The Western Ontario Series in Philosophy of Science

A Logical Approach to Philosophy

Essays in Honour of Graham Solomon

Edited by

David DeVidi and
Tim Kenyon

Springer

A LOGICAL APPROACH TO PHILOSOPHY

THE WESTERN ONTARIO SERIES IN PHILOSOPHY OF SCIENCE

A SERIES OF BOOKS
IN PHILOSOPHY OF SCIENCE, METHODOLOGY, EPISTEMOLOGY,
LOGIC, HISTORY OF SCIENCE, AND RELATED FIELDS

VOLUME 70

A LOGICAL APPROACH TO PHILOSOPHY

Essays in Honour of Graham Solomon

Edited by

DAVID DEVIDI
University of Waterloo, Canada

and

TIM KENYON
University of Waterloo, Canada

Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN-10 1-4020-3533-0 (HB)
ISBN-13 978-1-4020-3533-3 (HB)
ISBN-10 1-4020-4054-7 (e-book)
ISBN-10 978-1-4020-4054-2 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

Printed in the Netherlands.

CONTENTS

PREFACE

Graham Solomon, to whom this collection is dedicated, went into hospital for antibiotic treatment of pneumonia in October, 2001. Three days later, on Nov. 1, he died of a massive stroke, at the age of 44.

Solomon was well liked by those who got the chance to know him—it was a revelation to find out, when helping to sort out his affairs after his death, how many "friends" he had whom he had actually never met, as his email included correspondence with philosophers around the world running sometimes to hundreds of messages. He was well respected in the philosophical community more broadly. He was for several years a member of the editorial board for the Western Ontario Series in Philosophy of Science.

While he was employed at Wilfrid Laurier University in Waterloo, Ontario, several of us at the University of Waterloo always regarded our own department as a sort of second academic home for him. We therefore decided that it would be appropriate to hold a memorial conference in his honour. Thanks to the generous financial support of the Humphrey Conference Fund, we were able to do so in May 2003. Many of the papers in this volume were presented at that conference.

We decided to organize both the conference and this volume around an *approach* to philosophy, one which we knew was close to Graham's heart. In calling both the conference and the volume *A Logical Approach to Philosophy* we are trying to suggest what we have in mind. One slogan Solomon and DeVidi liked to toss out, because they almost believed it, was "If you're doing your philosophy right, sooner or later you end up doing algebra." (Though they'd be the first to insist that not all algebra is philosophy.) Somewhat more accurately, Graham was of the view that the methods and results of formal logic were central to the progress of philosophy in recent times in at least two ways.

First, much substantive progress made on traditional philosophical problems was dependent on the application of formal methods. He had a not-so-well-hidden streak of logical positivism, so was not surprised when some careful formal investigation showed a philosophical dispute to be the result of a confusion. Indeed, nothing pleased him better. But while he might have wished the positivists were right about metaphysics, he recognized that they were not. Some problems in metaphysics, he thought, call for solution rather than dissolution. This, coupled with his recognition of the fertility of formal methods, led to his general sympathy with a view, expressed by Dummett and others: that the way to make substantive progress possible on traditional metaphysical disputes, such as disputes about realism and anti-realism in a particular domain, is to recast them as disputes in philosophical logic. Perhaps more obviously, progress in understanding the nature of scientific theories required bringing to the investigation considerable logical and mathematical sophistication.

Secondly, many of the most interesting philosophical issues of the age arise from the need to come to terms with surprising technical results. Much of the best philosophy, but also much of the worst, comes from trying to sort out the lessons of important theorems due to Gödel, Tarski, Cohen, and others, or of Russell's paradox, or of some rather obvious theorems of quantified modal logic. Philosophy often begins just where the formal proof ends.

For the conference and the volume we invited contributions from a number of philosophers who seem to us to do philosophy that falls under this description—that is, their work is informed by the methods and results of formal logic, but is clearly philosophical work. Some of them do work that involves the proof of theorems, but in these cases the theorems are the product of investigation that is motivated by philosophical rather than merely mathematical concerns. Some of them typically do more traditional discursive philosophy, but the investigations bring the methods and results of formal logic to the centre of their discussions. Moreover, we invited contributions from philosophers whose published work we were in position to know for a fact Graham admired greatly.

We also include here a paper by one of the editors, David DeVidi, because he was Solomon's most frequent collaborator—presumably, then, Solomon at least admired those portions of DeVidi's published work for which he was primarily responsible! We also include a paper Graham co-authored, one that DeVidi had read at the Western Canadian Philosophical Association meetings in Regina, Saskatchewan, just two weeks before he died. We publish it in a form only slightly more polished than it was in then so that it can legitimately appear as Solomon's last publication in philosophy.

* * *

It would have been appropriate to have a memorial conference for Graham Solomon on any number of topics. His PhD thesis was on Leibniz and the history of topology, and he published interesting work out of it. He published also on Hume and Hobbes. His interests in the history of philosophy extended forward to central figures in 20th Century philosophy, and he published important work on Russell and Carnap. He was interested in Philosophy of Science, and his first publication was on space-time theories. He was interested in the history and philosophy of mathematics, and had a particular fascination with the histories of model theory and group theory. He co-authored articles on these topics with many different philosophers. We apologize to them for setting the parameters of this volume too narrowly to include contributions on these topics. We hope that the quality of the submissions we were able to include will lead them to agree that this volume is a fitting tribute to Graham.

* * *

We are grateful for financial support from the Office of the Dean of Arts at Wilfrid Laurier University, for the preparation of this volume; and from the Humphrey Conference Fund, for the conference on which it is partially based. Our thanks to Windsor Viney for preparing the index, and to Nancy Davies for helping with the copy editing. Both of us gratefully acknowledge the support of the Social Sciences and Humanities Research Council of Canada during the preparation of this volume.

David DeVidi and Tim Kenyon
Waterloo, June 2005

INTRODUCTION

David DeVidi and Tim Kenyon

The papers in this collection illustrate the manifold roles that logic plays in the most productive contemporary approaches to many traditional problems in philosophy. They range in style from John Bell's austerely formal "Choice Principles in Intuitionistic Set Theory," which includes a number of new technical results rich with philosophical significance, to B. Jack Copeland's charming discussion of the history of completeness proofs for modal logics. Some, such as Allen Hazen's "Do We have the Right Limitative Theorems?" take logic as subject matter, while others, such as Stewart Shapiro's "Externalism, Anti-Realism and the KK-Thesis" and William Demopoulos's "On the Rational Reconstruction of our Theoretical Knowledge," use it as a tool for investigating other areas of philosophy.

While the approaches taken and the problems explicitly addressed in these papers are far flung and varied, a number of recurring themes run through them. The purpose of this introduction is to indicate some of these common themes, and to set the papers in a somewhat broader context so that some of the goals of the authors will be clearer to readers who are not already intimate with the particular issues under investigation.

D. Devidi and T. Kenyon (eds.), A Logical Approach to Philosophy, 1-21.

1 Epistemology and Epistemic Logic

There is no surprise in the observation that logical techniques illuminate epistemology. It is in this field that Stewart Shapiro's lucid paper makes its primary contribution. What *is* surprising about Shapiro's paper is the conclusion he reaches—that epistemic externalism is compatible with the so-called KK-thesis.

The KK-thesis is simply the thesis that results when one adapts a familiar principle of modal logic, often called (4),

$$\Box P \rightarrow \Box\Box P,$$

to the epistemic case. This principle is valid in many modal systems, but is most familiar as the key principle of the modal system S4. At least, in standard formulations the principle (4) is what is added to the very weak system sometimes called KT, in which the principle often called (T),

$$\Box P \rightarrow P$$

is the only axiom scheme beyond those required to ensure that it is a *normal* modal logic we are dealing with.[1] These principles and the various modal systems they can be combined to form become relevant to epistemology once one attempts to formulate an *epistemic logic*—in other words, once one decides to interpret $\Box$ in some way having to do with *knowledge*. For instance, one might interpret $\Box P$ as "Person X knows that P," and so decide to write K in place of $\Box$. It is then relatively uncontroversial that the principle (T) should be valid in epistemic logic, for it amounts to no more than the requirement that knowledge is *factive*—that what is *known* must be *true*. Early investigations of epistemic logic, for instance (Hintikka 1964), argued that, given a suitable interpretation of the knowledge operator, the KK-principle is

[1] A normal system is one in which the scheme $\Box(P \rightarrow Q) \rightarrow (\Box P \rightarrow \Box Q)$ and the rule of *necessitation*, that $\Box P$ follows from P, are valid. The necessitation rule must of course be distinguished from the scheme $P \rightarrow \Box P$ which is rarely valid in interesting modal systems.

also valid. It is still regarded by many as part of the correct logic of *idealized knowability*.

Nowadays, however, it is probably fair enough to say that the KK-thesis is only likely to be acceptable to epistemic *internalists*. Internalism, in this context, is the view that in order for a belief to count as knowledge, among other things, the *knower* must be in possession of a justification for the belief. Obviously, "being in possession of" cannot mean that the justification is something the knower is now or ever has been aware of but, the internalist insists, the justificatory material must at least be *accessible* to the knower. That is, the knower must be able, in suitable circumstances, to bring the justificatory material to awareness. What this means isn't entirely clear, and the details are the subject of intense debate among internalist epistemologists. It is easy enough to see why an internalist might be inclined to accept the KK-thesis, however. For if the knower is in a position to bring to awareness the justification underlying a belief that φ, presumably the knower is likewise in a position to recognize that the material constitutes a justification, and so to know that φ is known.

As Shapiro notes, it is a widespread view that KK is a principle that *only* an internalist could love. Indeed, as he also notes, Simon Blackburn, in the *Oxford Encyclopedia of Philosophy*, comes close to *defining* externalism as the view that one can have $\mathrm{K}(\varphi)$ without $\mathrm{KK}(\varphi)$. Shapiro proposes instead to consider an account of the general structure of the conditions under which a statement of the form $\mathrm{K}(\varphi)$ might be true that is compatible with both internalist and externalist accounts: $\mathrm{K}(\varphi)$ amounts to $\varphi \wedge \mathrm{B}(\varphi) \wedge C_i(\varphi) \wedge C_e(\varphi)$, where $C_i(\varphi)$ are whatever internalist conditions are required for it to be true that $\mathrm{K}(\varphi)$ while $C_e(\varphi)$ are the external conditions. (Set $C_e(\varphi)$ to $\varnothing$ for all φ and one has an internalist account of knowledge. Otherwise we are dealing with some variety of externalism.) By analyzing the requirements for $\mathrm{KK}(\varphi)$ to be true under any such analysis, Shapiro is able to show that, contrary to "common sense," it is possible to be an externalist and yet to hold the KK-thesis. He also presents a plausible case that this is possible *only if* the externalist is willing to accept that the external conditions $C_e(\varphi)$ are *knowable by default*—for instance, that in the absence of reason for supposing that one's situation does not constitute a *normal case*, one *knows* that the situation is normal.

Shapiro offers little indication of how acceptable he takes such an externalist view to be. But he has shown that logical space exists for such a view where few suspected it, and has clearly indicated the philosophical price one must pay to occupy it.

2 Choice Principles and Logic

Two papers in this collection pay considerable attention to a class of recent results which deserve more attention from philosophers than they have hitherto received. These results make clear the close relationship between *choice principles* and certain classically valid but constructively invalid "logical principles." Here we will indicate one of the reasons not made explicit in these papers that these results deserve attention.

It is, we suppose, fairly widely known that Michael Dummett has advanced an account of what is at stake in the various debates between realists and anti-realists that abound in philosophy—realists vs. phenomenalists about the material world, realists vs. behaviourists about mental states, realists vs. nominalists about universals, realists vs. constructivists about mathematical objects, and so on. In rough-and-ready form, Dummett's contention is that *realism* in any one of these debates can be associated with acceptance of *bivalence* for discourse about the disputed putative entities. Slightly more precisely, the claim is that realists, and only realists, ought to hold that every well formulated (e.g., non-vague, non-ambiguous) statement having to do with the subject under consideration is determinately true or determinately false.

But there are other parts to Dummett's view that are perhaps less well known. In works such as *The Logical Basis of Metaphysics*, one finds Dummett suggesting that so-called *intuitionistic logic* is, in fact, logic of the most general sort. That is, the principles of intuitionistic logic are ones which apply in any discourse and regardless of one's metaphysical presuppositions. It is, so to speak, *metaphysically neutral*. Principles sometimes called "logical," for instance those principles of classical logic that are not intuitionistically valid, may well be correct in certain limited domains. But in such cases the *grounds* for accepting the principles will be *metaphysical* rather than logical.

However, whether it is presented by Dummett or his expositors, the explanation of how a metaphysical fact can justify a principle of reasoning can seem somewhat thin. The argument sometimes runs as follows: If a particular discourse is about a domain that is suitably *mind-independent*, then that mind-independent reality is available to fix the truth values of statements in the domain. Hence, provided the statements are suitably formulated, they will either be true or false, even if it is beyond our ability, even in principle, to determine which truth value a particular statement happens to have. Thus realism with respect to a particular subject matter, equated with belief that what is under discussion is a reality that is in some suitable sense mind-independent, is taken to imply bivalence for statements having to do with that subject matter. Finally, bivalence is well known to imply that logic is classical (subject only to a few technical provisos not obviously relevant to the present discussion).

The class of results previously mentioned offers us the prospect of a much tighter explanation of how one may buy "logical" principles with metaphysical coin. Perhaps the best known of these results is *Diaconescu's Theorem*, first proved in topos theory, but later shown to have versions that also apply in, for instance, intuitionistic set theory. The theorem states that in such intuitionistic mathematical theories *the Axiom of Choice implies the law of excluded middle*.

In recent years Bell and others have pursued the topic of the relationships between logical principles and choice principles. For there is a whole range of interesting principles valid in classical logic which fail in intuitionistic logic. Some famous examples are the De Morgan law that for every sentences α and β, $\neg(\alpha \wedge \beta) \rightarrow \neg\alpha \vee \neg\beta$, the principle $(\alpha \rightarrow \beta) \vee (\beta \rightarrow \alpha)$, sometimes called *Dummett's scheme*, or *linearity*, and such quantifier laws as $\neg\forall x\alpha(x) \rightarrow \exists x\neg\alpha(x)$ and $\exists x(\exists y\alpha(y) \rightarrow \alpha(x))$. There are also many choice principles besides the full Axiom of Choice. One interesting example is Hilbert's *logical choice operator* ε, which is governed by the law that for every predicate φ, if any object has φ then the epsilon operator identifies an object which has φ. In (Bell 1993*a*) this principle is shown to imply some of these laws, but absent some further *extensionality* assumptions, not some others.

Bell's paper in the present collection represents the current state of the art in these investigations, and includes many results not previously published. In particular, Bell shows almost all of the most interesting "logical" principles intermediate between intuitionistic and classical logic to be *equivalent to* (and not merely implied by) one or another choice principle.

Since choice principles have the logical structure of generalized statements of conditions sufficient to guarantee the existence of a particular type of entity (often a function of a particular sort, but in the case of ε an object of the sort appropriate to the property in question), a reasonable suggestion is that they encode *metaphysical* presuppositions in a way that is much more transparent than do the corresponding "logical" principles. At the least, these results can greatly refine the question of one's metaphysical commitments in adopting a particular principle, should one accept the Dummettian claim of metaphysical neutrality for intuitionistic logic. Thus these results have important implications for contemporary debates about realism and anti-realism.

3 Epistemology and Anti-Realism

DeVidi's paper takes up issues related to those raised by both Shapiro and Bell. He begins from an argument found in Timothy Williamson's *Knowledge and Its Limits*. Williamson's book is, in large measure, a sustained inquiry into the philosophical implications of externalism of a different sort, namely *semantic externalism*.

Semantic externalism is the view that meaning isn't (merely) in the head. Slightly more helpfully, it is the view that for at least some linguistic items, the meaning attached to those items is determined by factors which need not be, and in general are not, accessible to the users of the language. Most famously, it is taken to be the lesson of Putnam's Twin Earth thought experiments that the environment plays a crucial role in determining the meaning of terms like "water." Suppose an English speaker were miraculously transported to a planet superficially indistinguishable from Earth, where lakes and streams are filled with stuff phenomenally identical to, but chemically distinct from, water. Putnam's argument rests upon the intuition that such a speaker utters a falsehood when in his new environs he says things like, "My, that was a refreshing glass of water." The meaning of "water" in an Earth-English speaker's mouth seems to be partially determined by the nature of the stuff in earthly lakes, streams and plumbing. And the chemical structure of water is certainly not something in general accessible to every speaker who makes meaningful assertions using the word "water."

Williamson argues that this externalism about semantic content has many implications, one of them being that we must be epistemic externalists as well. Along the way, though, he offers a counter-argument to an influential argument due to Michael Dummett. Dummett's argument seems to cast doubt on the sorts of meaning theories favoured by semantic externalists. Dummett argues that what he calls "truth conditional" meaning theories cannot account for the linguistic *understanding* of competent speakers of a language, for if the truth conditions of a statement can transcend any possible evidence for the claims, then there is no prospect for an explanation of how those conditions, and hence the meanings in question, are *grasped* by the speaker.

Dummett suggests that an *assertion theoretic* account of meaning, where the meaning of a statement is determined by the conditions under which its assertion is warranted, is better placed to account for the grasp competent speakers have on meaning. Williamson argues that this is not so. Indeed, an "*X*-conditional" account of meaning, for any *X*, will fail to meet the standard to which Dummett wishes to hold the truth conditional account. Dummett's demand *cannot be met*, because it amounts to a demand that certain psychological states be *luminous*—that is, they must be states such that if one is in them, then one is "in a position to know" that one is in them. Williamson argues that *there simply are no non-trivial luminous states*. Since Dummett's demand is for something that no *X*-conditional meaning theory could provide, the problem must be with the demand, and not with the meaning theories that fail to meet it.

DeVidi's paper in part is the presentation of a counterexample to Williamson's argument. The mathematicians who probably take most seriously the idea that the *meaning* of a mathematical statement is given by its assertion conditions will, like Dummett, see a tight connection between meaning and *proof conditions*. For as almost all will agree, warrant to assert and possession of a proof are closely related in the mathematical case. The view that takes this most seriously is probably the *Propositions as Types* view, which equates each mathematical *proposition* with the *type of the proofs of that proposition*. DeVidi argues that this view offers the prospect of a theory of mathematical meaning that satisfies Dummett's constraint, but also is one on which the state of being warranted to assert need not be (non-trivially) luminous.

The second half of DeVidi's paper suggests that those hoping for an assertion theoretic account of mathematical meaning ought not to take too much heart from the first half of the paper. The first half of the paper argues that Williamson has not shown Dummett's requirements of a meaning theory to be unsatisfiable without appeal to luminous mental states. But the Propositions as Types view leaves one with a stock of mathematical propositions that is considerably sparser than a classical mathematician would be happy with. Of course, it's not a huge surprise that one ends up with some sort of mathematical constructivism if one begins from the equation of truth and provability. However, DeVidi uses the example of the Axiom of Choice to illustrate how, if one cleaves too closely to the Propositions as Types view, one winds up with a theory of mathematical propositions which has rather too few meaningful mathematical claims to keep even most constructivists happy.

4 Logic, History of Logic, and Philosophy

Three of the papers in the collection can plausibly be described as historical. But they illustrate excellence in history of philosophy, and in particular history of logic, in three strikingly different ways.

4.1 An Episode in the History of Modal Logic

B. Jack Copeland's contribution to the volume addresses a very specific historical question: Did Montague have a model-theoretic modal completeness proof as early as 1955? There is some reason to think, based on what he says in a work he published in 1960 about work done in 1955, that Montague at least claimed to have one.

It turns out that the answer is no, but the reason is subtle. More precisely, Copeland reports evidence uncovered in Montague's *nachlass* that Montague did indeed claim a completeness proof in 1955 for certain propositional modal logics, but not for quantified modal logic. But even here, caution is needed. For while the first complete semantics for modal logics had been algebraic, and Montague's semantics is not of that sort, Copeland persuasively argues against the claim that Montague had the first *model theoretic* completeness proofs even for the propositional case. For while Montague did indeed employ a relation between *models* in his interpretation of the modal operators, he nowhere suggested that the models might be interpreted as *worlds*, or *points of reference*, or anything of that sort, nor did he suggest any interpretation of the relation between models that he employed—it served merely as a technical device to ensure that the (K) scheme, $\Box A \rightarrow A$, did not turn out valid. Thus the appropriate conclusion is that Montague's 1955 work was "simply an extension of Tarski's model theory to languages containing modal operators."

