasioned by a denial of the general distinction between actuality and possibility. Because the same paradoxicality can be generated using Fitch's proof with results that do not require the truth of global anti-realism or any other type of anti-realism, no pursuit of a restriction strategy on the knowability assumption will effectively address the paradox. A proper appreciation of the depth of the paradox shows that restriction strategies are red herrings.

Once one gives up the idea that the proof is sophistical and sees the irrelevance of approaches that restrict the knowability assumption, the next step is to wonder whether there is something mistaken about the special rules being used for the apparatus in the proof beyond first-order theory. In particular, one might be suspicious of the rules for the knowledge operator, the rules that encode the ideas that knowledge implies truth and that knowing a conjunction constitutes knowledge of each of the conjuncts. Disputes about these rules take us beyond the suspicion of mere sophistry, but as we saw in Chapter 4, a solution to the paradox cannot be found by questioning the special rules for the knowledge operator.

From the original suspicion of sophistry, we thus come ineluctably to the suspicion that there is something wrong with classical first-order logic. This progression ought to be very surprising—one day we are thinking the proof is simply sophistical, and the next we are trying to figure out how to revise our logic. Besides being surprising, the revision

strategy does not look promising. If the revision is along intuitionistic lines, the lost logical distinction between known and knowable truth is replaced with a lost distinction between unknown and unknowable truth. No comfort will be found in this direction, however: all that results is that we substitue one account of a lost logical distinction between actuality and possibility for a different, or slightly reworded, one. In response to this last challenge, it is tempting to argue that our sense that there is such a distinction depends on reading the key logical formulas differently from the way intuitionists understand those formulas. This response raises a further problem, that of finding a way to keep the discussion from making philosophy of logic impossible. In the end, something very close to the Quinean claim that those defending alternative logics are merely changing the subject turned out to be true of the attempt to evade the paradox by replacing classical logic with something weaker. In the end, the change of logic does nothing whatsoever to ease the intellectual discomfort engendered by Fitch's result.

Once all of these avenues of escape are shut off, it becomes very difficult to find any hope of escaping the paradox, but we saw in the last chapter some indication of hope. The ray of light results from treating the paradox as a special instance of the fallacies involved in substituting into intensional contexts. The special case has to do with substituting quantified formulas into such contexts, for if quantifiers are modally indexical, then the substitution is illicit.

Perhaps the most pressing issue raised by our investigation of the knowability paradox concerns the solution to it presented in the last chapter. Even if it is granted that the solution succeeds in avoiding the paradox, why should we think it is correct to think of quantifiers as modally indexical? Isn't this just another *ad hoc* maneuver to save anti-realism?

This way of looking at the solution is inapt. As noted earlier, the phenomenon of substitutions into intensional contexts is laden with difficulty, and many of the standard substitutivity assumptions have to be jettisoned in order to avoid problematic inferences. The result is that restrictions on substitutions into intensional contexts can be justified in virtue of the problematic inferences they prohibit, even in the absence of some more general theoretical basis of the restriction. That is precisely the way the present view of quantifiers was defended in the last chapter. It prevents the problematic inference according to which there is no distinction between known truth and knowable truth.

But do we need anything more here than an insistence that substituting into an intensional context is invalid? That is, why not simply

point out the problems with such substitution and then declare the argument invalid, without adding the idea of the modal indexicality of quantification? The answer is that it is simply not enough to point out that there are problems with substituting into intensional contexts, and an adequate defense of the invalidity of Fitch's proof requires explaining exactly how the substitution is mistaken.

Consider, for example, the following argument: 6 is necessarily less than 9, and since 3 plus 3 just is 6, 3 plus 3 is necessarily less than 9. This argument substitutes into an intensional context, but shouldn't be rejected on that basis, since it is a valid argument. So some explanation of failure of substitutivity is required in order for that approach to the paradox to be adequate, and that is what the idea of the modal indexicality of quantification provides.

The difference between this proposal and, e.g., Tennant's proposal, which I have claimed is *ad hoc*, is that his restrictions are supposed to be derived from a general theory of truth, and they are not. When constructing a logic, one primary measure we make of a proposal is in terms of whether it gets the inferences right. If the logical apparatus certifies as valid obviously problematic inferences, then there is an argument against that proposal. Some ways of fixing the problem are obviously *ad hoc*. For example, if the remedy fixes the problem introducing an exception clause that applies only to that problem, then it is *ad hoc*. The present theory is, however, fully general even though it was developed in response to Fitch's result and I know of no troubling inference it prohibits other than those involved in the knowability paradox. These facts reveal the motivation for the theory; they do not show that it is *ad hoc*.

What else is needed besides the fully general nature of the theory to show that it is not *ad hoc*? I do not know what the minimal conditions on a solution are, but when the solution presents a coherent and fully general picture differing from the problematic one, we have a sufficient justification to reject the charge that the new proposal is *ad hoc* unless and until defeating information is presented to undermine this *prima facie* warrant.

I also argued that the development of the present approach is very much like the development of Kripke's semantics for quantified modal logic as a way to avoid a semantics that apparently implies that everything exists necessarily. By proposing that the truth conditions for quantified formulas be interpreted relative to the varying domains of modal contexts, that proposal fits well with our intuitive idea that quantified formulas should answer to the domain of the context in

question, not some larger domain that includes things that exist only in other possible worlds, and the problematic implication about necessary existence is avoided. The neo-Russellian view of quantification has a similar feature, in that it proposes that modal contexts introduce the need to tie the domain more closely to the thought expressed by a quantified sentence in order for what is known in one world to count as showing that a truth in another world is knowable.

In this way, another comparison is instructive. The neo-Russellian view of quantification compares to the standard, Fregean picture of quantification in modal contexts in the same way the multiple proposition theory compares to the single proposition theory regarding sentences used to express *de se* attitudes. The single proposition theory holds that 'I am here now' expresses the same proposition in every context of use, just as the standard, Fregean picture of quantification holds that 'Everything is F' expresses the same proposition in every modal context of use. The multiple propositions theory holds instead that the theory of expression is involved in context shifts, so that 'I am here now' expresses a different proposition in different contexts of utterance. Just so, the neo-Russellian view of quantification holds that the theory of expression is expected to handle the semantic import involved in modal context shifts, so that 'Everything is F' expresses, or may express, a different proposition in different modal contexts.