This bit of historical detective work is obviously of interest to those who care about the history of modal logic. But nowadays this is probably a larger class of philosophers than might be expected. Recent claims that Kripke's ideas about rigid designators had largely been anticipated by earlier modal logicians, especially Ruth Barcan, were met with angry denunciations. Our impression is that once the anger and vitriol had cleared from the air Kripke's claim to originality was essentially intact, but we were left with a much clearer picture of exactly what was striking and novel about his philosophical insight. Copeland's paper doesn't simply raise a similar priority question about a related accomplishment of Kripke's only to squash it. He provides the illumination without the need to detour through the nastiness, and provides also a compelling answer to the question of what was so special about Kripke's accomplishment in his completeness proofs for modal logic, when many others already had bits and pieces of what Kripke was later able to assemble.

4.2 Logic and the Sweep of History

If Copeland's paper is an example of how one can learn a lot from a careful investigation of a specific question about a particular development, William Demopoulos in "On the Rational Reconstruction of our Theoretical Knowledge," ambitiously traces a particular idea through nearly a century's worth of philosophical development. Demopoulos begins his story with one of Russell's attempts to apply his logical discoveries to solve philosophical problems—the application of the theory of descriptions in his explanation, at the time of *Problems of Philosophy*, of our ability to *refer to* objects with which we are not acquainted. While a lot of attention tends falls nowadays on the use of the theory of descriptions to explain our use of empty names, at least as important for Russell is their use in the defense of the *Principle of Acquaintance*: the claim that every expression we can understand must be composed wholly of constituents with which we are acquainted, in Russell's technical sense of that term. Manifestly, though, we often use, with apparent understanding, names of individuals with which we are not acquainted. By taking proper names to be descriptions and appealing to his well known theory of those, Russell put himself in a position to argue that this was possible because we *are* acquainted with something else: a (possibly complex) propositional function that only the individual in question satisfies. Thus the theory of descriptions is at the centre of Russell's account of our knowledge of the material world on the basis of only minimal assumptions about our experience.

Demopoulos finds similar motivations at work in both Ramsey's and Carnap's accounts of the content of theories. What we now call the *Ramsey sentence* of a theory is well known: it is found by conjoining the sentences of the theory, replacing the theoretical vocabulary by variables of the relevant arity and type, then prefixing to the result appropriate existential quantifiers to bind the introduced variables. The result is deductively equivalent to the original—at least, it is equivalent with respect to determining which sentences involving only observational vocabulary are derivable and which are not. The parallel to Russell's application of his theory of descriptions is obvious enough: the observational vocabulary presumably has something to do with the possibility of acquaintance, and Ramsey is showing the dispensibility of terms apparently referring to things with which we are not or cannot be acquainted. As Demopoulos argues, Ramsey and Russell differ in that there is no suggestion in Ramsey that the Ramsey sentence of a theory and the original theory are logically equivalent.—A fortiori the Ramsey sentence doesn't provide an *analysis* of what the statement of the original theory means.—But it nevertheless seems a likely candidate for an account of how our knowledge of matters theoretical is founded on things known by acquaintance.

Carnap's mature account of our theoretical knowledge is naturally seen as a next step in this tradition. Carnap, too, employs Ramsey sentences and takes them to encapsulate the "factual" content of a theory. The key proposal in his theory is that the "Carnap sentence," as it has come to be called, captures the *theoretical* content of a theory. If θ is the sentence giving the original theory, and $R(\theta)$ is its Ramsey sentence, $R(\theta) \rightarrow \theta$ is the Carnap sentence of the theory. It amounts to the statement that if *any* entities satisfy the conditions specified in the Ramsey sentence, then the entities referred to by the theoretical vocabulary of the theory do so.

Demopoulos has more than one goal with respect to Russell's account of our knowledge of the material world and Ramsey and Carnap's theories of theories. First, as we have explained, he points up the ways in which these authors share motivations. But in doing so he also clarifies the philosophical goals these authors have in mind. All these authors have in mind to address a fundamental epistemological problem to do with our knowledge of things with which we are not acquainted. This does not always seem to be recognized by commentators on their work. For instance, Carnap's discussions are often regarded as a specialized contribution to the philosophical analysis of areas of science that are ripe for axiomatization.

Important though these attempts may be, Demopoulos wants to isolate why these attempts need to be classified as *interesting* but *wrong*. To this end, he presents a clever argument appealing only to basic results of model theory to show that Carnap's account of the content of our theory leaves out something essential to our pretheoretical idea of what it means for a theory to be true. For supposing the theory θ to be consistent, Demopoulos shows that there must then be a model of θ in the "intended domain" of the theory. But, Demopoulos argues, this trivializes what it means for θ to be true. Hence Carnap's account leaves something crucial out of account.

Finally, Demopoulos does not want to leave us with the impression that these are worries that later philosophers of science have successfully left behind. In particular, he argues that Bas van Fraassen's constructive empiricism runs into precisely parallel difficulties. So it is not simply a matter of these problems being due to the attachment of Carnap and Ramsey to an outdated syntactical notion of theories: model theoretic accounts which share some of the key motivations of Carnap and Ramsey meet the same problems. In the end, Demopoulos takes himself to have shown not only that the account given by Russell, Ramsey and Carnap "possess internal coherence and elegance, but more importantly, a degree of philosophical motivation not matched by its rivals."

4.3 Why History didn't take a Different Turn

A.P. Hazen's paper takes on many tasks at once. In (Tennant 2000), Neil Tennant argues as follows: We now know that two fundamental desiderata of axiomatization, namely categoricity and deductive completeness, are not jointly realizable for the interesting cases. This is something that became clear to mathematicians and logicians in the wake of Gödel's Incompleteness Theorems. However, it *might* have been proved on the basis of fairly basic properties of the notions involved. Why, then, Tennant asks, didn't mathematicians discover this fact sooner? The ostensible purpose of Hazen's paper is to answer Tennant's provocative question. Indeed, Hazen offers three distinct answers to the question. We don't want to steal his thunder by giving them away here.

But the response to Tennant, while independently interesting, is really only a small part of what Hazen accomplishes in this paper. The paper is clearly written with a non-specialist audience in mind, and is largely devoted to a wonderfully lucid description of the key concepts and results needed to understand Tennant's claim and Hazen's reply. The paper includes: a brief but clear introduction to the history of axiomatization; an explanation of the concepts of deductive completeness and categoricity and why they matter; an explanation for the rise of first order axiomatics; an introduction to the Löwenheim–Skolem theorem; a clear explanation of why the Compactness Theorem is so important; and a clear introduction to the notion of a non-standard model. The clarity of the explanations and the clever interweaving of the parts make Hazen's paper a thing of beauty that logic teachers will want their students to read. ("*This* is why we spent so much time proving Compactness!")

In his reply to Tennant, Hazen makes inspired use of type theory, interpreted as a many-sorted first order language. Here again he does not just discuss Tennant, but takes the opportunity to treat matters that even many logicians could probably stand to have clarified for them, including the relationship between *higher-order* and *many-sorted first order* languages, and the question of what is lost and what preserved when one makes the standard translation from a many sorted first order language to a single sorted one.

5 Non-Classical Logics and Pluralism

The papers by Bell and DeVidi both are concerned with the philosophically pregnant relationship between intuitionistic and classical logic, and with differences between various intuitionistic systems. Several of the other papers in this collection address other issues surrounding non-classical logics.

5.1 Expressive Power, Paradox and Negations

In section 2, we briefly described Michael Dummett's views about the relationship between disputes between realists and anti-realists of various sorts and commitment to classical or non-classical logics. The connection between the metaphysical and logical issues arise because of Dummett's contention that certain traditional metaphysical debates are best regarded as debates about the correct *theory of meaning* for the language used in discussions in each of these disputed domains. Skipping the details that make the view interesting, Dummett suggests that advocates of realism have, in truth-conditional semantics, a relatively clear idea of what a theory of meaning should look like (though Dummett thinks there are serious objections to the idea that such a meaning theory will work). Anti-realists, on the other hand, typically lack any clear idea about how an anti-realist theory of meaning for the domain in question should look. The one well-developed anti-realist theory of meaning on offer, Dummett suggests, is provided by constructivist mathematics. According to Dummett, this means that anti-realists with respect to non-mathematical domains should take seriously the possibility that their own domain-specific anti-realist theory of meaning will look similar to constructivism, and so will require similar modifications to logic.

Solomon and DeVidi take up a difficulty that confronts those who hope to use constructive logics for something other than the purely mathematical-foundational purposes for which they were originally developed. If the *only* operators available are those of intuitionistic logic, then we seem to be left unable to express things we manifestly *can* express, and, indeed, must be able to express if the advocacy of intuitionistic rather than classical logic is to provide the philosophical payoff claimed for it. They quote Dummett on the need, even in accounts of constructive mathematics, to consider not merely the intuitionistic negation that occurs in mathematical statements, but also a sort of "empirical negation" that allows us to say, for instance, that before 1892 the statement that "π is transcendental" was *unproven* but nevertheless true, since provable. Intuitionistic negation cannot be involved in the claim of unprovenness, since intuitionistic negation carries with it the implication of some sort of *impossibility*, and clearly the above claim was not meant to convey that the proof was *impossible* before 1892.

Moreover, intuitionistic logic has often been recommended as a palliative for the philosophical distress caused by one or another paradox. It has, for instance, been recommended as part of the solution to the paradoxes of vagueness in (Putnam 1983, DeVidi to appear), and argued against in (Read & Wright 1985, Williamson 1996). In (DeVidi & Solomon 2001) it is recommended as a solution to the so-called "paradox of knowability." The heart of this paradox is that *given classical logic* and assuming certain uncontroversial principles about knowledge and necessity, the claim that all truths are knowable implies that all truths are known. Commitment to the latter claim is presumably distressing to anyone, as everyone recognizes the existence of truths that go forever unknown. But the sorts of anti-realists views we have been discussing are very much tied up with the claim that no truth is in principle unknowable. Given that the ties between anti-realism and intuitionistic logic are familiar nowadays, it is heartening for anti-realists that the derivation of the unwanted conclusion requires the application of a classical principle that fails in intuitionistic logic.

But if one has *only* the intuitionistic negation operator available, one will not be able to express the claim "there are truths that remain forever unknown," for just the same sort of reason one couldn't say that certain statements at certain times were unproven but not unprovable. Intuitionistic logic absolves the knowability thesis of its irksome classical consequence that there are no forever-unknown truths, but it seems to do so at the cost of making the worry itself inexpressible.

Solomon and DeVidi attempt to sketch a semantics for a second, "empirical" negation that can be added to intuitionistic logic and so provide the system with the expressive power that is lacking in its absence. Obviously, this negation needs to be something other than the usual classical negation, on pain of resurrecting the paradox. They also discuss a number of other philosophical issues that arise when one attempts to apply constructive logic in situations where constructivism is not being advocated.

J.C. Beall addresses philosophical issues that arise for philosophers who make a quite different sort of response to the existence of various paradoxes. The standard account of how to respond to paradox goes something like this: a paradox arises when seemingly reasonable presumptions allow the construction of seemingly valid arguments to seemingly absurd conclusions; so an adequate response to the paradox involves diagnosing which presumption is *only apparently true*, or why one of the arguments *only appears to be valid*, or why the conclusion only appears to be absurd. Each kind of diagnosis is philosophically familiar, including the third—as Quine noted decades ago, a result may be regarded by one generation as paradoxical and by the next as a theorem. But for some paradoxical conclusions this third type of resolution has seemed too radical, requiring the rejection of principles at the heart of basic intelligibility. An example is the Liar sentence (*This sentence is false*) which appears to be true if and only if it is false. Accepting that the Liar is both true *and* false is traditionally seen as obviously unworkable, a prospect eviscerating the concept of truth (and hence those of truth preservation and negation besides). Thus responses to the Liar and similar semantic paradoxes has been focused on responses of the first two sorts. One of the key motivations for dialetheism, the view Beall investigates in his paper, is the suspicion that the reason no satisfactory resolution of the semantic paradoxes has been forthcoming despite decades of intense effort is that *no diagnosis of either of those sorts is forthcoming. Dialetheists*, in effect, ask us to consider the possibility a response of the third sort even to the semantic paradoxes—perhaps the conclusion is only *apparently absurd*.

Dialetheism, to be precise, is the view that for some sentence P, both P and $\neg P$ are true. Of course, if one is to be a dialetheist without accepting the disastrous consequence that all sentences are true, one needs to reject the principle, sometimes called *explosion*, that everything follows from a contradiction. A logic that rejects explosion is *paraconsistent*, but not all advocates of paraconsistent logics are dialetheists. Relevance logics are paraconsistent, for instance, but as the paper by Stephen Read makes clear, not all advocates of relevance logic are dialetheists.

There seems to be plenty of room to debate what is involved in forming the *negation* of a statement, and for arguing that classical negation doesn't quite get things right. For instance, if we think back honestly most of us will recall finding explosion an outlandish principle when we first encountered it. So if explosion is part of the classical meaning of negation the relevance logicians at least have a prima facie case to make for their negation as closer to right. On the other hand, classical, intuitionistic and minimal negations all capture nicely the intuition that "not P" should be the *weakest statement incompatible with P*—that is, any statement incompatible with P implies "not P," so it can profitably be regarded as a sort of disjunction of all possible statements incompatible with P.

But the intuitiveness of this latter account of what's involved in negation is the source of the obvious objection to dialetheism. For if any intuition about negation seems most solid, it is that $\neg P$ must be *incompatible* with P, and this is precisely what the dialetheist must deny. A related objection to standard dialetheism is that is suffers similar "expressive limitations" to those outlined above for intuitionistic logic: how, for instance, does a dialetheist say, as we certainly appear able to do, that a particular sentence is *true, but not false* if the only negation available is a dialetheic one on which mere truth of a negation fails to rule out truth of what is negated?

Beall offers a novel "double aspect" account of dialetheic negation that, he argues, allows us to make some progress against these objections. The basic idea is that negation normally behaves classically, but in a certain class of abnormal "paradoxical set-ups" it exhibits different characteristics—it is "free floating" in a sense Beall tries to make precise. This allows an explanation of the strength of our intuitions about, for instance, the incompatibility of statements and their negations, since our intuitions about negation evolve in response to normal cases, not paradoxical set-ups. That is, our grasp of negation arises in response to the classically behaving cases, and not from a few "odd paradoxical sentences—sentences that arise out of mere grammatical necessity (or the odd contingent state of affairs)." And that, Beall argues, allows the dialetheist to circumvent the objection that a dialetheic language is unable to express claims like "*P* is not both true and false."

5.2 More than One Logic?

Beall has co-authored with Greg Restall a series of articles (Beall & Restall 2000, Beall & Restall 2001) devoted to a defense of a version of *logical pluralism*, i.e., of the view that there is, in some important sense, more than one correct logic. Likewise, Bell, DeVidi and Solomon co-authored a book titled *Logical Options* (Bell, DeVidi & Solomon 2001) whose title betrays their pluralistic inclinations. This collection closes with Stephen Read's vigorous defense of the contrary view, i.e., of the view that there is *one true logic*.

This description of Read's paper, not to mention its position in the book, seems to cast Read in the role of fighting a rearguard action in defense of the received view. The paper includes a direct attack on Beall and Restall's version of pluralism, arguing that it is incoherent. It is probably still true to call logical monism "the received view." And the main thrust of Read's diagnosis of where Beall and Restall go off the rails—that advocates of pluralism of this sort *illegitimately* assume a classical metalanguage in their analyses of various non-classical systems—has a certain pedigree. As well as being a sort of folklore objection to the enterprise of advocating non-classical logic for any philosophical purpose, it is a persistent sort of objection to the best known sort of pluralism from pluralism's previous heyday. In (Carnap 1937, § 17), we get the pithiest statement of the *principle of tolerance*, a principle that, one way or another, Carnap advocated throughout his philosophical career:

> In logic, there are no morals. Everyone is at liberty to build up his own logic, i.e. his own form of language, as he wishes. All that is asked of him is that, if he wishes to discuss it, he must state his methods clearly, and give syntactic rules instead of philosophical arguments.

Michael Friedman, in (Friedman 1999, p.230), argues that this position is unstable. For, Friedman suggests, in order to formulate "syntactic rules" for a classical language in a manner satisfactory to Carnap, as Carnap clearly intends us to be able to do, we must presume the resources of a classical metalanguage. But this is incompatible with the spirit of the principle of tolerance, which requires of us that we employ in the metalanguage in which we compare logical systems only principles acceptable to everybody.

But it would be a mistake to regard Read's argument as essentially conservative. For while Read advocates the view that there is one true logic, he holds that the true logic is *relevance logic*. He argues that taking a particular non-classical logic as correct—something he thinks advocates of relevance logic rightly do, and advocates of constructivism mistakenly do—ought to restrict one's meta-linguistic investigations to those in which that logic's principles are employed. With this methodological principle in place, he is able to explain why relevance logicians have had so much difficulty giving, for instance, a satisfactory formal semantics for the negation operator of their system. For getting the right validities to obtain when such accounts are given in classical metalanguage seems to require the invention of dubious "states" in which both A and $\neg A$ can hold. This move brings with it the need to distinguish these states from possible worlds (or fragments thereof), or else to swallow the bitter dialetheic pill that would allow both A and $\neg A$ to hold in a single *possible* world. If one instead insists on "doing one's semantics in the logic in which one believes," the negation clause can be the homophonic clause familiar from first courses in logic: $\neg P$ is true if and only if P is not true. The mess arises when one decides to speak as though classical "not" is intended on the right. One might think of this tendency as learned classicism making its way back into one's metalanguage—a sort of circular confirmation bias for the intuitiveness of classical logic.

6 A Logical Approach

The articles in this collection assign logic a central role in the practice of philosophy. The variety of areas this small number of articles reach into, and the quality of the contributions they make there, speaks to the productivity and importance of such an approach to philosophy. This probably won't be news to most people who are actually reading this book, since it's not news that the evolution of formal logic in the past 150 years has been a driving force in philosophy (and other fields) in that time. But pointing out what should be obvious is perhaps worthwhile in a time where one sometimes hears it claimed that logic is a spent force in philosophy, or even that its influence has somehow been more for ill than for good. The few articles in this collection make contributions in some areas in which the central role of logic goes without saying, such as philosophy of logic and history of logic. They contribute also to areas where it should come as little surprise, such as formal semantics, foundations of mathematics and philosophy of science. But they also contribute to areas that ramify into every branch of philosophy—epistemology, theories of meaning, realism and anti-realism, theories of truth. We therefore commend the logical approach to philosophy to you, whatever your preferred area of philosophical study.

Chapter

ONE

EXTERNALISM, ANTI-REALISM, AND THE KK-THESIS

Stewart Shapiro

1 The KK-thesis

The purpose of this paper is to study the connection between the so-called "KK-thesis" and externalist epistemology. The focus is on the epistemic status of the external conditions on knowledge. Do we know that the external conditions hold (when they do)? What sort warrant do they need? The point is broadly logical, and so I will try to be non-specific (and thus neutral) on the complex and subtle issues on which epistemologists disagree.

Focus attention here on a single subject, whom we shall name "Karl." Karl is an ordinary human being, with normal epistemic powers (whatever those may be) and a normal range of experience. We are concerned with theories and theses concerning Karl's knowledge at a given time. Let Φ be an arbitrary sentence, or proposition. For now, we follow standard views that "Karl knows Φ" entails that Φ is true, that Karl believes Φ, and that Karl's belief is justified. No doubt there are other necessary conditions needed to handle Gettier-type cases and other pathological situations. Epistemologists occupy them-

D. Devidi and T. Kenyon (eds.), A Logical Approach to Philosophy, 22-35.

selves with the nature of the justification required for Karl's knowledge, and the relationship between the justification of Φ and its truth. The dispute between externalists and internalists is a case in point.