The analogy with single and multiple proposition theories is instructive in another way. The multiple proposition theorist can complain about the explanation of action offered by a single proposition theorist. When Hume and Heimson both believe what is expressed by 'I am hungry', on the single proposition theory they both believe exactly the same thing. Moreover, what they believe is not distinctively about either of them, since both of them believe it. That is, the proposition in question does not make their belief a *de se* belief, a belief about oneself as oneself. To see this we should ask the following question. When Hume responds to his belief by taking action, why doesn't he put food in the mouth of Heimson or someone other than Hume? An answer must appeal to more than the propositional content in question, since both Hume and Heimson believe precisely identical propositions but do not behave in precisely identical ways. When Hume feeds Hume, he acts in a way similar to how Heimson acts when Heimson feeds Heimson. To act identically, they'd have to feed the same person. To explain the lack of strict identity in their actions, the single proposition theory must appeal to the idea that Hume believes the proposition in question to be true

of himself, and the same for Heimson. That is why he puts the food in his own mouth rather than in someone else's mouth.

There is a further problem. Suppose Heimson is wrong and Hume is right—that is suppose Heimson is not hungry but Hume is. The single proposition theory cannot explain this fact with reference only to the proposition in question. Instead, the theory will have to talk of the proposition being true with respect to Hume and false with respect to Heimson.

A standard semantical treatment can mask the need for complication here. Since one of the beliefs is Hume's belief, the semantical treatment of the truth of his belief will yield the answer 'true' for his belief because it automatically evaluates the proposition with respect to Hume. Such a semantics is misleading, however, for nothing about the proposition in question makes Hume the appropriate individual to look to in order to determine whether the proposition is true. After all, some of Hume's beliefs have a truth-value that is calculated in a way that does not involve Hume in any way at all, so we ought to be told why, in this case, things are different. The answer is that the semantical treatment in question masks the fact that we are implicitly assuming that Hume believes the proposition to be true of himself, rather than being a treatment that devolves from the mere fact that Hume believes that proposition.

In both areas, the multiple proposition theory has a simpler story to tell. On this theory, the propositional content has built into it whether it involves Hume or Heimson. This fact allows an explanation in terms of the content of belief why Hume puts food in his mouth rather than someone else's. Moreover, this same fact allows an explanation of how Hume is right and Heimson wrong that appeals only to the truth-value of the propositional content itself.

The neo-Russellian theory of quantification has analogous advantages. On the Fregean picture of quantification, two individuals, call them Hume and Heimson, in different possible worlds, believe what is expressed by a quantified sentence, and in doing so believe precisely the same proposition. Suppose Hume is right and Heimson wrong. We can't explain this feature solely in virtue of the propositional content of the two beliefs, but must rather take Hume to be believing that the propositional content in question is true of his world and Heimson to believe that the content is true of his world. That is why the domain of each world is relevant to the question of who is right and who is wrong.

Just as in the case of the single proposition theory, the standard semantical treatment of the quantifiers can mask the needed complexity.

The semantical treatment automatically adverts to the domain of the world in question, in spite of the fact that there is nothing in the proposition itself that directs things in this way. So, just as in the case of the single proposition theory, the semantics masks the requisite additional feature that the proposition is being believed to be true of the world in question.

In this way, the neo-Russellian view of quantification builds into the proposition expressed what is left as an additional *relatum* on the Fregean view. To the extent that one is attracted to the simpler story of the multiple proposition theory, one has a motivation for preferring the neo-Russellian view of quantification as well. On both the single proposition theory and the Fregean account of quantification, the standard semantical treatment can mask from our view the need for an additional *relatum*. A different approach addresses the need for such an additional *relatum* in the theory of expression itself, yielding a propositional content sufficient to drive the semantical calculation of a truth-value without the need for an additional *relatum* in the nature of belief beyond propositional content.

A critic still might ask whether I take my arguments to show that the neo-Russellian view is a uniquely correct solution to the paradox? I do not claim that. In the end, there may be more elegant and fecund proposals, but I do not know of any. Moreover, I do not even claim that the arguments given for the view are good enough arguments to show that the view is likely to be correct. In the end, the argument I give for accepting the modal indexicality view of quantification is that it is the only game in town. Without it, since there is no other game in town, we must endorse the idea that there is no distinction between knowable and known truth, and that is too much to ask. So, unless and until a better account of the mistake in Fitch's proof comes along, the reasonable conclusion to accept is that the best explanation of the illicit substitution involved in the proof adverts to the neo-Russellian account of quantification.

This point bears emphasis. Critics of anti-realism, such as Williamson, view the paradox as a refutation of (most versions of) anti-realism, with Fitch's proof simply a display of a surprising logical result to this effect.[2] Such approaches to the paradox, however, are not philosophically deep enough. What is paradoxical here is not that Fitch has discovered a proof

[2] A point he made most recently to me at the *Modalism and Mentalism* conference in Copenhagen at the end of January, and most clearly made in print in "On the Paradox of Knowability", *Mind* 96 (1987), 256–261 and in "Verificationism and Non-Distributive Knowledge."

that threatens anti-realism, but rather that Fitch has discovered a proof that threatens a logical distinction between actuality and possibility. One way to put this point is to notice that the knowability claim, if true, is supposed to be necessarily true: it is an implication of a proper understanding of the nature of truth. The claim that all truths are known is not likely to be true, especially if we restrict our domain to finite minds, but it could be true. So if it is true, it would be contingently true at best. Yet, given Fitch's proof, the two are logically equivalent, which they cannot be without have the same modal status. A satisfactory response to the paradox cannot simply swallow this loss without explanation. What we would like best is something like what we get in the case of the same lost distinction in the context of actual and possible necessity: we want a semantic explanation of why such a loss occurs, of the sort we get by appeal to possible worlds and accessibility relations on them.

It is important here to compare the situation involved in Fitch's proof with other proofs that generate surprising results, such as Gödel's incompleteness results. These results are surprising, and threaten important philosophical perspectives, such as Hilbert's formalism. These results themselves are not paradoxical, however. They present no challenge to anything like the edicts of common sense or the viewpoint of received opinion. Should this view of the incompleteness results be mistaken, to that extent we encounter paradox and the need for further explanation as to how the proofs could be correct. As noted already, the paradigm of such explanation is semantic, and it is precisely this lack that makes Fitch's proof paradoxical rather than merely surprising.

Consider for another example Vann McGee's apparent counterexample to modus ponens.[3] The result of McGee's arguments is not merely surprising, but paradoxical. Modus ponens is so well-entrenched a part of our ordinary view of things that our reaction to his arguments is that they must contain a mistake. Suppose, however, that we are wrong. If we are wrong, and McGee right, some explanation is in order. We need to know how it could be that our ordinary view of things could be so mistaken. It is worth noting that McGee attempts just such an explanation: logical rules, he claims, should be thought of as more akin to generalizations and law-like statements in science which can be useful and instructive even if not always completely accurate.

My intension in citing McGee's explanation is not to endorse it, nor to endorse his arguments that modus ponens is not an exceptionless

[3] Vann McGee, "A Counterexample to Modus Ponens", *Journal of Philosophy* 82.9 (September 1985), pp. 462–447.

logical rule.[4] The point is only that when a proof conflicts with ordinary understandings, a further explanatory burden must be shouldered. So it is not enough simply to accept the surprising character of Fitch's result. One must also shoulder the philosophical burden of explaining how the proof could be correct since it implies a lost distinction between actuality and possibility.

The most straightforward such explanation is a semantical one, as we find when explaining the lack of a distinction between actual necessity and possible necessity, and the closeness of this distinction to the distinction between universal actual knowledge and universal possible knowledge might incline us to look for a semantical explanation here as well. Perhaps, though, other explanations could succeed.

Perhaps, but it is not obvious how to give one. Suppose the explanation offered were purely syntactic, as in: the rules of (K-Dist) and (KIT), plus the metalinguistic rule of (RN) are so inherently plausible that the conflict they create with the intuitive logical distinction between possibility and actuality is not paradoxical at all. The results are surprising and unanticipated, but not paradoxical.

I think this answer must be coming from someone living in logical denial. No argument can conclusively show that this approach is mistaken, since the difference between what is paradoxical and what is merely surprising is, perhaps, only a difference in degree and not in kind. Even so, there is a distinction to be drawn here between the unanticipated and the seemingly contradictory, and Fitch's proof engenders the latter experience and not simply the former. It is not merely surprising when we are told that what looks like a necessary truth is logically equivalent to what looks like a contingent truth. We can't simply affirm the rules, and say, "I guess we were wrong; non-omniscience really is impossible." That's simply not an adequate explanation of what's gone wrong; more accurately, it is not an explanation at all.

[4] My own view of the matter is that it is preferable to abandon importation/exportation in response to his arguments. If his example is put in counterfactual form, this response becomes obvious: to say that if Reagan were to lose, then if Anderson were to lose, Carter would win, is to say something false; whereas to say that if both Reagan and Anderson were to lose, Carter would win, is to say something true. Only the former, however, is of any use to McGee's argument.

McGee's argument, of course, involves indicative conditionals rather than counterfactual ones. As a result, more argument is needed to get around his claims. The extra arguments needed, I believe, involve refusing to adopt the assertibility condition semantics he employs, but since that is beyond the scope of the present essay, I will leave that topic for another time and place.

One might carry the syntactic approach further, however. Instead of simply insisting on the plausibility of the above rules, one might try to portray the paradox as a special case of a more general phenomenon. This approach is a good one, and has a history going back to J. L. Mackie's early paper on the paradox.[5]

Here's an example of this approach.[6] The first step involves finding additional operators that generate a Fitch-like result. For example, consider the operator "it is written on my blackboard that" and the operator "it is true that", and the idea that anything true might be written correctly on my blackboard. The idea of such examples is to argue that there are many other operators that mimic the distributivity and factivity features upon which Fitch's proof relies, yielding similar paradoxical results, so that by repetition of result we might begin to be immunized to the thought that something paradoxical is occurring. We could thus add to our list of Fitch-like results those obtainable by using operators such as the "true belief" operator, the "truly wished for" operator, the "truly imagined" operator, the "truly desired" operator, etc. In each case, the existence of thing not truly X-ed will purportedly be incompatible with the idea that any truth can be truly X-ed.

Once we suspect that the Fitch result is an instance of a more general phenomenon, we will then attempt to characterize the more general phenomenon in order to complete the attempt to explain away the knowability paradox. To do so, we will need to find some general characterization of why all of these proofs work, in contrast to our intuitive judgment that there is a logical distinction to be maintained in these contexts. One such explanation we looked at in a previous chapter appealed to the notion of structural interference, claiming that the process of X-ing when applied to a conjunction can cause problems since in X-ing the first conjunct, one may affect the truth-value of the second.[7] Well, not quite, since the claim in question is an eternal truth if a truth at all (it quantifies over all individuals who X and all times), so a more careful claim would be that the X-ing of the first conjunct entails the falsity of the second.

So, the strategy is this: generalize the Fitch result, and then give a unifying explanation of what is occurring in all these instances. Suppose

[5] J. L. Mackie, "Truth and Knowability", *Analysis* 40 (1980), pp. 90–92.

[6] I owe a great deal to Michael Hand regarding this approach. In fact, I think it fair to say that I simply would not have seen the possibility or significance of proof-theoretic insight without the long discussions we have had together.