The KK-thesis is the main principle of the modal logic S4. In symbols it is

$$K(\Phi) \rightarrow KK(\Phi).$$

To get an epistemological thesis, of course, we must interpret the K-operator. One option would be to understand K(Φ) as "Karl knows that Φ." So the KK-thesis would be that if Karl knows that Φ then Karl knows that Karl knows that Φ. This is clearly false for some knowers and some sentences Φ. For example, Karl may know that Yankees won the World Series in 1929, but he may not know that he knows this for the simple reason that he never bothered to formulate an explicit belief that he knows that the Yankees won the World Series in 1929.

To obtain a plausible thesis, we have to idealize the outermost "K." The idea is that if Karl knows that Φ, then Karl *can come to know* that he knows that Φ, or, in other words, it is know*able* (by Karl) that Karl knows that Φ. What of the modality invoked here? I'd like to stay as neutral and noncommittal as possible, and yet still have a reasonable framework for epistemology. Let us say that Φ is knowable if Karl can come to know that Φ by ordinary reasoning (deduction and induction perhaps) and introspection. The crucial idea is that Φ is knowable if Karl can come to know that Φ without getting any more information about the external world.

We will also idealize the inner K-operator in the KK-thesis, since it is harmless and convenient to keep things uniform in this way. So the thesis is that, for any sentence Φ, if Φ is knowable, then it is knowable that Φ is knowable. In the jargon of possible worlds, the KK-thesis says that if there is an accessible world w in which Karl knows that Φ, then there is a world w' accessible from w in which Karl knows that Karl knows that Φ.

The accessibility relation here is transitive. Suppose that world w_2 is accessible from w_1, and that w_3 is accessible from

w_2. By hypothesis, in w_2 Karl has done some reasoning and introspection from what he knows in w_1, and in w_3 he has done some reasoning and introspection from what he knows in w_2. So by combining, we see that in w_3 he has just done some reasoning and introspection from what he knows in w_1.

The KK-thesis is held by many philosophers and logicians. In epistemic logic, for example, S4 is often regarded to be the logic of idealized knowability (see (Shapiro 1985)), either in mathematics or in general. The logic is employed in the modal translation of intuitionistic languages, due to (Gödel 1933). Transitive Kripke structures are complete for S4.

The KK-thesis is important for the general matter of semantic anti-realism. The anti-realist claims that all truths are knowable. Since anything known—or knowable—is true, we have that for the anti-realist, Φ is true if and only if Φ is knowable. It is a platitude that if Φ is true, then it is true that Φ is true. If this were not so trivial, we might give it a fancy name, like the "TT-thesis." Given the identification of truth with knowability, the anti-realist must hold that if Φ is knowable, then it is knowable that Φ is knowable. So the anti-realist must accept some form of the KK-thesis.

Prima facie, however, the knowability invoked in anti-realism is not the one involved in usual discussions of the KK-thesis, nor is it the idealization articulated just above. I suggested that for the KK-thesis, Φ is knowable if Φ can come to be known (from what is already known) on the basis of reasoning and introspection. When the anti-realist claims that all truths are knowable, she does not commit herself to the (patently false) thesis that for any truth Φ, if Φ is true, then we can come to know Φ by reasoning and introspection alone. Surely, there are some truths for which we need to do further observations to establish their truth.

There may not be such a gap in mathematics, however. The widely (but not universally) held thesis that mathematics is a priori suggests that if Φ is a knowable mathematical proposition, then Φ can become known on the basis of reasoning alone. In mathematics, the epistemic standard is *proof*. Of course, if logicism fails, then we cannot prove everything, but the traditional view is that the axioms of basic mathematical theories are known a priori. Most, if not all, of the rest is deduction from those axioms. So it does seem that knowable mathematical propositions can become known on the basis of reasoning (and introspection). So on the traditional conception of mathematics, it seems that the anti-realist is committed to the KK-thesis, in pretty much the present form.

2 Internalism

In the entry on epistemology in the second edition of *The Cambridge Dictionary of Philosophy*, Paul Moser formulates a

> …requirement that justificational support for a belief be *accessible*, in some sense, to the believer. The rough idea is that one must be able to access, or bring to awareness, the justification underlying one's beliefs. … *Internalism* regarding justification preserves an accessibility requirement on what confers justification, whereas epistemic externalism rejects this requirement. (p. 276)

Moser adds that "internalists do not yet share a uniform detailed account of accessibility." I do not think we will need one here.

The KK-thesis is often justified on internalist grounds. In mathematics, for example, it is commonly argued that if we have a proof of a mathematical proposition Φ, then we can inspect the proof (or introspect) and realize that it is a proof. Intuitively, this inspection is the accessibility needed for the original knowledge claim.

Let us begin with a very simple, perhaps naive internalism. Suppose we put aside pathological, Gettier-type cases and hold that knowledge is justified true belief. I am not sure that there are any Gettier cases in mathematics anyway, the case that interests me most.

Let $B\Phi$ abbreviate "Karl believes Φ" and let $J\Phi$ abbreviate "Karl is justified in believing Φ." So "Karl knows Φ" comes to $\Phi \wedge B\Phi \wedge J\Phi$.

For any proposition Φ, our internalist holds that by introspection and reasoning, Karl can assess whether he believes that Φ and whether he is justified in this belief. For the latter, he checks to see if his warrant is sufficient for knowledge. In mathematics, he just checks to see if his proof is good. Such a check seems to be a reasonable gloss on Moser's "accessibility," and it brings that notion in line with the accessibility relation sketched above.

There are two ways to show that our naive internalist is committed to the KK-thesis. First, assume $K(\Phi)$. So Φ is knowable. Let w be a world in which Karl knows Φ, recalling that the only difference between w and the actual world is that Karl does some more reasoning and introspection in w. In w, Karl believes that Φ, and this belief is justified. That is, in w, $\Phi \wedge B\Phi \wedge J\Phi$ holds. By the internalism, Karl can then perform some reasoning and introspection to learn that he believes Φ and that he is justified in this belief. That is, there is a world w' accessible from w in which $KB\Phi$ and $KJ\Phi$ both hold. Presumably, we also have that $K\Phi$ holds in w'. Karl surely cannot lose knowledge by the relevant introspection and reasoning (unless the introspection and reasoning confuse him, but let us ignore that by invoking the usual idealizations). So we have that in w', $K\Phi \wedge KB\Phi \wedge KJ\Phi$. This entails that in w', Karl knows that $\Phi \wedge B\Phi \wedge J\Phi$. So $K(K\Phi)$ holds in w.

This argument assumes that since $K\Phi$ is equivalent to $\Phi \wedge B\Phi \wedge J\Phi$, then $K(\Phi)$ is equivalent to $K[\Phi \wedge B\Phi \wedge J\Phi]$. This is an instance of the inference scheme, $\Phi \equiv \Psi \vDash K\Phi \equiv K\Phi$, which is not valid in general. There are, however, reasons to accept the inference in the present case. First, we might assume that Karl himself knows or can know (by introspection and reasoning) that knowledge is justified true belief. That is, we might assume $K[K\Phi \equiv \Phi \wedge B\Phi \wedge J\Phi]$. Then we can safely conclude that $K(K\Phi) \equiv K[\Phi \wedge B\Phi \wedge J\Phi]$, via the above reasoning.

The inference in question may be good even if Karl does not know that knowledge is justified, true belief. The naive internalism in question is not a mere statement that $K\Phi$ is materially equivalent to $\Phi \wedge B\Phi \wedge J\Phi$. The theory is that knowledge *just is* justified, true belief. Thus, on the theory in question, $KK\Phi$ says that what is knowable is that $\Phi \wedge B\Phi \wedge J\Phi$, for that is what $K\Phi$ is.

The matter may not be so straightforward. A triangle "just is" a plane figure whose angles sum to two right angles. It does not follow from "Karl knows that his block is a triangle" that "Karl knows that his block is a plane figure whose angles sum to two right angles." Perhaps one might say that the formula $\Phi \wedge B\Phi \wedge J\Phi$ gives the semantic content or the meaning of the statement $K\Phi$. Then if Karl believes that $K\Phi$, then he must know that $K\Phi$ is $\Phi \wedge B\Phi \wedge J\Phi$. Otherwise, Karl has a belief that he does not understand. However, this is a most implausible thesis. It entails that the only people that believe that they know something are those who have figured out that knowledge is justified, true belief. That is, nobody but the naive internalist can have beliefs about what she knows.

Presumably, the equation of $K\Phi$ with $\Phi \wedge B\Phi \wedge J\Phi$ is a proposed *philosophical analysis*. Philosophical analyses are presumably intermediate between claims of meaning equivalence and claims of necessary equivalence. The issue here is whether it follows from the claim of philosophical analysis that $K(\Phi)$ is equivalent to $K[\Phi \wedge B\Phi \wedge J\Phi]$. It seems that this depends on whether the analysis can be obtained by reasoning and introspection alone. If the analysis is a priori, then the foregoing inference is valid, but I do not venture a further opinion.

I went through this exercise because a similar situation will come up later with externalism (and with a less naive internalism). In the present case, however, the inference in question can be avoided. Again, let w be a world in which Karl knows Φ, so that $\Phi \wedge B\Phi \wedge J\Phi$ holds in w. By introspection, Karl can realize that his belief that Φ is justified, and he can formulate a belief that he knows $K\Phi$ (if he has not formulated this belief already). So there is a world accessible from w in which BKΦ holds. Moreover, this last belief can be justified: upon introspection, Karl's warrant for Φ becomes a warrant for his belief that he knows Φ. So there is a world w' in which BKΦ and JKΦ holds. If we assume, as above, that Karl still knows Φ in w', then we have that in w', KΦ is true, believed, and justified. So in w', Karl knows KΦ. So KKΦ holds in w.

3 Externalism

Perhaps the intuitions underlying the KK-thesis are internalist. If so, then if we opt for an externalist epistemology, we lose the basic motivation for the KK-thesis. Of course, this does not entail that the KK-thesis fails on a particular externalism. Nevertheless, there is a widespread view that the KK-thesis does fail on externalist epistemologies. Part of the entry on "externalism" in Simon Blackburn's *Oxford Dictionary of Philosophy* reads:

> In the theory of knowledge, externalism is the view that a person might know something by being suitably situated with respect to it, without that relationship being in any sense within his purview. The view allows that you can know without being justified in believing that you know. (p. 133)

According to Blackburn, then, the externalist holds that Karl can know Φ without being justified that he knows Φ. If we add that at least sometimes Karl cannot come to know that his belief in Φ is justified just by reasoning and introspection, we can have KΦ without KKΦ. In other words, the KK-thesis fails.

However, this is pretty close to just defining "externalism" as "an epistemology in which the KK-thesis fails." It would be more instructive to characterize externalism independently, and then explore what is required for a given epistemology to sanction or refuse the KK-thesis.

Let us settle on a generic formula for knowledge. Say that Karl knows Φ if and only if:

$$\Phi \wedge \mathrm{B}\Phi \wedge \mathrm{C}_i(\Phi) \wedge C_e(\Phi),$$

where $C_i(\Phi)$ gives the "internally available" conditions (if any) and $C_e(\Phi)$ gives the external conditions (if any) for knowledge. If $C_i(\Phi)$ holds, then Karl can determine that it does by routine reasoning and introspection. That is what we mean by saying that the conditions are internally available. With $C_e(\Phi)$, of course, we do not have this. The conditions may obtain without Karl being aware that they do, and it may be that no amount of introspection or reasoning will yield a sufficient warrant for $C_e(\Phi)$.

Of course, the details of C_i and C_e vary from epistemology to epistemology. The internalist holds that C_e is empty, or vacuous, while the externalist places substantial conditions in C_e. On some versions of externalism, C_i is empty, or all but empty, and $C_e(\Phi)$ is something like "Karl's belief in Φ was produced by a reliable mechanism," usually with some nuances added. Presumably, Karl cannot tell by reasoning or introspection alone that his mechanisms are functioning reliably. Thus, the condition is external. Other externalist epistemologies, concerned perhaps with Gettier-type cases, have $C_i(\Phi)$ much like the internalist's justification, something like "Karl has a strong, undefeated, internally accessible warrant to believe Φ." The added externalist clause C_e might be something like: "there are no truths that would sufficiently undermine Karl's warrant for Φ, were he to become aware of them." To invoke a standard example, suppose that Φ is "there is a barn here," and Karl's warrant is perception: he sees something that looks like a barn and there really is one before him causing the perception. Now suppose that there are a lot of fake barns in the vicinity. Then if Karl discovers this, his warrant for Φ would be undermined (even though Φ is true). Anyone who has good reason to believe that there are a lot of fake barns in the vicinity would see that Karl's perception is not a sufficient warrant for Φ in this case. So the external condition $C_e(\Phi)$ would hold only if there are not a lot of fake barns in the vicinity. Similarly, $C_e(\Phi)$ would entail that Karl is not dreaming, that no one placed LSD in his morning coffee, and that he has not been secretly transported to a twin earth, much like our own environment except .... Again, Karl cannot determine, by reasoning and introspection alone, that he is not dreaming, not hallucinating, and not visiting twin earth.

Some issues in the epistemology of mathematics also fit our mold. Let $C_i(\Phi)$ be "Karl has a derivation of Φ in theory T." We can assume that (by introspection and reasoning) Karl can determine that a given derivation in T is indeed a derivation in T. An externalist condition might be that theory T is natural, intuitive, consistent, true, conservative, etc., depending on one's philosophy of mathematics. Given the incompleteness theorems, it is it is too much to demand that Karl have an internally accessible *proof* that T is consistent or conservative.

Returning to the general case, suppose that Karl knows Φ in world w, i.e., suppose that in w, Φ, $B\Phi$, $C_i(\Phi)$, $C_e(\Phi)$ all hold. The KK-thesis entails that by reasoning and introspection alone Karl can come to know that he knows Φ. So let w' be a world accessible from w in which Karl knows that he knows Φ. In w', then, Karl knows that $\Phi \wedge B\Phi \wedge C_i(\Phi) \wedge C_e(\Phi)$. Presumably, there is no problem with Karl knowing Φ in w', since he knows it in w, and there is no problem with Karl knowing that he believes Φ nor with him knowing $C_i(\Phi)$, since those are internally accessible. If there is a problem here, it is with Karl knowing $C_e(\Phi)$ in w'.

One proposal is that if other things are equal, then Karl's original warrant for Φ (i.e., $C_i(\Phi)$) is automatically a warrant for $C_e(\Phi)$. Suppose, for example, that $C_e(\Phi)$ is, or entails, that there are no facts about the world that would undermine Karl's warrant for Φ (if he were to become aware of them), a standard remedy for Gettier-type pathologies. Then one might argue that Karl's internal warrant for Φ (i.e., $C_i(\Phi)$) is *itself* a warrant (or includes a warrant) that there are no facts that seriously undermine this warrant. Plausibly, if Karl ever becomes aware that his warrant $C_i(\Phi)$ is undermined, then he does not know Φ (unless he obtains some other warrant). Thus, $C_i(\Phi)$ is a good warrant only if Karl has every reason to think that the warrant is not undermined. So Karl knows Φ only if he has every reason to believe (this aspect of) $C_e(\Phi)$. This is a quick and crude version of Crispin Wright's (Wright 1992*b*, pp. 48–57) more subtle argument concerning the KK-thesis for the notion he calls "superassertability." In present terms, the proposal is that an unchallenged $C_i(\Phi)$ is itself a warrant for $C_e(\Phi)$. Karl is warranted in assuming that the conditions are normal—that there are no fake barns around for example—so long as there is no evidence to the contrary.

This proposal seems to confute fallible warrants with externalist epistemology. On the view in question, the stated condition $C_e(\Phi)$ is not really external, since Karl can (defeasibly) determine that $C_e(\Phi)$ holds—by introspection alone. He just checks to see if there is no reason to doubt that the condition obtains. The proposal is that if other things are equal, then his internal warrant for Φ is itself a sufficient, but defeasible, warrant for $C_e(\Phi)$, and thus for the knowability of Φ. But this sufficient, defeasible warrant is in fact internally accessible, albeit fallibly. It might be better to think of the proposal as an internalism that accommodates fallibility. If Karl can know Φ fallibly, then he can know that he knows Φ with the same fallibility. Indeed, the proposal is that Karl can in fact come to know that he knows, with essentially the same warrant. What we are calling $C_e(\Phi)$ should be part of $C_i(\Phi)$.

There is little sense in quibbling over the labels "internalism" and "externalism," but on a straightforward externalism, we should be able to rule out Karl having an internal warrant for the external condition $C_e(\Phi)$ in any accessible world, for at least some propositions Φ. As we put it above, the feature that makes $C_e(\Phi)$ an *externalist* constraint is that Karl need not be aware that $C_e(\Phi)$ holds, and may not be able to determine, *fallibly or otherwise*, that $C_e(\Phi)$ holds by reasoning and introspection alone. On such a view, no amount of ordinary reasoning and introspection will produce a sufficient warrant for $C_e(\Phi)$. Plainly, a sufficient warrant for $C_e(\Phi)$ is just not internally accessible.

The externalist who maintains the KK-thesis must hold that, nevertheless, whenever Karl knows Φ (in a world w), he can come to know that $C_e(\Phi)$ holds: there is a world w' accessible from w in which Karl knows Φ. In particular, applying our scheme, there is an accessible world w' in which $C_e(\Phi)$, $C_i[C_e(\Phi)]$ and $C_e[C_e(\Phi)]$ all hold. Of course, the internal and external conditions for knowledge of $C_e(\Phi)$ may not be the same as those for knowledge of Φ itself (contra the above proposal that accommodates fallibility). Nevertheless, we just saw that the constraints that make $C_e(\Phi)$ *externalist* rule out Karl obtaining an internally accessible warrant for $C_e(\Phi)$. Recall that the only difference between w and w' is that in the latter, Karl does some reasoning and introspection. By hypothesis, introspection and reasoning are not sufficient to obtain a warrant for $C_e(\Phi)$.

So an externalist who maintains the KK-thesis must hold that $C_i[C_e(\Phi)]$ is nil. That is, Karl can know $C_e(\Phi)$ without having any internally accessible warrant for this belief. How is this possible? As far as I can tell, our externalist must regard $C_e(\Phi)$ as known, or knowable, by default—at least defeasibly. The slogan might be that $C_e(\Phi)$ is innocent until proven guilty. For example, suppose that $C_e(\Phi)$ is, in part, that Karl's perceptual mechanisms are in good working order. Once Karl realizes that he has no reason to think that his perceptual mechanisms are not functioning properly, then he *knows* that they are functioning properly. The case considered above is similar. Suppose that $C_e(\Phi)$ is, or entails, that there are no facts that would undermine Karl's (internal) warrant $C_i(\Phi)$ for Φ, were Karl to become aware of them. That is, Karl's warrant would remain good no matter how much further knowledge he obtains. The present proposal is that Karl knows this by default. If $C_i(\Phi)$ is indeed sufficient for knowledge of Φ (provided $C_e(\Phi)$ holds), then in the normal cases, there are no undermining facts. And if there are no undermining facts, then Karl knows this, albeit defeasibly.

One might claim that $C_e(\Phi)$ is a *presupposition* of the relevant knowledge. It is plausible, perhaps, to maintain that Karl knows Φ only if he knows, or can know, that the presupposition holds. It is more plausible to maintain that Karl knows that he knows Φ only if he knows that the presuppositions (for knowledge of Φ) obtain. The proposal in question is that Karl knows the latter by default. He knows the presupposition holds as long as he has no reason to suspect otherwise. That is, Karl's belief in $C_e(\Phi)$ is justified to the extent that he has no reason to suspect that $C_e(\Phi)$ is false. Of course, if Karl did have a reason to doubt $C_e(\Phi)$, this would contribute to undermining his knowledge of Φ. This is the way of presuppositions.

Perhaps Karl can determine by introspection and reasoning that he has no reason to doubt $C_e(\Phi)$. This may count as some sort of (defeasible) internal but *negative* warrant for $C_e(\Phi)$, and we might say that the external component $C_e(\Phi)$ for knowledge of Φ only has, and only needs, this negative warrant. But, again, we should not quibble over the labels, or whether negative warrants count as warrants. I leave it to the reader to determine if an introspectively determined belief that we have no reason to doubt $C_e(\Phi)$ itself counts as an internal warrant for $C_e(\Phi)$.