[7] For the latter perspective on the paradox, Michael Hand, "Knowability and Epistemic Truth", *Australasian Journal of Philosophy* 81(2002), pp. 216–228.

we succeed in doing so in terms of the notion of structural interference—why would that be sufficient to explain away Fitch's result. We would have achieved a kind of fully general understanding of this result, but why would such a fully general understanding eliminate the paradoxicality of the result? After all, moving to this level of generality doesn't imply that the idea of universally known truth is logically necessary and it doesn't explain away our conviction that this idea is clearly contingent. There remains the intuitive idea that we ought to be able to explain how it can be known, from the perspective of one situation, that something is unknown from the perspective of another situation,[8] and the intuitiveness of this idea does not somehow magically disappear once we generalize the phenomenon involved in Fitch's proof. Even if appeal to the notion of structural interference were to convince us that Fitch's proof is an instance of a more general phenomenon, that still leaves us two options. One option is to conclude that no paradoxicality is present, but the other option is to conclude that the little leaven of paradoxicality introduced by Fitch has now been spread to a much larger loaf. And note that nothing in the appeal to structural interference prevents the paradoxical conclusion that the omniscience claim is necessarily true if true, and necessarily false if false. But how could that be? How could the claim that human beings will collectively become at some point omniscient be something beyond an overly-optimistic (contingent) prediction?

It is worth comparing the apparent contingency of this omniscience claim with a more famous claim reputed to have the same sort of property, the claim that God exists. One's intuitive, pre-philosophical attitude toward this claim should be the same as that toward the omniscience claim above: it is a contingent matter whether there is a God. There is a plausible path of reasoning to the denial of the contingency claim, however. It begins by claiming that God is the most perfect being, that He exemplifies maximal greatness. We thus obtain the logical equivalence of the claim that God exists with the claim that maximal greatness is exemplified. The final steps toward a denial of the contingency assumption is to clarify what maximal greatness involves (it is to display the maximal amount of any great-making property that has an intrinsic maxima) and argue that

[8] For the latter perspective, see Dorothy Edgington, "The Paradox of Knowability", *Mind* 94 (1994) pp. 557–568 and Jonathan L. Kvanvig, "The Knowability Paradox and the Prospects for Anti-Realism", *Noûs* 29 (1995), pp. 481–499; for the latter see Hand, "Knowability and Epistemic Truth", and Greg Restall, "Not Every Truth Can be Known: At Least, Not all at Once", draft.

modal stability is itself a great-making property whose intrinsic maxima is existence in all possible worlds.

There are two quite natural responses to this threat to the contingency of the theistic claim. The first is to question the proof itself, to question the implications of the concept of maximal greatness, especially to doubt whether maximal modal stability is itself a great-making property. In doing so, one may look for analogues of the property, or one may simply construct formal notions that are claimed to have the modal stability property of being necessarily instantiated if instantiated. This strategy has a long history of threatening the ontological argument, from Gaunilo's perfect island to Arnauld's existent lion.

This approach is like the proof-theoretic generalizations to the knowability paradox characterized above. It is, however, a more promising approach here, since the examples used do not simply mimic the problematic proof, but constitute reductios of it. What they show is that the proof contains a mistake, even if we cannot identify exactly where the mistake occurs, thereby reaffirming our intuitive sense of the contingency of the theistic hypothesis. In the mirroring of Fitch's proof above, the attempt is not to uncover a mistake in a proof we find discomfiting, but rather to convince us of its validity by finding other contexts in which the same kind of proof can be repeated. This approach will only leave us puzzled, however, since what we need is an account that explains away the apparent contingency of the non-omniscience claim.

Thus the first response to the argument provides no comforting analogue for the non-semantic attempt in question to disarm the knowability paradox, so let us look at the second kind of response. This response questions the account of the theistic claim itself. Why should we think that the claim that God exists is logically equivalent to the claim that maximal greatness is exemplified? After all, it's not as if the meaning (sense) of the term 'God' is the same as the meaning (sense) of the term 'maximally great being'.

Seeing what defenders of the argument do at this point shows why the non-semantic approach we are considering here is unsatisfying. Defenders of the ontological argument sometimes simply stipulate an understanding of 'God' in terms of maximal greatness. Such an approach leaves untouched the intuitive sense of the contingency of the theistic hypothesis, and thus provides no useful model for the pragmatic approach to the knowability paradox to emulate. What is needed instead is some way of explaining away some apparent contingency.

A different approach taken by defenders of the ontological argument attempts to do so. Such an approach takes the form of a reductio of the

denial of contingency, beginning with the the suppositions that there is a God and also that there is a distinct being greater than God. The argument then proceeds by asking what understanding of God one might have that would call for allegiance, or worship, or religious commitment to the lesser being.

This line of argument could be resisted by insisting that a proper conception of God has no religious significance whatsoever,[9] but that escape route will strike most as fairly extreme. My point here, however, is not to defend this approach, but to show it it can be used to explain away the apparent contingency of the theistic hypothesis. The central idea of the reductio strategy is to take other theological commitments one is likely to have and use those commitments to show that there couldn't be a being greater than God, and thus that the existence of God is either necessary or impossible.

To counter the intuitive nature of the contingency of non-omniscience, something similar would be useful. To offer an analogy to this approach, one will need to find an explanation independent of Fitch's proof to convince us that our non-omniscience is necessary. The present approach, in terms of structural interference, doesn't explain away the needed property. But the structural interference idea isn't suited to such an explanation. Instead, it is constructed to be able to explain how the confirmation difficulties involved in possible truths such as "I do not exist" and "no thinkers exist" do not threaten anti-realist assumptions about the epistemic nature of truth. (The first can't be confirmed by me, and the second can't be confirmed by anyone.) The problem is that even if this account is good at explaining these points, it has no resources whatsoever for explaining any mistaken assessment of modal status. It is not as if "I don't exist" was thought to be necessarily false and the notion of structural interference showed it to be only contingently false; nor did we think "no thinkers exist" was impossible and were corrected through the envisioned explanation. As such, the explanation proposed for dealing with the apparent contingency of non-omniscience does nothing at all like the reductio approach cited above for explaining our mistaken intuitions concerning the modal status of the claim that God exists.