I am not sure if we can force mathematical knowledge into this framework. Recall that the idea there is that $C_i(\Phi)$ is something like "Karl has a derivation of Φ in theory T." The external condition $C_e(\Phi)$ is something like "theory T is natural, intuitive, consistent, true, conservative, etc." Again, in light of the second incompleteness theorem, it is presumably too much to demand that Karl have an internally accessible proof that T is consistent, true, conservative, etc. On the proposal considered here, we would say that Karl knows that T is natural, intuitive, consistent, true, conservative, etc., by default. Since consistency and conservativeness, at least, are themselves mathematical matters, the proposal runs against the slogan that the standard in mathematics is *proof*. We claim to know something—that a theory is consistent—but have no proof of this. But, of course, we can't prove everything. The alternatives seem to be to hold that Karl knows the consistency, etc. of T by default, or to hold that he does not, and cannot, know what he knows in mathematics, i.e., the KK-thesis fails. I will not speculate on which of these pills is the more difficult to swallow.

The conclusion, so far, is that the externalist can maintain the KK-thesis by arguing—or just asserting—that the external conditions $C_e(\Phi)$ for knowledge are knowable by default. Either Karl needs no warrant for $C_e(\Phi)$ at all, or else he just needs the negative warrant that he knows of no reason to doubt $C_e(\Phi)$. I see no other route to the KK-thesis for the externalist.

Like the above case of naive internalism, the foregoing treatment assumes that since Karl knows Φ if and only if $\Phi \wedge \mathrm{B}\Phi \wedge \mathrm{C}_i(\Phi) \wedge C_e(\Phi)$, then $\mathrm{KK}\Phi$ if and only if $\mathrm{K}[\Phi \wedge \mathrm{B}\Phi \wedge \mathrm{C}_i(\Phi) \wedge C_e(\Phi)]$. Recall that this is an instance of the inference scheme $\Phi \equiv \Psi \vDash \mathrm{K}\Phi \equiv \mathrm{K}\Psi$, which is not valid in general. The inference is good here, however, if Karl knows, or can know, *that* he knows Φ if and only if $(\Phi \wedge \mathrm{B}\Phi \wedge \mathrm{C}_i(\Phi) \wedge C_e(\Phi))$. That is, the above reasoning holds if Karl can come to know the correct analysis of knowledge by reasoning (or introspection). This is not wholly implausible—to the extent that epistemology is an a priori enterprise.

Recall that the final proposal is that Karl knows the external conditions $C_e(\Phi)$ by default. Accordingly, the only (internal) warrant for $\mathrm{K}\Phi$ that Karl needs is that the external conditions $C_e(\Phi)$ hold in normal cases, and that he has no reason to think that the present case is not normal. Perhaps Karl can obtain this negative warrant without knowing, in detail, what the external conditions are. That is, Karl might be able to obtain the negative warrant with only a general grasp of what normal conditions are. If the situation is in fact normal, then all Karl needs is a general belief that his warrant for Φ is not undermined.

In any case, we can get most of the way with a weaker inference. Suppose the above formula holds. To repeat:

$$\text{(Karl knows } \Phi) \text{ if and only if } (\Phi \wedge \mathrm{B}\Phi \wedge \mathrm{C}_i(\Phi) \wedge C_e(\Phi)).$$

Assume the KK-thesis. Let w be a world in which Karl knows Φ. Then there is a world w' accessible from w (i.e., by introspection and reasoning alone) in which Karl knows that he knows Φ. So, in w':

$$\text{Karl knows } (\mathrm{K}\Phi \wedge \mathrm{B}(\mathrm{K}\Phi) \wedge \mathrm{C}_i(\mathrm{K}\Phi) \wedge C_e(\mathrm{K}\Phi)).$$

Again, the internal and external conditions for $\mathrm{K}\Phi$ need not be the same as those for Φ itself. What we need is an inference in the following form:

> t is a warrant for P; $P \equiv Q \wedge R$ (by analysis); therefore t is a warrant for R (or at least t can be turned into a warrant for R if we add some reasoning and introspection).

It seems plausible to maintain that a warrant for $\mathrm{K}\Phi$ would have to be a warrant for $C_e(\Phi)$, since this last is a component of $\mathrm{K}\Phi$ (whether Karl knows this or not). By hypothesis, there is no internal warrant available for $C_e(\Phi)$. So, as with the foregoing proposal, we must conclude that $C_e(\Phi)$ does not need an internal warrant. It is knowable by default, or at best via a negative warrant.[1]

[1]This paper was written after conversations with Sarah Sawyer and Crispin Wright, although they may not think I have it right yet. Thanks also to the audience at the workshop “A Logical Approach to Philosophy,” in memory of Graham Solomon, at the University of Waterloo, May 9-10, 2003.

Chapter

TWO

CHOICE PRINCIPLES IN INTUITIONISTIC SET THEORY

John L. Bell

In intuitionistic set theory, the law of excluded middle is known to be derivable from the standard version of the Axiom of Choice that every family of nonempty sets has a choice function. In this paper it is shown that each of a number of intuitionistically invalid logical principles, including the law of excluded middle, is, in intuitionistic set theory, equivalent to a suitably weakened version of the Axiom of Choice. Thus these logical principles may be viewed as choice principles.

We work in intuitionistic Zermelo–Fraenkel set theory **IST** (for a presentation, see (Grayson 1979), where it is called $\mathbf{ZF}_1'$). Let us begin by fixing some notation. For each set A we write $\mathcal{P}(A)$ for the power set of A, and $\mathcal{Q}(X)$ for the set of *inhabited* subsets of A, that is, of subsets X of A for which $\exists x(x \in A)$. The set of functions from A to B is denoted by B^A; the class of functions with domain A is denoted by $\mathrm{Fun}(A)$. The empty set is denoted by 0, $\{0\}$ by 1, and $\{0,1\}$ by 2.

We tabulate the following *logical schemes*[1]:

[1]In addition to these logical schemes there is also the scheme—called in (Lawvere & Rosebrugh 2003) the *higher distributive law*—

HDDL $\forall x[\alpha(x) \vee \beta(x)] \rightarrow \exists x\alpha(x) \vee \forall x\beta(x)$

D. Devidi and T. Kenyon (eds.), A Logical Approach to Philosophy, 36-44.

SLEM $\alpha \vee \neg\alpha$ (α any sentence)

Lin $(\alpha \to \beta) \vee (\beta \to \alpha)$ (α, β any sentences)

Stone $\neg\alpha \vee \neg\neg\alpha$ (α any sentence)

Ex $\exists x[\exists\alpha(x) \to \alpha(x)]$ ($\alpha(x)$ any formula with at most x free)

Un $\exists x[\alpha(x) \to \forall x\alpha(x)]$ ($\alpha(x)$ any formula with at most x free)

Dis $\forall x[\alpha \vee \beta(x)] \to \alpha \vee \forall x\beta(x)$ (α any sentence, $\beta(x)$ any formula with at most x free)

Over intuitionistic logic, **Lin**, **Stone** and **Ex** are consequences of **SLEM**; and **Un** implies **Dis**. All of these schemes follow, of course, from the full law of excluded middle, that is **SLEM** for arbitrary formulas.

We formulate the following *choice principles*—here X is an arbitrary set and $\varphi(x, y)$ an arbitrary formula of the language of **IST** with at most the free variables x, y:

AC$_X$ $\forall x \in X\exists y\varphi(x, y) \to \exists f \in \mathrm{Fun}(X)\forall x \in X\varphi(x, fx)$

AC$_X^*$ $\exists f \in \mathrm{Fun}(X)[\forall x \in X\exists y\varphi(x, y) \to \forall x \in X\varphi(x, fx)]$

DAC$_X$ $\forall f \in \mathrm{Fun}(X)\exists x \in X\varphi(x, fx) \to \exists x \in X\forall y\varphi(x, y)$

DAC$_X^*$ $\exists f \in \mathrm{Fun}(X)[\exists x \in X\varphi(x, fx) \to \exists x \in X\forall y\varphi(x, y)]$

The first two of these are forms of the *Axiom of Choice* for X; while classically equivalent, in **IST** **AC**$_X^*$ implies **AC**$_X$, but not conversely. The principles **DAC**$_X$ and **DAC**$_X^*$ are dual forms of the Axiom of Choice for X: classically they are both equivalent to **AC**$_X$ and **AC**$_X^*$, but in **IST** **DAC**$_X^*$ implies **DAC**$_X$, and not conversely.

It is not difficult to show that, over intuitionistic predicate logic, **Un** implies **HDDL** which in turn implies **Dis**. As is shown below, both **Un** and **Dis** are equivalent to choice principles; however, I have not been able to show that the same is true of **HDDL**.

We also formulate what we shall call the *weak extensional selection principle*, in which $\alpha(x)$ and $\beta(x)$ are any formulas with at most the variable x free:

WESP $\exists x \in 2\alpha(x) \wedge \exists x \in 2\beta(x) \to \exists x \in 2\exists y \in 2[\alpha(x) \wedge \beta(y) \wedge [\forall x \in 2[\alpha(x) \leftrightarrow \beta(x)] \to x = y]]$.

This principle asserts that, for any pair of instantiated properties of members of 2, instances may be assigned to the properties in a manner that depends just on their extensions. **WESP** is a straightforward consequence of $\mathbf{AC}_{\mathcal{Q}(2)}$. For taking $\varphi(u, y)$ to be $y \in u$ in $\mathbf{AC}_{\mathcal{Q}(2)}$ yields the existence of a function f with domain $\mathcal{Q}(2)$ such that $fu \in u$ for every $u \in \mathcal{Q}(2)$. Given formulas $\alpha(x)$, $\beta(x)$, and assuming the antecedent of **WESP**, the sets $U = \{x \in 2 : \alpha(x)\}$ and $V = \{x \in 2 : \beta(x)\}$ are members of $\mathcal{Q}(2)$, so that $a = fU \in U$, and $b = fV \in V$, whence $\alpha(a)$ and $\beta(b)$. Also, if $\forall x \in 2[\alpha(x) \leftrightarrow \beta(x)]$, then $U = V$, whence $a = b$; it follows then that the consequent of **WESP** holds.

We are going to show that each of the logical principles tabulated above is equivalent (over **IST**) to a choice principle. Starting at the top of the list, we have first:

PROPOSITION 2.1 **WESP** *and* **SLEM** *are equivalent over* **IST**.

Proof. Assume **WESP**. Let σ be any sentence and define

$$\alpha(x) \equiv x = 0 \vee \sigma \qquad \beta(x) \equiv x = 1 \vee \sigma.$$

With these instances of α and β the antecedent of **WESP** is clearly satisfied, so that there exist members a, b of 2 for which (1) $\alpha(a) \wedge \beta(b)$ and (2) $\forall x \in 2[a(x) \leftrightarrow b(x)] \to a = b$. It follows from (1) that $\sigma \vee (a = 0 \wedge b = 1)$, whence (3) $\sigma \vee a \neq b$. And since clearly $\sigma \to \forall x \in 2[\alpha(x) \leftrightarrow \beta(x)]$ we deduce from (2) that $\sigma \to a = b$, whence $a \neq b \to \neg\sigma$. Putting this last together with (3) yields $\sigma \vee \neg\sigma$, and **SLEM** follows.

For the converse, we argue informally. Suppose that **SLEM** holds. Assuming the antecedent of **WESP**, choose $a \in 2$ for which $\alpha(a)$. Now (using **SLEM**) define an element $b \in 2$ as follows. If $\forall x \in 2[a(x) \leftrightarrow b(x)]$ holds, let $b = a$; if not, choose b so that $\beta(b)$. It is now easy to see that a and b satisfy $\alpha(a) \wedge \beta(b) \wedge [\forall x \in 2[a(x) \leftrightarrow \beta(x)] \rightarrow a = b]$. **WESP** follows. $\square$

REMARK 2.1 The argument for **WESP** → **SLEM** is another "stripped down" version of Diaconescu's theorem that, in a topos, the axiom of choice implies the law of excluded middle. The result may be compared with that of (Bell 1993*a*) to the effect that the presence of extensional ε–terms renders intuitionistic logic classical.

Next, we observe that, while $\mathbf{AC}_1$ is (trivially) provable in **IST**, by contrast

PROPOSITION 2.2 $\mathbf{AC}_1^*$ *and* **Ex** *are equivalent over* **IST**.

Proof. Assuming $\mathbf{AC}_1^*$, take $\varphi(x, y) \equiv \alpha(y)$ in its antecedent. This yields an $f \in \text{Fun}(1)$ for which $\forall y \alpha(y) \rightarrow \alpha(f0)$, giving $\exists y[\exists y \alpha(y) \rightarrow \alpha(y)]$, i.e., **Ex**.

Conversely, define $\alpha(y) \equiv \varphi(0, y)$. Then, assuming **Ex**, there is b for which $\exists y \alpha(y) \rightarrow \alpha(b)$, so $\forall x \in 1 \exists y \varphi(x, y) \rightarrow \forall x \in 1 \varphi(x, b)$. Defining $f \in \text{Fun}(1)$ by $f = \{\langle 0, b\rangle\}$ gives $\forall x \in 1 \exists y \varphi(x, y) \rightarrow \forall x \in 1 \varphi(x, fx)$, and $\mathbf{AC}_1^*$ follows. $\square$

Further, while $\mathbf{DAC}_1$ is easily seen to be provable in **IST**, we have

PROPOSITION 2.3 $\mathbf{DAC}_1^*$ *and* **Un** *are equivalent over* **IST**.

Proof. Given α, define $\varphi(x, y) \equiv \alpha(y)$. Then, for $f \in \text{Fun}(1)$, $\exists x \in 1 \varphi(x, fx) \leftrightarrow \alpha(f0)$ and $\exists x \in 1 \forall y \varphi(x, y) \leftrightarrow \forall y \alpha(y)$. $\mathbf{DAC}_1^*$ then gives

$$\exists f \in \text{Fun}(1)[\alpha(f0) \rightarrow \forall y \alpha(y)],$$

from which **Un** follows easily.

Conversely, given φ, define $\alpha(y) \equiv \varphi(0, y)$. Then from **Un** we infer that there exists b for which $\alpha(b) \rightarrow \forall y \alpha(y)$, i.e. $\varphi(0,b) \to \forall y\varphi(0,y)$. Defining $f \in \mathrm{Fun}(1)$ by $f = \{\langle 0,b\rangle\}$ then gives $\varphi(0,f0) \to \exists x \in 1\forall y\varphi(x,y)$, whence $\exists x \in 1\varphi(x,fx) \to \exists x \in 1\forall y\varphi(x,y)$, and **Un** follows. □

Next, while $\mathbf{AC}_2$ is easily proved in **IST**, by contrast we have

PROPOSITION 2.4 $\mathbf{DAC}_2$ *and* **Dis** *are equivalent over* **IST**.

Proof. The antecedent of $\mathbf{DAC}_2$ is equivalent to the assertion

$$\forall f \in \mathrm{Fun}(2)[\varphi(0,f0) \vee \varphi(1,f1)],$$

which, in view of the natural correlation between members of Fun(2) and ordered pairs, is equivalent to the assertion

$$\forall y\forall y'[\varphi(0,y) \vee \varphi(1,y')].$$

The consequent of $\mathbf{DAC}_2$ is equivalent to the assertion

$$\forall y \in Y\varphi(0,y) \vee \forall y' \in Y\varphi(1,y').$$

So $\mathbf{DAC}_2$ itself is equivalent to

$$\forall y\forall y'[\varphi(0,y) \vee \varphi(1,y')] \to \forall y\varphi(0,y) \vee \forall y'\varphi(1,y').$$

But this is obviously equivalent to the scheme

$$\forall y\forall y'[\alpha(y) \vee \beta(y')] \to \forall y\alpha(y) \vee \forall y'\beta(y'),$$

where y does not occur free in β, nor y' in α. And this last is easily seen to be equivalent to **Dis**. □

Now consider $\mathbf{DAC}_2^*$. This is quickly seen to be equivalent to the assertion

$$\exists z\exists z'[\varphi(0,z) \vee \varphi(1,z') \to \forall y\varphi(0,y) \vee \forall y'\varphi(1,y')],$$

i.e. to the assertion, for arbitrary $\alpha(x)$, $\beta(x)$, that

$$\exists z\exists z'[\alpha(z) \vee \beta(z') \to \forall y\alpha(y) \vee \forall y'\beta(y')].$$

This is in turn equivalent to the assertion, for any sentence α,

$$\exists y[\alpha \vee \beta(y) \to \alpha \vee \forall y \beta(y)]. \tag{$*$}$$

Now $(*)$ obviously entails **Un**. Conversely, given **Un**, there is b for which $\beta(b) \to \forall y \beta(y)$. Hence $\alpha \vee \beta(b) \to \alpha \vee \forall y \beta(y)$, whence $(*)$. So we have shown that

PROPOSITION 2.5 *Over* **IST**, ***DAC***$_2^*$ *is equivalent to* **Un**, *and hence also to* **DAC**$_1^*$.

In order to provide choice schemes equivalent to **Lin** and **Stone** we introduce

ac$_X^*$ $\exists f \in 2^X[\forall x \in X \exists y \in 2\varphi(x, y) \to \forall x \in X\varphi(x, fx)]$.

wac$_X^*$ $\exists f \in 2^X[\forall x \in X \exists y \in 2\varphi(x, y) \to \forall x \in X\varphi(x, fx)]$ provided $\vdash_{\mathbf{IST}} \forall x[\varphi(x, 0) \to \neg\varphi(x, 1)]$.

Clearly **ac**$_X^*$ is equivalent to

$$\exists f \in 2^X[\forall x \in X[\varphi(x, 0) \vee \varphi(x, 1)] \to \forall x \in X\varphi(x, fx)]$$

and similarly for **wac**$_X^*$.

Then

PROPOSITION 2.6 *Over* **IST**, **ac**$_1^*$ *and* **wac**$_1^*$ *are equivalent, respectively, to* **Lin** *and* **Stone**.

Proof. Let α and β be sentences, and define $\varphi(x, y) \equiv x = 0 \wedge [(y = 0 \wedge \alpha) \vee (y = 1 \wedge b)]$. Then $\alpha \leftrightarrow \varphi(0, 0)$ and $\beta \leftrightarrow \varphi(0, 1)$, and so $\forall x \in 1[\varphi(x, 0) \vee \varphi(x, 1)] \leftrightarrow \varphi(0, 0) \vee \varphi(0, 1) \leftrightarrow \alpha \vee \beta$. Therefore

$$\begin{aligned}
\exists f \in 2^1 \quad [\forall x \in 1 \quad & [\varphi(x, 0) \vee \varphi(x, 1)] \to \forall x \in 1\varphi(x, fx)] \\
\leftrightarrow \quad & \exists f \in 2^1[\alpha \vee \beta \to \varphi(0, f0)] \\
\leftrightarrow \quad & [\alpha \vee \beta \to \varphi(0, 0)] \vee [\alpha \vee \beta \to \varphi(0, 1)] \\
\leftrightarrow \quad & [\alpha \vee \beta \to \alpha] \vee [\alpha \vee \beta \to \beta] \\
\leftrightarrow \quad & \beta \to \alpha \vee \alpha \to \beta.
\end{aligned}$$

This yields $\mathbf{ac}_1^* \to \mathbf{Lin}$. For the converse, define $\alpha \equiv \varphi(0,0)$ and $\beta \equiv \varphi(0,1)$ and reverse the argument.

To establish the second stated equivalence, notice that, when $\varphi(x,y)$ is defined as above, but with β replaced by $\neg\alpha$, it satisfies the provisions imposed in $\mathbf{wac}_1^*$. As above, that principle gives $(\neg\alpha \to \alpha) \vee (\alpha \to \neg\alpha)$, that is, $\neg\alpha \vee \neg\neg\alpha$. So **Stone** follows from $\mathbf{wac}_1^*$. Conversely, suppose that φ meets the condition imposed in $\mathbf{wac}_1^*$. Then from $\varphi(0,0) \to \neg\varphi(0,1)$ we deduce $\neg\neg\varphi(0,0) \to \neg\varphi(0,1)$; now, assuming **Stone**, we have $\neg\varphi(0,0) \vee \neg\neg\varphi(0,0)$, whence $\neg\varphi(0,0) \vee \neg\varphi(0,1)$. Since $\neg\varphi(0,0) \to [\varphi(0,0) \to \varphi(0,1)]$ and $\neg\varphi(0,1) \to [\varphi(0,1) \to \varphi(0,0)]$ we deduce $[\varphi(0,0) \to \varphi(0,1)] \vee [\varphi(0,1) \to \varphi(0,0)]$. From the argument above it now follows that $\exists f \in 2^1[\forall x \in 1[\varphi(x,0) \vee \varphi(x,1)] \to \forall x \in 1\varphi(x,fx)]$. Accordingly $\mathbf{wac}_1^*$ is a consequence of **Stone**. $\square$

In conclusion, we show how certain of the principles we have introduced can be derived in the presence of *term-forming operators*.

The ε– and τ-operators are term-forming operators yielding, for formulas $\alpha(x)$, terms $\varepsilon_x\alpha$ and $\tau_x\alpha$ in which the variable x is no longer free; they are introduced in conjunction with the axioms—the ε– and τ*-schemes*:

$$\exists x\alpha(x) \to \alpha(\varepsilon_x\alpha) \qquad \alpha(\tau_x\alpha) \to \forall x\alpha(x).$$

It is an easy matter to derive **Un** from the τ-scheme when τ is merely allowed to act on formulas with at most one free variable. When τ's action is extended to formulas with two free variables, the τ-scheme applied in **IST** yields the full dual Axiom of Choice $\forall X\mathbf{DAC}_X$. For under these conditions we have, for any formula $\varphi(x,y)$,

$$\forall x \in X[\varphi(x,\tau_y\varphi(x,y)) \to \forall y\varphi(x,y)]. \qquad (**)$$

Let $t \in \mathrm{Fun}(X)$ be the map $x \mapsto \tau_y\varphi(x,y)$. Assuming that $\forall f \in Y^X\exists x \in X\varphi(x,fx)$, let $a \in X$ satisfy $\varphi(a,ta)$. We deduce from $(**)$ that $\forall y \in Y\varphi(a, y)$, whence $\exists x \in X\forall y \in Y\varphi(x, y)$. The dual Axiom of Choice follows.

In the case of the ε-operator, the number of free variables in the formulas on which the operator is allowed to act is an even more sensitive matter. If ε is allowed to act only on formulas with at most one free variable (so yielding only closed terms), the corresponding ε-scheme applied in **IST** is easily seen to yield both **Ex** and $\mathbf{ac}_1^*$, and so also **Lin**. But it is (in essence) shown in (Bell 1993*b*) that, if only closed ε-terms are admitted, **SLEM** is not derivable, and so therefore neither is **WESP**. The situation changes dramatically when ε is permitted to operate on formulas with two free variables. For then from the corresponding ε-scheme it is easy to derive $\mathbf{AC}_X$ for all sets X, and in particular $\mathbf{AC}_{Q2}$, and hence also **SLEM**.

I have found three ways of strengthening, or modifying, the single-variable ε-scheme so as to enable it to yield **SLEM**. The first, presented originally in (Bell 1993*a*), is to add to the ε-scheme Ackermann's Extensionality Principle, viz.,

$$\forall x[\alpha(x) \leftrightarrow \beta(x)] \rightarrow \varepsilon_x\alpha = \varepsilon_x\beta.$$

From these WESP is easily derived, and so, a fortiori, SLEM.

The second approach is to take the ε-axiom in the (classically equivalent) form

$$\alpha(\varepsilon_x\alpha) \vee \forall x\neg\alpha(x). \tag{†}$$

From this we can intuitionistically derive **SLEM** as follows: Given a sentence β, define $\alpha(x)$ to be the formula

$$(x = 0 \wedge \beta) \vee (x = 1 \wedge \neg\beta).$$

Then from (†) we get

$$[(\varepsilon_x\alpha = 0 \wedge \beta) \vee ([(\varepsilon_x\alpha = 1 \wedge \neg\beta)] \vee \\ \forall x\neg[(x = 0 \wedge \beta) \vee (x = 1 \wedge \neg\beta)],$$

which implies

$$[\beta \vee \neg\beta] \vee [\forall x\neg(x = 0 \wedge \beta) \wedge \forall x\neg(x = 1 \wedge \neg\beta)],$$

whence

$$[\beta \vee \neg\beta] \vee [\neg\beta \wedge \neg\neg\beta],$$

winding up with

$$\beta \vee \neg\beta.$$

The third method is to allow ε to act on pairs of formulas, each with a single free variable. Here, for each pair of formulas $\alpha(x)$, $\beta(x)$ we introduce the "relativized" ε-term $\varepsilon_x\alpha/\beta$ and the "relativized" ε-axioms

$$(1) \quad \exists x\beta(x) \rightarrow \beta(\varepsilon_x\alpha/\beta)$$
$$(2) \quad \exists x[\alpha(x) \wedge \beta(x)] \rightarrow \alpha(\varepsilon_x\alpha/\beta).$$

That is, $\varepsilon_x\alpha/\beta$ may be thought of as an individual that satisfies β if anything does, and which in addition satisfies α if anything satisfies both α and β. Notice that the usual ε-term $\varepsilon_x\alpha$ is then $\varepsilon x\alpha/x = x$. In the classical ε-calculus $\varepsilon_x\alpha/\beta$ may be defined by taking

$$\varepsilon_x\alpha/\beta = \quad \varepsilon_y[[y = e_x(\alpha \wedge \beta) \wedge \exists x(\alpha \wedge \beta)] \vee [y = \varepsilon_x\beta \wedge \neg\exists x(\alpha \wedge \beta)]].$$

But the relativized ε-scheme is not derivable in the intuitionistic ε-calculus since it can be shown to imply **SLEM**. To see this, given a formula γ define

$$\alpha(x) \equiv x = 1 \qquad \beta(x) \equiv x = 0 \vee \gamma.$$

Write a for $\varepsilon_x\alpha/\beta$. Then we certainly have $\exists x\beta(x)$, so (1) gives $\beta(a)$, i.e.

$$(3) \quad a = 0 \vee \gamma.$$

Also $\exists x(a \wedge \beta) \leftrightarrow \gamma$, so (2) gives $\gamma \rightarrow \alpha(a)$, i.e.

$$\gamma \rightarrow a = 1,$$

whence

$$a \neq 1 \rightarrow \neg\gamma,$$

so that

$$a = 0 \rightarrow \neg\gamma.$$

And the conjunction of this with (3) gives $\gamma \vee \neg\gamma$, as claimed.

Chapter
THREE

ASSERTION, PROOF, AND THE AXIOM OF CHOICE

David DeVidi

Graham Solomon was a Teaching Assistant in the first philosophy course I ever took, when he was a Master's student and I a first year undergraduate at Carleton University. My main memory of him at this time is this: during a small group tutorial session, I was waxing sceptical, making (no doubt aggravating) use of arguments I'd borrowed from the course lectures. In particular, I was trying to convince my classmates that I (and they) really couldn't be sure that a particular piece of chalk sitting on the table existed. *Pling*, the chalk bounced off the top of my head. "Does that clear things up a bit?" Graham asked.

A few years later, when I was beginning my Master's work at the University of Western Ontario, Graham was in the final stages of writing his doctorate. My one significant encounter with him that year was an evening that involved he, a fellow student named David McCallum—known that year as McAnimal—and I wandering the streets of London, Ontario, stopping here and there to consume a few too many refreshing beverages at certain of our favourite local businesses. This turned out to be an important night for me, as

D. Devidi and T. Kenyon (eds.), A Logical Approach to Philosophy, 45-76.

the two of them convinced me that Pynchon and Elias Cannetti were all very well, but I really did need to read Don DeLillo, Philip Roth, Joan Didion. Over the years consumption of refreshing beverages went way down for all of us, but Graham's literary advice continued to be a reliable guide for me when selecting replacement addictions. The guy read everything, and had great taste.

Our paths crossed again a few years later, when I was in the final stages of my own doctoral dissertation. I moved to Waterloo, Ontario, because my spouse had landed a good job there. I was writing a thesis in logic, but this was a mid-program change of plans, as I had originally intended to write a thesis on Carnap's *Logical Syntax of Language*. After completing his degree, Graham had taken at job at Wilfrid Laurier University in Waterloo. When he was back in London talking to his friend and former supervisor Bill Demopoulos about Carnap, Bill suggested that since I was in the same town and had been doing work along the same lines, the two of us should get together and talk *Syntax*. Graham and I soon figured out that we worked very well together, and we became frequent collaborators. We started by turning the paper that was the proto-thesis for my abandoned Carnap project into something publishable. We found working together so congenial that we turned to working on other projects. In the process, we became good friends. His stoical temperament helped make us a good team, where my job was to be the one likely to fly off the handle. Whether as a teacher, a mentor, a friend, or merely an advisor on literary matters, he was always someone with whom it was good to spend time. He is missed, and those who didn't get to meet him won't know what they missed.

1 Introduction

What I hope to do in this paper is to show that two seemingly quite separate but (I hope) independently interesting developments in philosophy are importantly related, and in particular that the second has important implications for evaluating the first.

The first topic is this: Among the many interesting and important new arguments in Timothy Williamson's *Knowledge and its Limits* (Williamson 2000) is one which he takes to undermine Michael Dummett's influential and oft-presented argument against truth-conditional theories of meaning and in favour of assertibility conditional meaning theories. Since this is a linchpin of Dummett's celebrated argument that anti-realism is, so to speak, the default position for particular domains of discussion, and that realism is something one must "win through to" by explaining what realist truth conditions can consist in for a particular domain, there are important consequences for contemporary metaphysical discussions if Williamson is right.

The second development arises in foundations of mathematics. By the early 1990s one could find in the mathematical logic literature two distinct research programs, each calling itself, at least sometimes, "Intuitionistic type theory" (e.g., (Martin-Löf 1984), (Bell 1993*b*)). Despite sharing a name, these programs have quite different philosophical starting points. One begins with the recognizably constructivist goals the "intuitionistic" in the name would lead you to expect, and has evolved into an explicit attempt to devise workable constructivist foundations for mathematics. The other approach evolves out of some of the least constructive reaches of mathematics, including algebraic topology, set theory, and the other fields behind the development of category theory and the isolation of the notion of an elementary topos. It turns out that the *logic* of elementary toposes, which is to say the system comprising the principles which are correct in every topos, is an intuitionistic type theory. But here the "intuitionistic" in the name doesn't indicate constructivist motivations, but the fact that the logic is recognizably akin to the formal logic developed by Arend Heyting which goes by the name "intuitionistic logic," but which has found a variety of (non-constructivist) philosophical and mathematical applications. The most striking contrast between these similarly named type theories is in the status each assigns to choice principles like the Axiom of Choice, as we shall see.

We proceed as follows. First, I will sketch Williamson's case against Dummett's argument for assertion theoretic semantics. In particular, I will attempt to draw out the assumptions which lead Williamson to conclude that any such semantics depends on the existence of "luminous" states. Secondly, after briefly introducing the two versions of intuitionistic type theory, I will rehearse some of the key points of the first, the Propositions as Types version of constructivism. This will put us in a position to argue, against Williamson, that assertion theoretic semantics isn't *dependent* on the existence of luminous states, or at least that the luminous states required are of a sort to which his argument fails to apply, and which he explicitly allows for. In the third section of the paper, though, we will contrast the Propositions as Types view with Local Set Theories, the intuitionistic type theories which arise in topos theory.[1] In particular, by contrasting what needs to be said about choice principles in the two sorts of type theories we will be able to point out some limitations an advocate of a strictly assertion theoretic semantics is liable to run up against, even within mathematics, which is supposed to be the ground where it has the best prospects for success. Thus we arrive at alternative reasons internal to a broadly constructivist view for supposing that this is too restrictive a perspective to provide an adequate theory of meaning even for mathematics.

2 Luminosity

Williamson's anti-Dummettian argument is an application of a general argument against what he calls "luminosity." He

[1]Much of this discussion will be drawn from the much more detailed investigation to be found in (DeVidi 2004).

notes the “constant temptation in philosophy to postulate a realm of phenomena in which nothing is hidden from us” (Williamson 2000, p. 92), and his argument is that it is always a mistake to give in to this temptation. He has in mind not just traditional “given” or “transparent” states (for instance, *being in pain*, or *having it appear to one that*), states about which some philosophers have supposed us to be infallible or omniscient. Rather, Williamson wants to consider also more sophisticated accounts where we *might be mistaken about the state* we are in if we are, for instance, deceived by ourselves or others, or *unaware of being in the state* if, for instance, we’re too busy with something else to reflect on the fact. He therefore considers “luminous” any state which is such that if you are in it, you *could* (in a suitable sense of “could”) know that you are in it. He offers as his official definition the following: A conditions C is *luminous* if and only if:

(∗) For every case α, if C obtains in α, then in α one *is in a position to know* that C obtains. (p. 95)

Being in pain and like mental states, as traditionally conceived, satisfy this definition, but so too would certain other states according to some authors. Williamson quotes Dummett giving an example:

> It is an undeniable feature of the notion of meaning—obscure as that notion is—that meaning is *transparent* in the sense that, if someone attaches a meaning to each of two words, he must know whether these two meanings are the same. ((Dummett 1978, p. 131), quoted at (Williamson 2000, p. 96))

Obviously, Dummett can’t mean by “transparent” anything like *infallible* or *omniscient*, since he’ll want to account for the fact that the question of whether two terms mean the same is very often going to be met with either verbal fumbling or a blank stare. Dummett presumably means that given sufficient careful reflection someone will be able to figure out whether the two terms mean the same.

Williamson's contention is that there are no *non-trivial* cases of conditions C which satisfy ($*$), i.e., there are no *interesting* luminous conditions. By *non-trivial*, Williamson means a condition which one is sometimes in, sometimes not, both in the sense that in some cases the condition obtains while in others not, and in the sense that one can *change* from being in the condition to not being in the condition. The claim is that any such condition cannot be luminous, for the supposition of luminosity joined with non-triviality reduces to absurdity. We consider the argument only briefly, since while the details are interesting, they are not directly relevant to the present project. (Since we are concerned with the consequences Williamson contends follow from this conclusion, not with his argument for it.)

Imagine, as we are supposing is possible, an agent undergoing a *change* from being in a state C to not being in it, and that this transition has been divided into finitely many stages. Since C is luminous, at every stage the agent is *in a position* to know that C obtains. Suppose that at every stage the agent is doing whatever is required to know it.

The next crucial premise is the following general claim:

(**I**) If at stage i one knows that C obtains, then at stage $i+1$ C obtains.

Williamson defends this claim for reasons having to do with what he takes it to *mean* to say that someone *knows* something. In particular, he takes it to follow from the requirement that to count as knowledge one's beliefs must be *reliably based*, coupled with the specific features of the description of this case. Again, we shall not pursue these details.

Of course, given the description of the case and principle (**I**), problems ensue. For since the agent is doing whatever is required at the first stage to know that C obtains, he knows that it does. It follows from (**I**) that the agent is in C, and we are supposing the agent to be still doing whatever is required to know that, as well. And so on. It follows that C obtains at every stage, and and we get a contradiction with the suppositition that this is a case of a transition from being in C to not being in C.

2.1 Anti-Anti-Realism

Williamson takes his conclusion to undermine the influential Dummettian case for an anti-realist theory of meaning. Dummett's argument is designed to show that realist meaning theories on which the meaning of a declarative sentence is to be accounted for in terms of its truth conditions are unacceptable. They should be replaced by theories on which the meaning of a sentence is to be given instead in terms of the conditions under which *assertion* of the statement would be *warranted*. Williamson argues that unless Dummett can appeal to (non-trivial) luminous conditions, his assertion-conditional semantics is in no better position than truth-conditional semantics to respond to the charge Dummett takes to be fatal to the latter. Since if Williamson is right the preceding argument shows that there are no non-trivial luminous conditions, assertion conditional meaning theories are no better off than truth-conditional ones. Moreover, any X-conditional semantics where the X does not appeal to a luminous condition will fall prey to Dummett's argument against truth-conditional semantics. So, presumably, the problem is with *the requirement* Dummett complains truth-conditional semantics fail to meet—since no other sort of meaning theory could meet it either—and *not with the various meaning theories which fail to meet it.*

What is the Dummettian complaint against realist versions of truth conditional theories of meaning? It begins from the suggestion that a theory of meaning for some part of a language must explain what is *understood* by a competent speaker of that bit of language. And understanding implies *knowing the meaning*. If meanings are truth conditions, then one who understands, for instance, a sentence must know its truth conditions. What makes these *realist* truth conditions is that one allows that these truth conditions might transcend any possible evidence—the truth values of the sentences are fixed by a mind-independent reality quite apart from our intellection. This leaves open the possibility that for some such claims it is impossible, even in principle, to determine whether they hold or not. The problem for the realist is to explain in what a speaker's knowledge that a sentence had this sort of truth condition could consist. What does it mean, in particular, to know that a sentence has a particular truth condition rather than some alternative truth condition when there is no possible evidence which would distinguish the two cases? If anything *could* constitute such knowledge, it seems unlikely that it could be anything determined by the *use* of the relevant bits of language in a language community—breaking the bonds between *meaning* and *use* is, for Dummett, enough to render a proposed theory of meaning certainly beyond the pale.

Replacing truth conditions by assertibility conditions is supposed to avoid this problem. That is, the meaning of a declarative sentence, on this account, is determined by the conditions under which its assertion would be warranted. To understand the sentence therefore requires *knowing* the conditions under which its assertion would be warranted. The claim is that, unlike the realist who cannot explain what knowledge of truth conditions consists in, it is possible to explain in what the knowledge of assertibility conditions consists—roughly, in the disposition to assert the sentence when and only when assertion is warranted, other things being equal.

Williamson objects that the assertion theoretic account is better off than the realist semantics *only if assertibility conditions are luminous*. But if Williamson is right, the argument against luminosity extends to assertibility conditions, so the assertibility theorist is in no better position than the realist to meet Dummett's challenge. The upshot is that it is Dummett's requirement that is unacceptable—that is, if there are to be systematic meaning theories, they will be theories which do not satisfy the condition that understanding a sentence requires knowing its meaning, at least if knowing the meaning involves being able to recognize when the conditions determining the meaning obtain. If Williamson is right, this opens up room for semantic theories which either reject the "meaning is determined by use" mantra, as perhaps certain causal theorists might be happy to do, or at least construe it very differently than Dummett does.[2]

3 Two Flavours of Type Theory

We now shift gears to consider the second topic of interest, namely the two versions of intuitionistic type theory. A useful place to begin is with something curious that happens in W.W. Tait's searching and illuminating investigation of some fundamental concepts of constructive mathematics in his 1994 paper "The Law of Excluded Middle and the Axiom of Choice" (Tait 1994). Tait argues that the Axiom of Choice "is indeed a law of logic," at least from the *construction theoretic* point of view he regards as the correct one for foundations of mathematics. Given this view of the Axiom of Choice, it is not surprising that he views the extensive 20th Century debate on the status of that principle as largely misguided. He argues that it is actually the *law of excluded middle* (LEM) that lies behind various bits of philosophical awkwardness often attributed to the Axiom of Choice, and so it is Excluded Middle which ought to be rejected. Tait provides a sketch of a type theory which is supposed to capture his construction theoretic view, one in which, he claims, the Axiom of Choice holds, but the LEM need not.

For a classically-minded logician LEM is a principle of logic while the Axiom of Choice is notoriously controversial, so some classically-minded logicians will no doubt be surprised

[2]Williamson takes the latter course in at least some places. For instance, in some of his defenses of the claim that vague predicates have precise boundaries which we are in principle incapable of identifying, he holds to the view that meaning is determined by use while also saying things like "Meaning may supervene on use in an unsurveyably chaotic way." (Williamson 1997, p. 175).

by Tait's inversion of the status of the principles, which is presumably what Tait intended. But another group of logicians were likely to be surprised by Tait's claim, a group that also works with a sort of type theories, sometimes called "local set theories." Local set theories are the internal languages of toposes, and one of the most interesting results in topos theory is Diaconescu's theorem, which is often stated as follows: *the Axiom of Choice implies the Law of Excluded Middle*.