At the risk of being overly repetitive, I want to call attention once more to what an appeal to structural interference is good for. The appeal to structural interference is best suited to explain why anti-realists should

[9] See, e.g., Richard Taylor, *Metaphysics* (Prentice-Hall, 1992). Taylor endorses arguments for the existence of God, but takes this result to have no religious significance at all.

not be so quick to endorse the universal knowability assumption. The fact that some claims exhibit features that structurally prevent their verification shows that, if Fitch's proof is in fact a successful proof, anti-realists should be free unapologetically to reject the knowability assumption of that proof. The concept of structural interference, if successful, shows both how "I don't exist" can be unverifiable by me and yet contingent and how "No thinkers exist" can be unverifiable by anyone and yet not be necessary. So if Fitch's proof is successful, then we can, perhaps, apply the concept of structural interference to these results to conclude that anti-realism can escape refutation by that proof.

On this score, the structural interference proposal is not the only attractive proposal to consider to supplement the account presented here to save anti-realism from the threat posed to it by Fitch's proof. Even though Edgington's proposal was found wanting in Chapter 3, her work has spawned a body of literature attempting to use her insights to find a solution to the paradox.[10] Another attempt along these lines is proposed by Rabinowicz and Segerberg. This approach (hereafter "the RS approach") employs a two-dimensional semantics that distinguishes the world being referred to or described from a perspective world in the semantical treatment of the knowledge operator. The ordinary use of appeals to perspectives, or contexts, is in terms of a function from sentences to propositions expressed, i.e., Kaplanian character, but here the perspectives are shifted into the content of the operator itself. In this way, the RS approach involves a kind of contextualism about the knowledge operator.

A similar approach without the contextualism is taken by Sten Lindström, who employs a semantics involving situations rather than worlds (as in the RS approach).[11] What is common to both derives from the insight that a formula says something about a world or situation referred to or described, but what it says about that world or situation depends on the perspective from which the formula is expressed, i.e., from the context of utterance itself.

Put in this way, there are strong similarities between these approaches and the position adopted in the last chapter as a solution to the paradox. The difference is that both approaches just described take as their goal a

[10] See, e.g., Wlodek Rabinowicz and Krister Segerberg, "Actual Truth, Possible Knowledge", *Topoi* 13 (1994), pp. 101–115; and Sten Lindström, "Situations, Truth and Knowability: A Situation-Theoretic Analysis of a Paradox by Fitch", in E. Ejerhedoch and S. Lindström, eds., *Logic, Action and Cognition: Essays in Philosophical Logic* (Dordrecht, 1997). [11] Lindström, "Situations, Truth and Knowability."

clarification of the anti-realist claim about truth within the general strategy outlined by Edgington of treating the universal knowability claim as a careless formulation of the anti-realist or verificationist understanding of truth. Thus, when applying their semantical apparati to the problem at hand, both approaches follow Edgington in reformulating the universal knowability claim in terms of a formula involving an actuality operator. From the perspective of the present approach to Fitch's proof, my canonical claim is this: Fitch's proof creates a paradox as well as a problem for anti-realism, and *these two issues are simply not the same problem*. The fundamental problem addressed here is the paradox itself, and no solution to the problem for anti-realism yields, by itself, a resolution of the paradox.

Seeing this point helps us to appreciate exactly where an appeal to structural interference or a two-dimensional semantics is relevant. To ask such approaches to solve the paradox is to ask too much of them. They have a conditional virtue: assuming their adequacy, they possess the conditional virtue of being able to explain how anti-realism can escape refutation if Fitch's proof is valid. So they are interesting from the perspective of the problem created for anti-realism by Fitch's proof. The paradox itself, however, is a different problem, a problem that arises directly from the question of whether or not that proof is valid (not from the question of whether an anti-realist should endorse the universal knowability assumption). So before applying the concept of structural interference or some version of a two-dimensional semantics to see if anti-realism can be preserved, we first need a solution to the paradox itself.

In short, the paradox should disturb us all, anti-realists and realists alike. It is true that the difference between a paradox and a merely surprising logical result is often not a difference in kind but only a difference in degree. Even so, there are distinctive marks of each that we look for when assessing what kind of a result we have achieved. If, for example, the result is merely one that we had no reason to think was true, we should classify such a result as a surprise. Or, again, if the result is merely one that threatens a particular philosophical perspective, such as anti-realism, we should still classify the result as merely surprising. But when the result threatens some aspect of received opinion, especially received opinion on logical matters themselves, the result will often move beyond the merely surprising to the truly paradoxical. In the present case, the lost logical distinction is part of a firmly entrenched understanding of the nature of the modalities of necessity, possibility, and actuality. It is not a partisan distinction that only certain philosophical perspectives could endorse, and in this way, it is paradoxical to

face a derivation that undermines the distinction, in the same way it is paradoxical to be told that two grains of sand constitute a heap or that motion is impossible. The perplexity engendered by Fitch's proof is paradoxical, and the paradox cannot be addressed either by embracing Fitch's proof as a refutation of anti-realism or by finding a version of anti-realism that involves no commitment to the knowability claim itself. What we need is either an explanation of the failure of Fitch's proof or an explanation of why there is, in fact, no logical distinction between actuality and possibility in the domain in question. Nothing short of that constitutes a proper philosophical response to the paradox.

It is in light of this assessment of the paradox that I submit the semantical way out as the only game in town. No other approach to the paradox offers any hope of addressing the fundamental paradoxicality involved in asserting a lost logical distinction between actuality and possibility. Furthermore, this defense of the solution in the last chapter fits with the quite different perspective on the paradox than that which drives the literature on the topic. The paradox is not a special problem for those who embrace anti-realist conceptions of truth, or who hold to positions that fit naturally with such a conception. In the end, what is paradoxical about the knowability paradox has nothing whatsoever to do with one's conception of truth. Instead, the paradoxicality in question is found in the absurd collapse of the distinction between actual and possible universally known truth.