Tait's paper illustrates something interesting that seems to have been happening in the early 1990s. There seem to have been two distinct research programs on the go, both sometimes traveling under the name "intuitionistic type theory" (compare, e.g., the titles of (Martin-Löf 1984) and (Bell 1993*b*)), in one of which the Axiom of Choice is a valid *logical* principle, in the other of which it implies the law of excluded middle, and so all of classical logic (and so, obviously, cannot be generally valid if it's really *intuitionistic* type theory). Moreover, it seems each group was only vaguely aware of the work of the other, as is witnessed by the fact that someone of Tait's stature would write a paper called "The Law of Excluded Middle and the Axiom of Choice" without so much as mentioning Diaconescu's result.

The type theory Tait describes is an example of what we might call a *Martin-Löf type theory*, or as we shall say here, a *Propositions as Types Theory* (something we'll often abbreviate to PAT).[3]

[3] 'Martin-Löf type theory' is something of a loosely defined notion, not least because Martin-Löf has proposed more than one type theory. Moreover, this sort of type theory is part of an ongoing research program, and so is subject to modifications to make it suitable for a variety of purposes, many having to do with making it more suitable as a programming language in computer science. See the discussion in (Thompson 1991), esp. Ch. 7 and 9 and (Bridges & Reeves 1999). A defining feature of this sort of type theory is its commitment to the "Propositions as Types" view, sometimes called the *Curry-Howard Isomorphism*. See (Howard 1980). An easy to read introduction to the rudiments, as they are understood by the originator, is supplied in (Martin-Löf 1984). Rather than trying to deal with a somewhat amorphous class in a general way, I will usually deal with the specific system sketched by Tait (with occasional references

What makes PAT theories relevant to present concerns is that the Propositions as Types view, in its various guises, represents the state of the art for systematic and sustained attempts to work out the consequences of ideas central to the constructivist semantics that characterize anti-realist meaning theories for Dummett and Williamson. For the central idea is that a mathematical proposition is to be identified with "the type of its proofs."—It is true exactly if there is an object of the appropriate type, which is to say if and only if there is a proof for the proposition.—So truth amounts to provability. But this is not a matter of stipulation, a decision to restrict what we shall be willing to count as a true mathematical proposition. Rather, it is supposed to follow from an analysis of what the meanings of mathematical claims could consist in. Martin-Löf is particularly adamant about distinguishing mathematical *propositions* from mathematical *judgements*. Propositions are types, which is to say (constructable) syntactical objects on which further mathematical operations may be performed. Judgements, on the other hand, are actions by thinking subjects, and involve the mentalistic notions of *understanding* and *grasping the meaning of*. In particular, if **A** is a proposition, "**A** is a proposition" and "**A** is a true proposition" are both judgements. The former judgement is warranted only when one knows what would count as a (canonical) proof of **A**, while the latter would be warranted only when one knows how to find such a proof. Thus, as Tait puts the point, "the force of the identification of truth with provability is simply that the only warrant for asserting **A** is a proof of **A**" (Tait 1994, p. 52).

We will therefore briefly sketch some of the rudiments of the Propositions as Types views. Next we consider how one

to (Martin-Löf 1984)), assuming that it will serve as a useful, philosophically motivated exemplar of the class. Note that I do not mean to suggest that Tait was inspired by, or that his views are somehow derivative from Martin-Löf's, as Tait has been a leading investigator in this field for a long time. The name 'Martin-Löf type theory' simply seems to have gained a certain currency in the literature. See, for instance, the historical discussions in the popular (Lambek & Scott 1986).

argues, from this basis, that the Axiom of Choice is a valid logical principle. For it is in fact quite commonly argued that there are good constructive reasons for accepting all sorts of choice principles, having to do with the constructivist interpretation of the quantifiers, especially the existential quantifier. Some of these arguments are better than others. But it will be worth our while to get one such argument on the table for when we return later to contrast PAT theories with the Local Set Theories in which the "the Axiom of Choice" has dramatically non-constructive consequences.

3.1 Propositions as Types

We will need to introduce some notation before we can talk sensibly about these matters. We will introduce it in bits as they become necessary, though we will continue to be informal. We will for the most part follow Tait's terminology, though we will alter his notation somewhat.

For convenience we will assume that each type $\mathbf{A}$ has its own stock of variables, $x_{\mathbf{A}}, y_{\mathbf{A}}, z_{\mathbf{A}}, \ldots$, so that, e.g., $\forall x_{\mathbf{A}}$ means 'for all objects x of type $\mathbf{A}$.' However, for convenience we will omit subscripts when the context makes clear what the type of a variable is.

In a *Propositions as Types* view, as the name suggests, one identifies *propositions* with *types*. As one might expect of a constructivist view, a *warrant* for asserting a proposition must be a *proof* of the proposition, which is in turn to be indentified with an *object* of that type—that is, just as the type and the proposition might be thought of as two ways of looking at the same thing, the objects of the type and the proofs of the proposition are two ways of looking at a single object (i.e., construction).

This, in turn, gives rise to a correspondence between logical operations on propositions and mathematical operations on types. For example, the obvious correspondence between proofs c of $\mathbf{A} \wedge \mathbf{B}$ and pairs $\langle a, b\rangle$, where a is a proof of $\mathbf{A}$ and b a proof of $\mathbf{B}$, shows that we can identify the proposition

$\mathbf{A} \wedge \mathbf{B}$ with the product $\mathbf{A} \times \mathbf{B}$ of the types **A** and **B**. Similarly, given the constructive understanding of $\vee$, a proof of $\mathbf{A} \vee \mathbf{B}$ is either a proof of **A** or a proof of **B** along with the information that it proves either the first or the second disjunct, and that a proof of either disjunct yields a proof of the disjunction. We can therefore identify $\mathbf{A} \vee \mathbf{B}$ with the disjoint union (i.e., coproduct) $\mathbf{A} + \mathbf{B}$.

A proof of $\mathbf{A} \to \mathbf{B}$, from a constructive point of view, is a method which, given a proof of **A**, yields a proof of **B**. So from the present point of view a proof f of $\mathbf{A} \to \mathbf{B}$ is a *function* from **A** to **B**. That conditional proposition is therefore identified with the type $\mathbf{A} \Rightarrow \mathbf{B}$ (also written $\mathbf{B}^{\mathbf{A}}$) of functions from **A** to **B**. We postulate the existence of an absurd proposition $\bot$, which is of course unprovable. We therefore identify $\bot$ with the empty type $\varnothing$. We then define, in the time honoured manner of intuitionistic logic, $\neg\mathbf{A}$ with $\mathbf{A} \to \bot$, and so with the type $\mathbf{A} \Rightarrow \varnothing$.

Finally, we consider the quantifiers. The constructivist understanding of $\forall$ parallels the understanding of $\to$ (whereas the classical logician is more likely to describe a parallel to $\wedge$). To supply a proof of $\forall x_{\mathbf{A}}\Phi(x)$ is to give a method by which one can supply, for any given $x_{\mathbf{A}}$, a proof of $\Phi(x)$. In the Propositions as Types systems, this is taken to mean that such a proof is a function f with domain **A** such that, for each $x_{\mathbf{A}}$, $f(x)$ is of type $\Phi(x)$. So $\forall x_{\mathbf{A}}\Phi(x)$ is the type comprising all such functions, and so is (isomorphic to) the product $\prod_{x_{\mathbf{A}}} \Phi(x)$. As for the existential quantifier, a proof of $\exists x_{\mathbf{A}}\Phi(x)$ has two parts, an object c of type **A** and a proof of $\Phi(c)$, i.e., an object of type $\Phi(c)$, while any such pair yields a proof of $\exists x_{\mathbf{A}}\Phi(x)$. Thus $\exists x_{\mathbf{A}}\Phi(c)$ is the type of all such pairs, which is (isomorphic to) the coproduct (i.e. disjoint union) $\sum_{x_{\mathbf{A}}} \Phi(x)$.

We can now formulate the principle called the Axiom of Choice in the Propositions as Types arrangement:

$$\forall x_{\mathbf{A}}\exists y_{\mathbf{B}}\Phi(x, y) \to \exists f_{\mathbf{A}\Rightarrow\mathbf{B}}\forall x_{\mathbf{A}}\Phi(x, f(x)). \qquad \text{(AC)}$$

AC is *provable*. To see this, first note that the operation of taking two given individuals a and b and forming the ordered pair $\langle a, b\rangle$, and the operation of "projecting out" the components a and b of a given an ordered pair $\langle a, b\rangle$, are obviously constructively acceptable and so are assumed to be present in the system. (This amounts to supposing that given types A and B, there is a product type $A \times B$, which as we have seen in the present setup amounts to supposing we have the operations of forming and simplifying conjunctions of propositions.) We will write π_1 and π_2 for the projection operations.

So, if u is a proof of $\forall x_{\mathbf{A}} \exists y_{\mathbf{B}} \Phi(x, y)$, then for each $x_{\mathbf{A}}$, $u(x)$ is of type $\exists y_{\mathbf{B}} \Phi(x, y)$, so $u(x)$ is a pair $\langle b, c\rangle$ with b_B and c a proof of $\Phi(x, b)$. So $\pi_1(u(x))$ is of type $\mathbf{B}$ and $\pi_2(u(x))$ is of type (i.e., is a proof of) $\Phi(x, \pi_1(u(x)))$. So we have a function

$$x \mapsto \pi_1(u(x))$$

of type $\mathbf{A} \Rightarrow \mathbf{B}$, call it $s(u)$. Furthermore, the function

$$x \mapsto \pi_2(u(x))$$

is a proof of $\forall x_{\mathbf{A}} \Phi(x, \pi_1(u(x)))$. Call this function $t(u)$. Then the function

$$u \mapsto \langle s(u), t(u)\rangle$$

is a proof of (AC).

4 The Problems with PAT: Luminosity?

As we have seen, Martin-Löf, in particular, enforces a distinction between judgements and propositions which seems to put him squarely in Williamson's sights. For he is quite explicit about requiring that all judgements be accompanied by a full *explanation of their meanings*, and he requires that the judgement "**A** is a proposition" is warranted only if one knows what a canonical proof of **A** would be. If, as Williamson frequently says, judgement is merely the internal counterpart of assertion, we have here, essentially, an assertion theoretic meaning theory. Does it commit Martin-Löf to luminosity? I shall argue that if any sort of luminosity is involved, it is of the sort Williamson allows for, viz *trivial* luminosity.

The crux of the matter seems to me to be naturally captured in this remark about Williamson's preferred account of assertibility, which is that one is warranted in asserting p if and only if one *knows* that p:

> The knowledge account therefore implies that we are not always in a position to know whether we have warrant to assert p [because we don't always know that we know p]. We are liable to error and ignorance about warrant, just as we are about everything else. This view of warranted assertibility is in sharp contrast with its treatment in anti-realist theories of meaning to which the notion of assertibility conditions of sentences is crucial. Such theories characteristically assume that one has no difficulty in knowing whether one has warrant to assert p. Independently of the knowledge account, there is reason to doubt that there could be a norm of the kind postulated by anti-realist theories. (Williamson 2000, p. 258)

The independent reason is, of course, the anti-luminosity argument. And the crux of the matter is the liability to error and ignorance about warrant that is allegedly available to Williamson's view, but not to the anti-realist.

Where does the room to allow for error and ignorance about warrant come from for the knowledge account?

> …to a first approximation, in mathematics one has warrant to assert p if and only if one has a proof of p. On the knowledge account, that is so because, to a first approximation, in mathematics one knows that p if and only if one has a proof of p. …Those are just first approximations, but where

> having warrant to assert p diverges from having a proof of p, so does knowing p. Conversely, where knowing p diverges from having a proof of p, so does having a warrant to assert p. Having warrant to assert p and knowing p do not diverge from each other; the knowledge account is confirmed. (p.263)

In other words, it is because there are possible circumstances in which we can be in possession of a proof *without* having warrant to assert p that the knowledge account is to be preferred. For in these cases, Williamson contends, we also possess a proof without possessing mathematical knowledge, but there are no possible situations in which we have knowledge of p without having warrant to assert p.

How could we have a proof of p without having either warrant to assert p or knowledge of p? The key to seeing how this can be so, says Williamson, is to recognize a distinction between *defeasibility* of a warrant for p and *ceasing to be a proof of* p: new information can take away one's *warrant* to make an assertion without turning a proof of p into a non-proof. "One can have warrant to assert a mathematical proposition by grasping a proof of it, and then cease to have a warrant to assert it merely in virtue of gaining new evidence about expert mathematicians' utterances, without forgetting anything" (p. 265). That is, if one hears many better mathematicians than oneself express scepticism about whether an alleged proof of p (which is, perhaps, very long and complicated) is correct, then this might be enough to remove one's warrant to assert p, by removing one's *knowledge* that p, even if the proof is correct and if one has grasped it.

What is unclear, however, is whether Williamson provides grounds to think that someone like Martin-Löf or Tait cannot account for these phenomena. Tait says about the relationship between assertibility and warrant: "the only warrant for asserting **A** is a proof of **A**." This statement doesn't obviously commit him to the claim that assertion of **A** is warranted *if and only if* one has a proof of **A**, so it's not obvious that the sort of case Williamson describes needs to be a problem for him, even if we take this to be an accurate statement of his considered view. More importantly, though, the equivalence of assertibility and possession of a proof is likely to be a "first approximation" for a constructivist no less than for Williamson. Martin-Löf's standard for when it is warranted to judge "**A** is a true proposition" seems a more accurate statement of the view of assertibility conditions for a constructivist: one can do so *when one knows how to find a canonical proof of* **A**. This is well known to allow the possession of a *non-canonical* proof to count as grounds for assertion, where a non-canonical proof amounts, roughly, to a demonstration that a canonical proof exists. But it surely also allows for garden variety "knowing how to find a proof": an undergraduate sees a theorem stated and proved in a textbook, and says to a friend, "Golly, Bob, there are infinitely many prime numbers. Did you know that?" Surely her assertion was warranted, in spite of her having read the theorem but not the proof. And if knowledge implies belief, as Williamson agrees, one might well not *know* how to find a proof of **A** even if the proof is sitting directly in front of you, if the testimony of other mathematicians is enough to cause you to no longer believe that it is a proper proof of **A**.

In short, then, the "knows how to find a canonical proof of **A**" standard of assertibility is much closer to the knowledge account of assertibility Williamson touts. Indeed, the obvious move of equating mathematical knowledge with "knowing how to find a canonical proof," which might be a first approximation of a view a constructivist might be happy to defend, makes it a *version* of Williamson's view, though admittedly with a story about mathematical knowledge of which he probably wouldn't approve. Moreover, it allows us to be "liable to error and ignorance about warrant, just as we are about everything else." For we can be mistaken about whether we know how to find a proof even if the proof lies in front of us, if circumstances conspire against us. There's no obvious reason the constructivist is committed to luminosity here.

But perhaps one needs to beat the bushes a little more to find where the commitment to luminosity is hiding. The Propositions as Types literature is full of subtle distinctions, one of which is between basic and non-basic sets. A *basic set*, in the words of (Bridges & Reeves 1999) is "a set for which no computation is necessary to demonstrate that an element belongs to it," so for instance one needn't prove that "0 is a natural number" before asserting it,[4] while for non-basic sets one may assert $x \in \mathbf{A}$ only after proving it. Does this betray a commitment to luminosity, at least for a class of atomic statements about these basic sets?

I think two things need to be said here. First, it's not obvious that *not requiring a proof* implies *impossible to be mistaken about whether one is warranted in asserting*. One might, for instance, mistakenly fail to recognize that the natural numbers are a basic set, and so suppose one needs to prove that "17 is a natural number" before asserting it, but have no idea what such a proof would look like. Secondly, even if it is the case that for these basic sets one wants to say something like "anyone who genuinely grasps the meanings of the words "natural number" and "17" is in a position to know, i.e., will recognize if he pays suitable attention, that "17 is a natural number" is true, and so defend some sort of luminosity for these claims, this is a sort of luminosity Williamson allows for.

> Is a condition that obtains in every case, the necessary condition, luminous too? It is luminous as presented in a simple tautological guise, if cases are restricted to those in which the subject has the concepts to formulate the tautology. It is not luminous as presented in the guise of an a posteriori necessity, or an unproved mathematical truth, or if the cases include some in which one lacks appropriate concepts. (p. 108)

[4]This claim needs qualifying: one needn't make such a demonstration if the elements are presented in canonical form. We leave aside the obvious qualifications that follow from this in what follows.

Williamson insists that this concession is harmless, that it will not provide anti-realists with a "cognitive home." But if what we have said above is correct, at least in constructivist mathematics, the anti-realist doesn't *need* such a home. And the limited sorts of luminosity they arguably commit themselves to are the sorts of stipulative claims that abound in mathematical practice on most anybody's story.

5 Problems with PAT: Is Choice Constructive?

We have seen how the PAT view leads naturally to the acceptance of a principle often called the Axiom of Choice. Tait presents essentially same proof we gave above. When doing so he says that by giving it is he is "making rigorous the argument previously given that Lebesgue should have accepted this principle" (p. 59) That argument, on page 49, runs as follows:

> First, if a proof of $[\exists y_{\mathbf{B}}\Psi(y)]$ must involve defining a witness b (i.e. such that $\Psi(b)$) and a proof of $\Psi(b)$, then a proof of the antecedent of AC:
>
> $$[\forall x_{\mathbf{A}}\exists y_{\mathbf{B}}\Phi(x,y)] \qquad (1)$$
>
> must yield a definition, for each x of type $\mathbf{A}$, of the corresponding witness y and a proof of $\Phi(x,y)$. For certainly a proof of (1) should yield, for each such x, a proof of $[\exists y_{\mathbf{B}}\Phi(x,y)]$. But then let $f(x)$ be the witness y so defined by this proof. So then we have a proof of $[\Phi(x,f(x))]$ as a function of x. Thus we have a means of transforming any proof of (1) into a definition of a function f of type $\mathbf{B}^{\mathbf{A}}$ and for each x, a proof of $[\Phi(x,f(x))]$. In other words, we have a means of transforming any proof of (1) into a definition of f and a proof of $[\forall x_{\mathbf{A}}\Phi(x,f(x))]$. We have defined a means of transforming an arbitrary proof of the antecedent

of AC into a proof of its conclusion:

$$\exists f_{\mathbf{A}\Rightarrow\mathbf{B}} \forall x_{\mathbf{A}} \Phi(x, f(x))].$$

But what more should be required of a proof of AC?

It seems to me that there is a subtle error in this reasoning if it is taken as an argument for the view that constructivists ought to accept the Axiom of Choice.[5] This is not to contend that there is anything wrong with the formal proof that AC is valid from the Propositions as Types point of view, but rather that the Propositions as Types reasoning does *not* precisely mirror analogous reasoning conducted in English.[6]

According to any constructivist understanding of the universal quantifier, the method which we must have if we have a proof of the antecedent of AC must indeed yield, for each $x_{\mathbf{A}}$, a pair consisting of a $y_{\mathbf{B}}$ and a proof of $\Phi(x, y)$. However, we so far do not have enough information to ensure that we can get an f of type $\mathbf{A} \Rightarrow \mathbf{B}$ by simply specifying that $f(x)$ is to be the term of type $\mathbf{B}$ supplied by applying the method at hand to x. $\mathbf{A} \Rightarrow \mathbf{B}$ is a type of *functions*, so such an f must be functional. In particular, then, if we can prove $x_{\mathbf{A}} = z_{\mathbf{A}}$, then we must also be able to prove $f(x) = f(z)$. And the constructivist reading of the universal quantifier does not by itself guarantee that *this* will be the case.

Tait is aware of this problem, and himself points out a good example: "[O]ne example is that for every real number x there is an integer y that is greater than x. In fact, we can constructively prove that, for any Cauchy sequence $[x]$, there is an integer y which is greater than (the real number represented by) x. But the proof does not yield the same y for sequences x and x' which represent the same real number"

[5]A more elaborate presentation of the argument given in this section and the next can be found in (DeVidi 2004).