In the face of this conception of the paradoxicality of the problem discussed here, the attempts to "solve" the paradox by adopting restrictions of one sort or another on the knowability claim are simply non-starters. In the early discussions of the paradox, anti-realists sometimes would bristle at the second assumption used in the proof, the assumption that there are some unknown truths, as if it would somehow help avoid the paradox if one simply rejected that assumption. Such an approach reveals a fundamental confusion about how *reductio* arguments work—the assumptions involved in such a proof do not have to be accepted by anyone for them to reveal logical commitments of other assumptions. Another confusion is displayed by thinking some restriction strategy could solve the fundamental problem raised by Fitch's proof. It is true that Fitch's proof provides a weapon for showing that certain anti-realist views of truth are problematic, but I have argued that the heart of the paradox is found elsewhere. Once one appreciates the heart of the paradox, restriction strategies are as much a non-starter as

are quibbles about whether anti-realists should endorse the additional assumptions used in the proof.

The proper conclusion to draw, then, is that the only way of escape from the paradox is to find a flaw in its logic. That means that either the rules of first-order theory are mistaken, or the special rules for the K-operator should be rejected, or there is a problem of substituting into a modal context. The final diagnosis all should accept is that it is all about the logic, so that if actual and possible known truth are truly distinct, it is only some mistaken facet of the logical assumptions that we are making that can sustain the distinction.

Perhaps the most surprising thing in this entire investigation is the double mistake of the direction of the literature on this problem. The first mistake I have dwelt on extensively, the mistake of trying too much to save anti-realism from the paradox by focusing on restriction strategies regarding the claim that all truths are knowable. The second mistake is equally surprising however. It is the mistake of thinking that a solution to the paradox will help anti-realism. In tandem, these two mistakes lead naturally to a conception and approach to the paradox that focuses fairly exclusively on the defensibility of anti-realism. Avoiding both mistakes puts us in the following position. We shouldn't begin by trying to save anti-realism and where we end up doesn't guarantee the defensibility of anti-realism, either. On the approach to the paradox defended here in terms of failure of substitutivity in modal contexts, the solution to the paradox does not free anti-realism from the variety of threats arising from Fitch's proof. But for those whose primary concern is the paradox itself, the philosophical issues and difficulties for anti-realism can be left for another time and place. The failure of Fitch's proof in terms of a failure of substitutivity in modal contexts saves the logical distinction between actual and possible universal knowledge, whether or not it relieves anti-realists of all the philosophical anxieties that have arisen on the path to this solution.

Bibliography

Alston, William P., 1989. "Levels Confusion in Epistemology." *Epistemic Justification*. Ithaca, N.Y.: Cornell University Press, pp. 153–171.

Beall, JC, 2000. "Fitch's Proof, Verificationism, and the Knower Paradox." *Australasian Journal of Philosophy* 78, pp. 241–247.

Berkeley, George, 1710. *A Treatise Concerning the Principles of Human Knowledge*. Chicago: Open Court, 1986.

Boër, Steven E. and Lycan, William G., 1980. "Who, Me?", *The Philosophical Review* 89, pp. 427–466.

—— 1986. *Knowing Who*. Cambridge, Mass.: Bradford Books.

BonJour, Laurence, 1985. *The Structure of Empirical Knowledge*. Cambridge: Harvard University Press.

Brogaard, Berit and Salerno, Joseph, 2002. "Clues to the Paradoxes of Knowability: Reply to Dummett and Tennant." *Analysis* 62, pp. 143–150.

—— 2002. "Fitch's Paradox of Knowability." *The Stanford Encyclopedia of Philosophy (Winter 2002 Edition),* Edward N. Zalta (ed.) <http://plato.stanford.edu/archives/win2002/entries/fitch-paradox/>.

Chisholm, Roderick, 1976. *Person and Object: A Metaphysical Study*. London: Allen & Unwin.

—— 1981. *The First Person*. Minneapolis: University of Minnesota Press.

Church, Alonzo, 1936. "A Note on the Entcheidunsproblem." *Journal of Symbolic Logic* 1, pp. 40–41.

Cohen, Stewart, 1988. "How to be a Fallibilist." *Philosophical Perspectives* 2, pp. 581–605.

—— 1991. "Skepticism, Relevance, and Relativity." Brian P. McLaughlin, ed., *Dretske and his Critics*. Oxford: Blackwell, pp. 17–37.

—— 1998. "Contextualist Solutions to Epistemological Problems: Scepticism, Gettier, and the Lottery." *Australasian Journal of Philosophy* 76, pp. 289–306.

Cozzo, Cesare, 1994. "What We Can Learn from the Paradox of Knowability." *Topoi* 13, pp. 71–78.

Davidson, B. L., 1985. "Belief *De Re* and *De Se*." *The Australasian Journal of Philosophy* 63, pp. 389–406.

Davies, Martin and Humberstone, Lloyd, 1980. "Two Notions of Necessity." *Philosophical Studies,* pp. 1–30.

DeRose, Keith, 1992. "Contextualism and Knowledge Attributions." *Philosophy and Phenomenological Research* 52, pp. 913–929.

—— 1995. "Solving the Skeptical Problem," *The Philosophical Review* 104, 1, pp. 1–52.

DeVidi, David and Kenyon, Tim, 2003. "Analogues of Knowability." *Australasian Journal of Philosophy* 81, pp. 481–495.
DeVidi, David and Solomon, Graham, 2001. "Knowability and Intuitionistic Logic." *Philosophia* 28, pp. 319–334.
Dewey, John, 1929. *The Quest for Certainty.* New York: Milton, Balch.
—— 1957. *Reconstruction in Philosophy*. Boston: Beacon Press.
Dretske, Fred, 1970. "Epistemic Operators." *Journal of Philosophy* 67, pp. 1007–1023.
—— 1971. "Conclusive Reasons." *Australasian Journal of Philosophy* 49, pp. 1007–1023.
—— 1981. "The Pragmatic Dimension of Knowledge." *Philosophical Studies* 40, pp. 363–378.
Dummett, Michael, 1976. "What is a Theory of Meaning? (II)." G. Evans and J. McDowell (eds.), *Truth and Meaning,* ch. 4. Oxford: Clarendon Press.
—— 1978. "The Philosophical Basis of Intuitionistic Logic." *Truth and Other Enigmas*. Cambridge: Harvard University Press.
—— 1991. *The Logical Basis of Metaphysics*. Cambridge: Harvard University Press.
—— 2001. "Victor's Error." *Analysis* 61, pp. 1–2.
Edgington, Dorothy, 1985. "The Paradox of Knowability." *Mind* 94, pp. 557–568.
Field, Hartry, 1980. *Science Without Numbers*. Oxford: Blackwell Publishing Co.
Fitch, Frederic, 1963. "A Logical Analysis of Some Value Concepts." *Journal of Symbolic Logic* 28, pp. 135–142.
Foley, Richard, 1986. *The Theory of Epistemic Rationality*. Cambridge: Harvard University Press.
Frankfurt, Hary, 1969. "Alternative Possibilities and Moral Responsibility." *Journal of Philosophy* 66, pp. 829–839.
Garson, James, 2001. "Modal Logic." *The Stanford Encyclopedia of Philosophy (Winter 2001 Edition),* Edward N. Zalta (ed.) <http://plato.stanford.edu/archives/win2001/entries/logic-modal/>.
Greco, John, 2000. *Putting Skeptics in Their Place*. New York: Cambridge University Press.
Grice, Paul, 1968. "Logic and Conversation." *Studies in the Way of Words*. Cambridge: Harvard University Press.
Hand, Michael, 2002. "Knowability and Epistemic Truth." *Australasian Journal of Philosophy* 81, pp. 216–228.
Hand, Michael and Kvanvig, Jonathan L., 1999. "Tennant on Knowability." *Australasian Journal of Philosophy* 77, pp. 422–428.
Hart, W. D., 1979. "The Epistemology of Abstract Objects: Access and Inference." *Proceedings of the Aristotelian Society* supplementary, 53, pp. 153–165.
Hart, W. D. and McGinn, Colin, 1976. "Knowledge and Necessity." *Journal of Philosophical Logic* 5, pp. 205–208.

Hawthorne, John, 2004. "Is Knowledge Closed Under Deduction?" *Debates in Epistemology,* Matthias Steup and Ernest Sosa, eds. Oxford: Blackwell Publishing.

—— 2004. *Knowledge and Lotteries*. Oxford: Oxford University Press.

Hintikka, Jaakko, 1961. "*Cogito Ergo Sum*: Inference or Performance?" *Philosophical Review* 72, pp. 3–32.

—— 1962. *Knowledge and Belief*. Ithaca, NY: Cornell University Press.

Hunter, Geoffrey, 1973. *Metalogic An Introduction to the Metatheory of Standard First Order Logic*. Berkeley, Ca.: University of California Press.

James, William, 1907. *Pragmatism*. Cambridge: Harvard University Press, 1975.

—— 1909. *The Meaning of Truth*. Cambridge: Harvard University Press.

Kirkham, Richard, 1995. *Theories of Truth*. Cambridge: MIT Press.

Kripke, Saul, 1963. "Semantical Considerations on Modal Logic." *Acta Philosophica Fennica* 16, pp. 83–94.

—— 1982. *Naming and Necessity*. Cambridge: Harvard University Press.

Kvanvig, Jonathan L., 1986. *The Possibility of an All-Knowing God,* ch. 2. London: Macmillan.

—— 1989. "The Haecceity Theory and Perspectival Limitation." *Australasian Journal of Philosophy* 67, pp. 295–305.

—— 1995. "The Knowability Paradox and the Prospects for Anti-Realism." *Noûs* 29, pp. 481–499.

—— 2003. *The Value of Knowledge and the Pursuit of Understanding*. New York: Cambridge University Press.

Kyburg, Henry, 1970. "Conjunctivitis." Marshall Swain, ed., *Induction, Acceptance, and Rational Belief,* pp. 55–82. Dordrecht: Reidel Publishing Co.

Lewis, David, 1973. *Counterfactuals*. Oxford: Blackwell.

—— 1986. *On The Plurality of Worlds*. Oxford: Blackwell.

—— 1996. "Elusive Knowledge." *Australasian Journal of Philosophy* 74, pp. 549–567.

Lindström, Sten, 1997. "Situations, Truth and Knowability: A Situation-Theoretic Analysis of a Paradox by Fitch." in E. Ejerhedoch and S. Lindström (eds.), *Logic, Action and Cognition: Essays in Philosophical Logic,* pp. 183–210. Dordrecht: Kluwer Academic Publishers.

Luper, Steven, 2002. "The Epistemic Closure Principle." *The Stanford Encyclopedia of Philosophy,* Edward N. Zalta (ed.) <http://plato.stanford.edu/archives/spr2002/entries/closure-epistemic/>.

Lycan, William G., 1996. "Plantinga and Coherentisms." Jonathan L. Kvanvig, ed. *Warrant in Contemporary Epistemology,* pp. 3–23. Totowa, N.J.: Rowman & Littlefield.

Mackie, John L., 1980. "Truth and Knowability." *Analysis* 40, pp. 90–92.

—— 1974. *The Cement of the Universe: A Study of Causation*. Oxford: Oxford University Press.

McDowell, John, 1994. *Mind and World*. Cambridge: Harvard University Press.

McGee, Vann, 1985. "A Counterexample to Modus Ponens." *Journal of Philosophy,* 82, pp. 462–71.

McGinn, Colin, 1984. "The Concept of Knowledge." *Midwest Studies in Philosophy IX 1984: Causation and Causal Theories,* Peter French, Theodore E. Uehling, Jr., and Howard K. Wettstein, eds., pp. 529–554. Minneapolis: University of Minnesota Press.

Melia, J., 1991. "Anti-Realism Untouched." *Mind* 100, pp. 341–342.

Menzel, Christopher, 2003. "Actualism." *The Stanford Encyclopedia of Philosophy (Spring 2003 Edition),* Edward N. Zalta (ed.) <http://plato.stanford.edu/archives/spr2003/entries/actualism/>.

Nozick, Robert, 1981. *Philosophical Explanations,* ch. 3. Cambridge: Harvard University Press.

Peirce, Charles Sanders, 1931–58. *Collected Papers of Charles Sanders Peirce,* 8 vols. Edited by Charles Hartshorne and Paul Weiss (vols. 1–6) and A. W. Burks (vols. 7–8). Cambridge: Harvard University Press.