[6]One finds very similar presentations of both the formal and informal versions of this argument, along with the claim that they present "the same idea," in (Martin-Löf 1984, p. 50).

(Tait 1994, p. 59). However, Tait argues that this "is not really a counterexample to our previous argument that AC is a theorem of constructive logic." The reason is that since "x is a Cauchy sequence" is not a constructively decidable property of numerical functions, there is no way to formulate the statement "for every Cauchy sequence there is an integer which is greater" in the language of Tait's type theory in a way which gives us an instance of the antecedent of AC. In short, in this, as in other purported counterexamples which might be offered, "the scope of x is not a type in our sense but rather is a set" (Tait 1994, p. 59).

There are a couple of things that need to be said about this. First, by 'sets' Tait means *collections of objects of a given type*, citing in support of this notion the contention, which he attributes to Gödel, that "the original notion of set is that of a set of objects of some type" (Tait 1994, p. 46). However, Tait is not consistent in this usage. For he also includes in his system, for each type **A**, the type $\mathcal{P}(\mathbf{A})$ of all subsets of the type **A**. Tait takes as a primitive type of his system the two valued type $\mathbf{2} = \{\top, \bot\}$. We then define $\mathcal{P}(\mathbf{A})$ to be the type of maps $\mathbf{A} \to \mathbf{2}$, that is, a subset of **A** is a map defined on the objects in **A** such that each such object gets either the value $\top$ or the value $\bot$. So, if f is of type $\mathcal{P}(\mathbf{A})$, then we define

$$x_{\mathbf{A}} \in f \text{ iff } f(x) = \top.$$

In short, this definition has precisely the same form as the familiar identification of subsets with classical characteristic functions, though, of course, in the constructive setting the population of such functions will be considerably sparser. Elements of $\mathcal{P}(\mathbf{A})$ are only going to be determined by *decidable* predicates $\Phi(x_{\mathbf{A}})$, so the scope of x in his example is not even going to be a *set* in this sense. Evidently there are at least two senses of "set" on the go in Tait's discussion, depending on how seriously one takes the identification of sets with objects of type $\mathcal{P}(\mathbf{A})$, and in one of these senses the scope of x is not a type in Tait's sense, and isn't a set either.

More importantly, though, even if we grant that this is not a counterexample to the correctness of the formal proof, it surely is a counterexample to the idea that the informal analogue of the proof shows that the Axiom of Choice or anything like it follows from a constructive understanding of the quantifiers. For the counterexample shows, at least arguably, that the notions of *constructivity* and *provability in this sort of type theory* come apart. For the identification of $\forall x_{\mathbf{A}} B(x)$ with $\prod_{x_{\mathbf{A}}} B(x)$ is justified by taking proofs of the former, i.e., elements of the latter, to be *functions*. But this is a specialization of the usual notion of constructive proof of a universal proposition, which requires only a *method*, and not necessarily a *functional method*. The result is that certain proofs one might be inclined to classify as constructively acceptable *cannot be so classified, if we take the PAT perspective as our touchstone for constructivity*. If that's right, then it seems that the validity of AC comes with a cost, namely a rather severe restriction on what counts as legitimate mathematics.

I think a familiar lesson lurks here. Tait asks rhetorically of his informal argument that Lebesgue's understanding of mathematical existence should commit him to AC, "What more should be required of a proof of AC?" The answer, it seems, is some reason to suppose that every "method" establishing a universal claim must be functional. This is an assumption which is built into the machinery of the PAT view, which is why the formal proof of AC in PAT is correct. But it is very easy to overlook this additional requirement for the validity of AC, and so to think that the validity of AC in PAT allows one to argue in ways which it doesn't. AC, as a principle of PAT, is a weaker claim than it at first seems, one that even in the presence of LEM probably doesn't warrant the acceptance of the Axiom of Choice as that name is usually understood. If that's right, then the familiar lesson is that it's easy to overlook such qualifications when a formal proof is glossed in natural language which *looks* for all the world to be a simple restatement of it, and so that it's remarkably easy to mislead oneself.

5.1 AC and The Axiom of Choice

Of course, constructivists are not unacquainted with the suggestion that their strictures have the effect of misclassifying some legitimate mathematics as illegitimate. So one might consider the question: is this a restriction someone tempted by the PAT view could live with?

The answer seems to me to be: probably not. One piece of indirect evidence is that Tait cannot help giving arguments which range outside this box as he makes his case for it. Consider, for instance, Tait's argument that Lebesgue's rejection of the Axiom of Choice makes his view incoherent. Tait distinguishes AC from the Axiom of Choice as formulated by Zermelo, which we'll call ZAC;

> For every set u of non-empty sets, there is a choice function for u, that is, f is defined on u and, for all $x \in u$, $f(x) \in x$.

Note that ZAC refers to *sets*, and not just to types, and that while Tait thinks AC is constructively valid, he holds no brief for ZAC.

Lebesgue's complaint against ZAC, as it is presented by Tait, is the familiar one that it unacceptably asserts the existence of a function f without specifying how $f(x)$ is to be defined for each value of x. In this Tait detects a commitment to a constructivist understanding of the quantifiers. Hence, by the informal argument above, he's also committed to AC. But elsewhere Tait detects a commitment to double negation elimination, which is constructively equivalent to LEM. And, he argues, AC + LEM is equivalent to ZAC. So Lebesgues view is incoherent.

There seem to be some pretty compelling indirect reasons for suspecting that something must have gone wrong here. If Tait's argument is correct, what room remains for the existence of *classical* models in which ZAC fails? These are thought by many to exist in abundance, as was shown, e.g., by Cohen's independence proofs. Tait certainly looks to be committed to the claim that anyone accepting that the principles of constructive logic are valid and who accepts LEM, is thereby committed to holding ZAC valid, on pain of incoherence. Thus the least we can say for Lebesgue in thinking he could accept LEM, presuming that he did, while questioning ZAC, is that he has a lot of company in his incoherence.

Presumably if one doesn't want to give Lebesgue so much company in his incoherence, one will need to distinguish two senses of "accepting constructive principles." There's the obvious sense in which all who accept classical logic do so, since every valid principle of constructive logic is likewise classically valid. But there's a stricter sense in which one accepts constructive reasoning only if one also buys the restrictions that advocates of constructive mathematics advocate. Perhaps Lebesgue, by seeming to insist that existence claims require the production of a witness for their legitimacy, seems to be buying into constructive reasoning in this stricter sense. And it's only such people who will be committed to accepting AC, and so if they also accept LEM will end up committed to ZAC. But if this is Tait's view, then it's not easy to see what the detour through AC and ZAC is supposed to add to the charge of incoherence he levels at Lebesgue—for LEM added to constructive logic implies the validity of existential claims for which we can produce no witness, and so in that sense it's incoherent to advocate this sort of requirement about existence along with LEM, and the discussion of axioms of choice is a red herring.

Tait's argument that if we assume LEM, then AC implies ZAC is this: Suppose $\forall x_{\mathcal{P}(\mathbf{B})}(x \in u \to \exists y_{\mathbf{B}}(y \in x))$, i.e., that u is a set of non-empty sets. Then $\forall x_{\mathcal{P}(\mathbf{B})} \exists y_{\mathbf{B}} \Phi(x, y)$, where Φ is the formula $(x \in u \to y \in x)$. (It is here that Tait says

LEM is required.)[7] So, by AC, $\forall x_{\mathbf{A}}\Phi(x, f(x))$, i.e., f is a choice function for u.

What ought we to say about Tait's argument that ZAC follows from AC and LEM? The proof works by constructing an antecedent of an instance of AC, with Φ being $x \in u \to y \in x$, where u is a subset of a power type, i.e., a set of sets. But there are two notions of "set" at work in Tait's writing, as we've seen. If we mean by "set" *an element of* $\mathcal{P}(\mathbf{A})$ *for a type* $\mathbf{A}$, then this claim involved the 'implicit assumption' that $x \in u$ must be a decidable formula. So u must be a *decidable* subset of that power type, and so what Tait has shown to follow from AC and (†) is not ZAC, which is emphatically a statement about *arbitrary* sets of non-empty sets, but some less general principle. On the other hand, if we mean by "set" that u is an *arbitrary* collection of elements of some power type, then we have no reason to suppose that we can construct the required instance of AC. The upshot of this is that neither Lebesgue nor anyone else needs to accept ZAC simply as a consequence of accepting the principles of constructive logic and LEM.

An indirect lesson can be learned here. It seems to me that the slide back and forth between two notions of "set" detectable here is the likely product of any attempt to think of "sets" entirely in accordance with the notion of "set" licensed within the PAT view. There are simply too many claims involving sets that even the advocates of such a view recognize as meaningful that cannot be formulated within that framework.

[7]Notice that this proof actually appeals only to the principle that, if y is not free in Φ, then,

$$\Phi \to \exists y \Psi(y) \text{ implies } \exists y(\Phi \to \Psi(y)). \quad (\ddagger)$$

This principle is strictly weaker than LEM, which casts some doubt on the equivalence claim as Tait formulates it. But (‡) is easily seen to be equivalent to $(\exists x(\exists y \Phi(y) \to \Phi(x))$, which is well known to be nonconstructive(see (DeVidi 2004) for discussion of this principle), so if Tait's arguments are correct, the point remains that while ZAC constructively implies AC, AC implies ZAC only if we assume some non-constructive principle.

5.2 Problems with PAT: Choice *is not* Constructive

Let us finally turn to a sketch of the proof of Diaconescu's proof that the Axiom of Choice *implies* LEM in intuitionistic set theories. Since we want to consider why the proof goes through in some theories but not others, it is useful to begin with a "stripped down" version of the proof presented by John Bell, in (Bell 1993*a*). In any intuitionistic theory in which: (1) there are two terms d and c such that $\vdash d \neq c$, and (2) we can find, for any formula A of the language, two terms s and t such that

$$\vdash (s = c \vee A) \wedge (t = d \vee A) \qquad (*)$$

and

$$\vdash A \rightarrow s = t, \qquad (**)$$

LEM is valid (Bell 1993*a*, p. 7). For since the distributive law is valid in intuitionistic logic, $(*)$ yields

$$\vdash (s = c \wedge t = d) \vee A.$$

Since we assume $\vdash c \neq d$,

$$\vdash s \neq t \vee A.$$

But from $(**)$ we have $\vdash s \neq t \rightarrow \neg A$, and so

$$\vdash \neg A \vee A.$$

To sketch the proof for any other theory, then, it suffices to consider what about the theory allows us to prove instances of $(*)$ and $(**)$ for each A. Let us consider an intuitionistic set theory. No real familiarity with intuitionistic set theory is needed for this sketch; one can simply think of it as a theory in which the usual principles of set theory hold (minus the Axiom of Choice of course), but the underlying logic is intuitionistic rather than classical.

So, in an intuitionistic set theory, 0 and 1 will obviously serve as the terms d and c, since $\vdash 0 \neq 1$. To prove $A \vee \neg A$, choose a y not free in A and define $B(y) \equiv A \vee y = 0$

and $C(y) \equiv A \vee y = 1$. Since this is a set theory, there will be something akin to the Axiom of Unordered Pairs which guarantees the existence of the set $\{0,1\}$, and the Axiom of Separation which gives us the sets $z = \{y \in \{0,1\} \mid B(y)\}$ and $w = \{y \in \{0,1\} \mid C(y)\}$. Let f be a choice function on $\mathcal{P}(\{0,1\})$, the power set of $\{0,1\}$. Clearly $\vdash A \to z = w$, by the Axiom of Extensionality. Moreover, z and w are both *non-empty* subsets of $\{0,1\}$. Thus by using $f(z)$ and $f(w)$ as the terms s and t we have: $(**)$, i.e., $\vdash A \to f(z) = f(w)$; and $(*)$, since $\neg A$ implies both $z = \{0\}$ and $w = \{1\}$.

Next, let's consider the question of why this proof cannot be carried out in PAT. It is useful to consider this matter in two different ways, roughly corresponding to considering the approaches to PAT taken by Martin-Löf and Tait in turn. First,[8] suppose that in the course of the argument we prove that both z and w are non-empty. Since A will not, in general, be provable, and a proof of $\exists x.x \in z$ involves presenting a witness that we can demonstrate to be in z, this proof will, in general, employ 0 as that witness. Likewise, the proof of $\exists x.x \in w$ will normally employ 1 as witness. The reason this argument cannot be carried through in the Martin-Löf versions of PAT is that f is a function not merely of the properties which yield non-empty subsets of a type, but also on the *proof* that the property does so. Thus we can set $s = f(z, 0)$ and $t = f(w, 1)$ and prove $(*)$ (modulo the qualification in footnote 8). But now, of course, $z = w$ needn't entail $s = t$, and so we cannot derive $(**)$. Thus AC doesn't imply LEM in this sort of PAT theory because *the identity of subsets is not an extensional matter*. And, as Bell notes, it is shown in (Maietti & Valenti 1999) that if extensional power sets are added to this sort of PAT, logic becomes classical.

On the other hand, Tait's version of PAT takes identity of sets to be an extensional matter. In this case we must recall

[8]This very useful presentation of the matter is borrowed from (Bell to appear). Bell also notes that, at least in Martin-Löf's system, $(*)$ may not be derivable, since its proof depends on the validity of the principle $a \in \{x|\Phi(x)\} \to \Phi(a)$, which fails in that system.

instead Tait's terminological stipulation that the subsets of a type **A** should be identified with the members of the type $\mathcal{P}(\mathbf{A})$, where $\mathcal{P}(\mathbf{A})$ is understood in Tait's sense. Officially, elements of $\mathcal{P}(\mathbf{A})$ are only determined by *decidable* predicates $\Phi(x_{\mathbf{A}})$. Then the problem with the proof of Diaconescu's theorem is obviously that z and w will not, in general, determine subsets of **2**, since Φ will not in general be decidable. On the other hand, as we have seen, Tait sometimes seems to speak as though it is legitimate to talk about the set of $x_{\mathbf{A}}$ with some property, even though we might have no proof that this corresponds to an element of $\mathcal{P}(\mathbf{A})$. If we allow ourselves to use the term 'set' in this looser sense, the reason we cannot show that AC implies LEM using the proof of Diaconescu's theorem is that it appeals to ZAC and not merely to AC. For in this case we are really appealing to AC to get a choice function on $\{\{z\}, \{w\}\}$, for z and w are indeed non-empty sets in the current sense, and we need a choice function on this set of non-empty sets to get the proof to go through. AC doesn't guarantee the existence of such a choice function because unless the predicates defining z and w are decidable, we're not going to get an instance of the antecedent of AC for the very reasons Tait gave in rejecting the purported counterexample to AC discussed above. And, as we have seen, ZAC is a strictly stronger principle than AC.

In both these sorts of PAT theory, then, the notion of subset heavily revised, preventing the derivation of Diaconescu's theorem. That is, we either restrict ourselves to decidable subsets, or we give up the idea that identity of subsets is an extensional matter. It is worth contrasting Local Set Theories with PAT in these respects.

First, subsets are determined extensionally in Local Set Theories, unlike the Martin-Löf PAT theories. On the other hand, one sometimes hears it said that in intuitionistic systems the power set of **2** can be very large, and that is precisely because each class of provably co-extensive predicates with a free variable of type **2** determines a subset of **2**, namely $\{\, x_{\mathbf{2}} \mid \Phi(x)\,\}$. But for Tait the only subsets of **2** are $\varnothing$, $\{0\}$,

$\{1\}$ and $\{0,1\}$, just as in classical mathematics, and so the existence of a choice function on $\mathcal{P}(\{0,1\})$ is likewise the same triviality for Tait as it is in classical set theory. But unless z and w can be proved to be among these sets, we cannot use a choice function on $\mathcal{P}(\mathbf{A})$ to choose elements from z and w, and so we cannot carry the proof through.

5.3 Some Philosophical Lessons

The most obvious lessons in all this have to do with just how subtle and prone to mislead some of the claims bandied about in discussions of constructivism can be. For if what I have suggested above is right, it is a mistake to think the 'witnessing' requirement involved in a constructive reading of the existential quantifier makes choice principles constructively correct.[9] Indeed, the argument that the Axiom of Choice is constructively valid appeals not only to the constructive understanding of the existential quantifier, but also of the universal quantifier. And, as we have seen, there is temptation to fail to keep in mind the distinction between the *method* required by a constructive demonstration of a universal claim and a *functional method*, as is required by the Axiom of Choice.

I would also like to suggest that the restrictions required to make AC come out valid ought to have us conclude that it is not, properly speaking, something we should regard as a *version* of the Axiom of Choice at all. For as we have seen, in order to make AC come out valid some significant modification of the standard notion of *set* is required. We might move, with Martin-Löf, to a notion of subset which is non-extensional. Or we might move, with Tait, to a notion of set where subsets are determined only by *decidable* predicates, if we take the notion of set to be the one internal to his version of PAT. I want to suggest that either move does too much damage to the original understanding of the Axiom of Choice, which in its natural formulation is a claim about the existence

[9]This is the main theme of (DeVidi 2004).

of a choice function for *an arbitrary collection of non-empty sets*, where sets are precisely the sorts of entities whose identity is determined extensionally and which are determined, as nearly as avoiding paradox allows, by the principle of comprehension.

Now, I don't want this to reduce to a claim that these features are somehow "analytic of the notion of set," though I suspect that might be true. It's enough to note that this notion of set informs theorists' understanding of the Axiom of Choice. Like named axioms based on crucially distinct notions are a recipe for equivocation and mistaken conclusions. Moreover, I think that this notion of set is in any case *unavoidable* for saying much that we want to say in doing mathematics, whether we are constructivists or classical mathematicians. We see this in Tait's need to appeal to this sort of collection when he quantifies, at least in English, over all Cauchy sequences. We see it in our need to quantify over all the terms in a language in the statement of the witnessing requirement in any explanation of the nature of the constructive existential quantifier. We see it, indeed, in the need to distinguish "basic sets" from other sets, as constructivists such as Bridges and Reeves do. To rule such collections "non-mathematical" is simply too much of a restriction. It seems to me, then, that if we must countenance some discussion of such collections, even if they are not the primary subject of our investigation, it is most appropriate to call them sets, and to find another name for other sorts of collections which are, for instance, not determined extensionally or which are only determined by decidable predicates, or whatever.

But if that is what sets are, and if they are unavoidable, then even if one accepts a principle like AC for some other class of entities, there is always the natural question of whether a similar principle holds for arbitrary collections of non-empty *sets*. It's perhaps not surprising that *this* is a principle that constructivists are not going to be happy with. But it restores things to their rightful order by making the Axiom of Choice into the non-constructive principle most everybody always thought it was. And Diaconescu's theorem remains a remarkable discovery of just how non-constructive it is, as it shows that the Axiom of Choice, properly so-called and not to be confused with similar looking restricted choice principles which are constructively valid, implies all of classical logic.

The problems with the Propositions as Types view as a theory of meaning, even of mathematical meaning, are not the ones Williamson points to. Rather, as we saw in the discussion of the Axiom of Choice, the Propositions as Types view simply has far too few propositions available to be a satisfactory theory of meaning, even in the restricted domain of mathematical discourse. As we have seen, even constructivists will want to say many things for which there will be no corresponding *propositions*, if propositions are types. An advocate of the PAT view as an account of mathematical meaning could say that such claims are meaningless after all, though this seems rather drastic. On the other hand, he might say they are meaningful, but not mathematical, and insist that it's only an account of mathematical propositions that's on offer. But this sort of persuasive definition of "mathematics" is unlikely to satisfy anyone, either. It seems hard to avoid the conclusion that the advocate of a Propositions as Types view is not going to be able to avoid saying something like this: not all meaningful mathematical claims are propositions. But this is to admit that there are meaningful mathematical claims which are not given by the conditions of provability, at least as those are spelled out in the PAT systems.