Percival, Philip, 1990. "Fitch and Intuitionistic Knowability." *Analysis* 50, pp. 182–187.

——1991. "Knowability, Actuality and the Metaphysics of Context-Dependence." *Australasian Journal of Philosophy* 69, pp. 82–97.

Perry, John, 1979. "The Problem of the Essential Indexical." *Noûs* 13, pp. 3–21.

Plantinga, Alvin, 1974. *The Nature of Necessity*. Oxford: Clarendon Press.

——1978. "The Boethian Compromise." *American Philosophical Quarterly* 12, pp. 429–438.

——1982. "How to be an Anti-Realist." *Proceedings of the American Philosophical Association* 56, pp. 47–70.

——1993. *Warrant: The Current Debate*. Oxford: Oxford University Press.

Pollock, John, 1986. *Contemporary Theories of Knowledge*. Totowa, N.J.: Rowman & Littlefield.

Putnam, Hilary, 1975. "The Meaning of 'Meaning' ". *Minnesota Studies in the Philosophy of Science VII: Language, Mind, and Knowledge,* pp. 215–271, Keith Gunderson, ed. Minneapolis: University of Minnesota Press.

——1981. *Reason, Truth, and History*. New York: Cambridge University Press.

——1983. "Reference and Truth". *Realism and Reason: Philosophical Papers, Volume 3,* pp. 69–86. New York: Cambridge University Press.

Quine, W. V. O., 1970. *Philosophy of Logic*. Englewood Cliffs, N.J.: Prentice-Hall.

Rabinowicz, Wlodek and Segerberg, Krister, 1994. "Actual Truth, Possible Knowledge." *Topoi* 13, pp. 101–115.

Rasmussen, S. A. and Ravnkilde, J., 1982. "Realism and Logic." *Synthese* 52, pp. 379–437.

Rorty, Richard, 1981. *Philosophy and the Mirror of Nature*. Princeton: Princeton University Press.

Rosenkranz, S., 2004. "Fitch Back in Action Again?" *Analysis* 64, pp. 67–71.

Rückert, H., 2003. "A Solution to Fitch's Paradox of Knowability." in Gabbay, Rahman, Symons, Van Bendegem, eds., *Logic, Epistemology and the Unity of Science*. Dordrecht: Kluwer Academic Publishers.

Russell, Bertrand, 1912. *The Problems of Philosophy*. Oxford: Oxford University Press.

Salerno, Joseph, 2000. "Revising the Logic of Logical Revision." *Philosophical Studies* 99, pp. 211–227.

Salmon, Nathan, 1986. *Frege's Puzzle*. Cambridge: MIT Press.

Smart, J. J. C., 1963. "Materialism." *Essays Metaphysical and Moral*. Oxford: Blackwell, 1987.

Sosa, Ernest, 1983. "Consciousness of the Self and the Present." J. E. Tomerlin (ed.), *Agent, Language and the Structure of the World,* pp. 131–143. Indianapolis: Hackett Publishing Co.

——1996. "Proper Functionalism and Virtue Epistemology." Jonathan L. Kvanvig, ed., *Warrant in Contemporary Epistemology,* pp. 253–270. Totowa, N.J.: Rowman & Littlefield.

——1996. "Postscript to 'Proper Function and Virtue Epistemology'." Jonathan L. Kvanvig, ed., *Warrant in Contemporary Epistemology,* pp. 271–280. Totowa, N.J.: Rowman & Littlefield.

——2004. "The Place of Truth in Epistemology." Linda Zagzebski and Michael DePaul, eds., *Intellectual Virtue: Perspectives from Ethics and Epistemology,* pp. 155–180. New York: Oxford University Press.

Stanley, Jason and King, Jeffrey C., forthcoming. "Semantics, Pragmatics, and the Role of Semantic Content." Zoltán Szabó, ed., *Semantics vs. Pragmatics*. New York: Oxford University Press.

Stanley, Jason and Szabó, Zoltán, 2000. "On Quantifier Domain Restriction." *Mind and Language* 15, pp. 219–261.

Stich, Stephen, 1983. *From Folk Psychology to Cognitive Science*. Cambridge: MIT Press.

Tennant, Neil, 1997. *The Taming of the True*. Oxford: Oxford University Press.

——2001. "Is Every Truth Knowable? Reply to Williamson." *Ratio* 14, pp. 263–280.

——2001. "Is Every Truth knowable? Reply to Hand and Kvanvig." *Australasian Journal of Philosophy* 79, pp. 107–113.

——2002. "Victor Vanquished." *Analysis* 62, pp. 135–142.

van Benthem, Johan, 2004. "What One May Come to Know." *Analysis* 64, pp. 1–12.

van Fraassen, Bas, 1989. *Laws and Symmetry*. Part II. Oxford: Oxford University Press.

Wansing, Heinrich, 2002. "Diamonds are a Philosopher's Best Friend: The Knowability Paradox and Modal Epistemic Relevance Logic." *Journal of Philosophical Logic* 31, pp. 591–612.
Williamson, Timothy, 1982. "Intuitionism Disproved?" *Analysis* 42, pp. 203–207.
—— 1987. "On the Paradox of Knowability." *Mind* 96, pp. 256–261.
—— 1987. "On Knowledge of the Unknowable." *Analysis* 47, pp. 154–158.
—— 1988. "Knowability and Constructivism." *Philosophical Quarterly* 38, pp. 422–432.
—— 1992. "On Intuitionistic Modal Epistemic Logic." *Journal of Philosophical Logic* 21, pp. 63–89.
—— 1993. "Verificationism and Non-Distributive Knowledge." *Australasian Journal of Philosophy* 71, pp. 78–86.
—— 2000. "Tennant on Knowable Truth." *Ratio* 13, pp. 99–114.
—— 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.
Wright, Crispin, 1987. *Realism, Meaning and Truth*. Oxford: Blackwell.
—— 1992. *Truth and Objectivity*. Cambridge: Harvard University Press.
—— 2000. "Truth as Sort of Epistemic: Putnam's Peregrinations." *Journal of Philosophy* 97, pp. 335–364.

This page intentionally left blank

Index