6 Conclusion

Where does this leave us? First, Williamson has not shown that an assertion theoretic account of meaning is impossible because of a commitment to luminosity. For as the PAT view demonstrates by its existence, it is possible to hold to a view that the meaning of a proposition is determined by its assertion conditions without thereby committing oneself to the luminosity of those assertion conditions. What Dummett insists on, and what he claims a realist, truth-conditional theory of meaning cannot obviously explain, is that a speaker should know the meanings of the sentences understood. So what is required is knowledge of the assertion conditions of these sentences. That, as we've seen, can be formulated in the manner employed by Martin-Löf and Tait, under which such knowledge doesn't require luminosity, i.e., one in which it is possible that a speaker be mistaken in all the expected ways about whether those conditions obtain in a particular case.

However, as we have also seen, there are serious deficiencies with PAT if it is offered as an account of mathematical meaning. In particular, the supposition that all propositions have their meanings given by what counts as a canonical proof seems unlikely to be able to give us a story about the meaning of everything which ought to count as a meaningful mathematical statement. While this version of the assertion theoretic account of the meaning of mathematical sentences doesn't fall prey to Williamson's objections, I think the facts reviewed above suggest that it's a much more seriously constrained account of mathematical meaning than is sometimes recognized.

Chapter
FOUR

MONTAGUE'S MODAL COMPLETENESS THEOREM OF 1955

B. Jack Copeland

Graham Solomon and I met when he spent a short sabbatical in New Zealand in 1996. Graham's interest was piqued by my assertion that Erewhon, high in the foothills of the Southern Alps, is an important site for anyone interested in the history of Artificial Intelligence. Erewhon's first owner was the novelist and critic of Darwin Samuel Butler, who ran sheep there during the earliest days of European settlement. Living in complete isolation in a small hut, Butler passed his time writing and playing a piano that he had carted into the wilds on a bullock dray. His *Erewhon; or, Over the Range*, and especially its section "Book of the Machines," should be mentioned on the reading list of every course dealing with the philosophy or history of AI.

Erewhon lies at the end of a long and rugged mountain track, with several river crossings. Graham proved himself to be a phlegmatic travelling companion. Once I missed the route and water rose to the bottom of the doors of our four-wheel drive, but Graham did not so much as pause in his discussion of Demopoulos on Newman on Russell (Demopoulos

D. Devidi and T. Kenyon (eds.), A Logical Approach to Philosophy, 77-83.

& Friedman 1985). Our conversation did mostly concern the history of logic, despite our destination. We were both interested in Richard Montague's 1960 paper "Logical Necessity, Physical Necessity, Ethics, and Quantifiers" (Montague 1960). This is based upon a talk that Montague gave at UCLA in 1955. In the paper Montague remarks that his modal deductive system is complete (p. 264) and in a footnote he says this about the paper:

> I … did not initially plan to publish it. But some closely analogous, though not identical, ideas have recently been announced by Stig Kanger (in "The Morning Star Paradox" (Kanger 1957*a*) and "A Note on Quantification and Modalities" (Kanger 1957*b*) and by Saul Kripke (in "A Completeness Theorem in Modal Logic" (Kripke 1959*a*)). In view of this fact, together with the possibility of stimulating further research, it now seems not wholly inappropriate to publish my early contribution. (p. 269)

Montague's editing left it unclear whether the claim about completeness was part of the 1955 material or whether it had been added at the same time as this footnote. Did Montague have a model-theoretic modal completeness proof as early as 1955, some three years before Kripke? Graham was acquainted with Montague's student Charles Silver and on the way back from Erewhon we decided to find out what Silver could tell us about this fascinating question. (Kripke submitted his completeness proof for the modal system S5 with quantifiers to the *Journal of Symbolic Logic* in the spring of 1958 (Kripke 1959*a*) and had a completeness result for quantified S4 by the summer of that year, although this was not published until 1963 (Kripke 1963); see further my article "The Genesis of Possible Worlds Semantics" (Copeland 2002).)

This is what we learned from Silver:

> After praising Kripke's work one time, Richard Montague mentioned that he too had a modal system

> as early as 1955, which was similar to Kripke's. After saying this, he paused, looked down at the table sadly and said softly, "but no completeness proofs."

A year later, hoping to find out more, I spent a few days examining Montague's *nachlass* at UCLA. I found Montague's handwritten notes for the 1955 talk. This manuscript overlaps yet is not identical with the version published in 1960. Among the differences are the presence in the 1955 manuscript of two short sections not appearing in the 1960 paper, entitled "Ethics" and "Quantifiers," and the inclusion in the 1960 paper of a section entitled "A Missing Law" not appearing in the 1955 manuscript.

Montague deals with a modal system that becomes equivalent to S5 upon the addition of the M-principle $\Box A \rightarrow A$. (Montague explained in "A Missing Law" why he omitted the M-principle: $\Box A \rightarrow A$ fails for both ethical obligation and physical necessity (p. 268).) Montague's purpose in the 1955 talk was to extend a Tarskian definition of satisfaction-in-a-model to the modal case. He defined a model as an ordered triple $\langle D, R, f \rangle$, where D is a domain, R is a function that assigns an appropriate extension (from D) to each predicate and individual constant, and f is a function that assigns to each individual variable a member of D. The treatment that he offered of logical necessity is this:

> it seems reasonable to consider "it is logically necessary that φ as asserting that φ holds under every assignment of extensions to its descriptive constants [predicate and individual constants].

Montague borrowed the satisfaction clauses for atomic statements and truth-functional compounds "without alteration" from Tarski. He noted that Tarski's satisfaction clause for "$\forall x$" may be thought of as involving a binary relation Q between models. $\langle D, R, f \rangle Q \langle D', R', f' \rangle$ if and only if $D = D'$, $R = R'$, and $f'(\alpha) = f(\alpha)$ for every individual variable α

different from x. Reading "$\Box$" as "for all x," the satisfaction clause for the universal quantifier becomes:

> $\langle D, R, f \rangle$ satisfies $\Box\varphi$ if and only if, for every model M such that $\langle D, R, f \rangle QM$, M satisfies φ.

Montague generalized this idea, allowing arbitrary binary relations between models. Where X is any such relation, *Satisfaction*$_X$ is defined by replacing "satisfies" by "satisfies$_X$" in the clauses for the truth-functions and by adding the following clause for the operator $\Box$:

> $\langle D, R, f \rangle$ satisfies$_X$ $\Box\varphi$ iff for every model M such that $\langle D, R, f \rangle XM$, M satisfies$_X$ φ.

Montague then stated a soundness theorem in terms of conditions on his binary relation. Where φ is any formula derivable in his deductive system, every model satisfies$_X$ φ, provided only that X fulfills the following conditions:

1. for all M, there is an N such that MXN
2. for all M, N, P, if MXN and NXP, then MXP, and
3. for all M, N, P, if MXN and MXP, then NXP.

Examination of the 1955 manuscript revealed that Montague did indeed claim a completeness proof for his propositional deductive system (although he did not exhibit the proof):

> a formula is valid if and only if it is a theorem. Furthermore, there is a *decision method* for the class of valid formulas …

A later version of the manuscript, partly typewritten, contains the amplification:

> A proof of the completeness and decidibility of the system can be obtained without much difficulty from the ideas in the article of Wajsberg cited above.

(This article by Wajsberg (Wajsberg 1933) contained an extended calculus of classes equivalent to the modal logic S5. Wajsberg simulated the necessity operator in his calculus by means of an expression $|X|$; this notation, originally introduced in (Hilbert & Ackermann 1928), indicates that the predicate X "applies to all objects.")

What Montague did *not* claim is a completeness result for the *quantified* form of the system. In the section of the 1955 manuscript entitled "Quantifiers" (omitted from the 1960 paper), he remarked:

> It has been seen that [□] can be eliminated in favour of quantifiers in the second-order predicate calculus. …In fact, the theory which contains quantifiers and [□] (and no other modal operators) seems to lie between the first-order and the second-order predicate calculus in power of expression. The first-order calculus can be completely axiomatized; the second-order calculus cannot. There is hope that the theory with [□] can be completely axiomatized.

The previously mentioned footnote 5 of the 1960 paper sheds additional light. There Montague said that his completeness result is equivalent to the following:

> A formula φ (of the language S) is a theorem … if and only if φ is satisfied by every complete model. (p. 269)

It is made clear that language S contains "no quantifiers," only individual variables and constants, predicates, truth-functional connectives, and the modal operator □. Later in the 1955 manuscript Montague did extend the language in various ways, adding quantifiers and identity, and allowing more than one primitive modal operator; but there were no completeness claims.

The earliest completeness results for modal systems—predating Montague's talk—were algebraic, not model-theoretic.[1]

[1] Some important articles in the algebraic literature are: (McKinsey & Tarski 1948, Scroggs 1951, Jónnson & Tarski 1951, Jónnson & Tarski 1952)

Did Montague have, therefore, the earliest completeness proof for a propositional modal logic relative to a model-theoretic semantics interpreted in terms of possible worlds and an accessibility relation between worlds? In a famous review of Kripke's 1963 paper "Semantical Analysis of Modal Logic I," Kaplan stated that in 1955 "Montague suggested the interpretation of modal calculi in terms of a relation between worlds" (Kaplan 1966, p. 122). This is highly misleading, since Montague's binary relation is a relation between models, and at no point in the 1955 manuscript did Montague suggest that the relation be understood as holding between worlds. The notion of a possible world was simply absent. (Montague mentions Carnap's 1946 interpretation in terms of state descriptions (Carnap 1946) only to reformulate Carnap's account in terms of models.)

Montague's binary relation functions only to ensure that $\Box A \rightarrow A$ is not always satisfied (and there is no discussion of other systems weaker than S5). Montague offered no interpretation of the binary relation, either in the 1955 manuscript or the 1960 version. Montague himself emphasized in later work that the binary relation of his 1955 talk was a relation between models, not a relation between "points of reference" (Montague 1974, p. 109). (Montague continued: "accessibility relations between points of reference … appear to have been first explicitly introduced [by] Kripke [(Kripke 1963)]." This is not correct, however. A binary relation between points of reference interpreted as worlds appeared in the earlier work (Meredith & Prior 1956).) indexPrior, A.N.

David Lewis wrote to me as follows (in 1996) concerning Montague's views on models versus worlds:

> I don't know about 1955–60, but in later years I think Montague would have opposed taking models as worlds for a familiar technical reason. Two

> worlds might be just alike in their domains and in the extensions they assigned to predicates—alike *qua* models—yet not alike in their accessibility relations. Likewise, still more obviously, in the case of times; and it tended to be thought that worlds and times would be treated alike. So identifying worlds (or times) would have amounted to imposing a troublesome and unmotivated constraint on model structures.

Montague's 1955 theory, then, is probably best regarded not as an early example of possible worlds semantics as such, but simply an extension of Tarski's model theory to a language containing modal operators. In the years that followed, Montague worked on the completeness problem for propositional modal logics in association with Kalish[2]. The two obtained "many partial results." These results were to be presented at an APA meeting in December 1959. Shortly before the meeting, Montague and Kalish saw an abstract by Kripke in the December issue of the *Journal of Symbolic Logic*, announcing completeness results for a wide range of modal systems (including M, S2, S3, S4, S5, S6, S7, S8, E2, E3, E4, E5′, related systems intermediate between M and S2, systems using the Brouwersche axiom, and various systems of deontic logic) (Kripke 1959*b*). Astonished, Montague and Kalish simply withdrew their paper. Kaplan recalls that Montague was curious to know whether S. Kripke was a man or a woman; everyone was surprised when Kripke turned out to be a child.

[2]This paragraph is based on my conversations and correspondence with Kalish (1998). Quotation marks indicate Kalish's words.

Chapter
FIVE

ON THE RATIONAL RECONSTRUCTION OF OUR THEORETICAL KNOWLEDGE

William Demopoulos

1 Introduction

My focus in this paper is the rational reconstruction of physical theories initially advanced by F.P. Ramsey, and later elaborated by Rudolf Carnap. As will become clear in what follows, the Carnap–Ramsey reconstruction of theoretical knowledge is a natural development of classical empiricist ideas, one that is informed by Russell's philosophical logic and his theories of propositional understanding and knowledge of matter; as such, it is not merely a schematic representation of the notion of an empirical theory, but the backbone of a general account of our knowledge of the physical world. Nor is it merely an interesting episode in the history of the philosophy of science; Carnap–Ramsey is an illuminating, if not ultimately satisfying, approach to epistemological problems that remain with us.

To give a preliminary overview, the classical epistemologi-

William Demopoulos, "On the Rational Reconstruction of our Theoretical Knowledge," *British Journal for the Philosophy of Science*, **54(3)** (2003), pp. 371–403. Reprinted by permission of Oxford University Press.

D. Devidi and T. Kenyon (eds.), A Logical Approach to Philosophy, 84-127.

cal issue to which Russell sought to apply his logical discoveries was that of showing how, on the basis of minimal assumptions regarding the scope of our experience, one might articulate an account of our knowledge of the material world. A first step toward a solution would have to address the fact that we succeed in *understanding* propositions which transcend the limitations of our experience. It is how Russell addressed this first step that is of primary interest for what follows. In the reconstructive program of Carnap and Ramsey, Russell's problem is transformed into an issue in the theory of theories: on the basis of a particular choice of a minimal non-logical vocabulary, to recover our theoretical knowledge within an expressively equivalent framework, a framework that preserves the characteristic features of our pre-analytic applications of the concepts of logical consequence, reference and truth. It is a condition of adequacy accepted by both programs that they should recover many pre-analytic intuitions about our theoretical knowledge. Thus, although both the classical and reconstructive programs have foundationalist overtones, it would be a mistake to view them as motivated by skeptical doubts concerning our theoretical beliefs.

In light of the extreme generality of the Carnap–Ramsey reconstruction, it might seem tendentious to characterize it as a reconstruction of *physics*. Indeed, the highly schematic and abstract style that the approach exemplifies has given way to more specialized foundational investigations of particular classes of physical theories. Unfortunately, this shift was accomplished without sufficient appreciation of what the reconstructive program we will be examining sought to accomplish. The aims of the Carnap–Ramsey reconstruction mandated that it should be stated with great generality; even though the reconstruction does not depend on the characteristics of any special class of physical theories, its applicability to physics is essential to its epistemological point. I believe—and will try to show—that insuperable difficulties confront the Carnap–Ramsey reconstructive program. But I hope also to make it clear that Carnap's and Ramsey's reconstructions possess not only internal coherence and elegance, but more importantly, a degree of philosophical motivation not matched by rival accounts. Let me begin with a review of the relevant Russellian background to the program I will be exploring.

2 Russell 's Theory of Propositional Understanding

Russell's first reasonably well-articulated application of his theory of descriptions to a traditional epistemological problem—namely that of determining the nature and scope of our knowledge of matter—occurs in *The Problems of Philosophy* where the theory is extended by the addition of the description theory of names and deployed in support of an exceptionally simple theory of propositional understanding or theory of meaning. Putting to one side the issue of vacuous names, Russell's theory of meaning tells us that if a sentence $S(n)$ contains a name n for an individual with whom we are not acquainted, the proposition expressed by the sentence cannot contain the bearer of the name among its constituents. We must instead imagine that the name is short-hand for a description. This description is in turn analyzed—'contextually defined' after the fashion of the theory of descriptions—into expressions for individuals and propositional functions which *are* proper constituents of the proposition expressed. The individuals and propositional functions are so chosen that the resulting proposition satisfies what in *Problems* Russell called 'the fundamental principle in the analysis of propositions containing descriptions: Every proposition which we can understand must be composed wholly of constituents with which we are acquainted' (Russell 1912, p. 58).

Notice that the point of the theory of *Problems* is not to eliminate what a non-vacuous name stands for, but to explain, compatibly with the fundamental principle, how a sentence containing the name is understood.[1] For example, I understand the sentence, 'Bismarck was an astute diplomat' because I am acquainted with a propositional function that only Bismarck satisfies. A useful way of putting the matter is to say that although the proposition I *assert* with this sentence contains Bismarck as a constituent, and I succeed in saying something *about* him, he is not a constituent of the proposition I *express*.[2] By hypothesis, the bearer of the name 'Bismarck' exists and is the unique individual satisfying some (possibly complex) propositional function-expression. How-

[1]The exposition of this point is made difficult by the fact that Russell is not always consistent about his use of 'contextual definition.' In fact, Russell can be quite *un*clear on the distinction between the contextual analysis of an incomplete symbol and the explicit definition of an entity or 'complete symbol'—occasionally even equating the two notions. Cf., e.g., the following passage from *Logical Atomism*: 'One very important heuristic maxim which Dr.Whitehead and I found, by experience, to be applicable in mathematical logic, and have since applied in various other fields, is a form of Ockham's razor. … The principle may be stated in the form: "Wherever possible, substitute constructions out of known entities for inferences to unknown entities." … A very important example of the principle is Frege's definition of the cardinal number of a given set of terms as the class of all sets that are "similar" to the given set. … Thus a cardinal number is the class of all those classes which are similar to a given class. This definition leaves unchanged the truth-values of all propositions in which cardinal numbers occur, and avoids the inference to a set of entities called "cardinal numbers," which were never needed except for the purpose of making arithmetic intelligible, and are now no longer needed for that purpose. … Another important example concerns what I call "definite descriptions," i.e., such phrases as "the even prime," "the present King of England," "the present King of France." There has always been a difficulty in interpreting such propositions as "the present King of France does not exist." The difficulty arose through supposing that "the present King of France" is the subject of this proposition. … The fact is that, when the words "the so-and-so" occur in a proposition, there is no corresponding single constituent of the proposition, and when the proposition is fully analyzed the words "the so-and-so" have disappeared.' (Russell 1924, pp. 326–28) Russell has here overlooked the fact that there must be some independent motivation for treating something as an incomplete symbol—something more than the mere applicability of the method of contextual definition.

[2]For a fuller elaboration of this distinction and the role it plays in Russell's theories of propositional understanding, see (Demopoulos 1999).

ever, not being known by acquaintance, Bismarck is not *himself* a constituent of the proposition expressed. He is nevertheless someone to whom we are able to refer and make assertions *about* because of our acquaintance with a property only he has. If, for example, Bismarck is identified as the first Chancellor of the German Empire, then I succeed in making assertions about Bismarck because, among other things, I am acquainted with the relation expressed by 'x is Chancellor of y.'

Russell's argument against Berkelian idealism—clearly one of the central lessons of *Problems*—is based on the observation we have just reviewed regarding his description theory of names. On Russell's reconstruction, Berkeley fallaciously assumed that the fundamental principle restricts what we can have knowledge about, what propositions we can assert; but in fact, it restricts only the propositions we can express. The application of Russell's new theories of propositions and denoting to our knowledge of the material world proceeds from three explicit assumptions and one tacit assumption. The explicit assumptions are: (*i*) we are not acquainted with matter; but (*ii*) it is always possible to formulate a description which is uniquely satisfied by the material object to which we take ourselves to refer; and (*iii*) these descriptions involve only propositional functions and individuals with which we are acquainted. The tacit assumption is that (*iv*) the propositional functions with which we are acquainted can be so chosen that their logically primitive constituents apply only to terms with which we are acquainted. Without the tacit assumption (*iv*), the critique of Berkeley and the epistemological significance of Russell's theory would be severely limited, since it could be objected that the view only secures realism about one part of the material world relative to realism about another.

Precisely how the tacit assumption is to be satisfied is something *Problems* only hints at, for example, when Russell writes:

> … if a regiment of men are marching along a road, the shape of the regiment will look different from

> different points of view, but the men will appear arranged in the same order from all points of view. Hence we regard the order as true also in physical space, whereas the shape is only supposed to correspond to the physical space so far as is required for the preservation of the order. (Russell 1912, pp. 32-33)

Russell's full elaboration of this idea is given in *The Analysis of Matter*, where his 'structuralism' is articulated at length.[3] What the response to Berkeley requires is a theory that will allow us to dispense with primitive nonlogical vocabulary items which name or indicate anything with which we are not acquainted—and, in the case of primitive predicates, a theory that admits only primitive non-logical predicates that are true of things with which we are acquainted—while allowing that we can have knowledge about things which fall outside the realm of our acquaintance. In particular, in its application to our knowledge of matter, we demand a theory that will explain how our ability to formulate propositions which express truths about the material world need not in any way require our acquaintance with that world. It is these desiderata that Russell's structuralism was intended to fulfill.

To see at least in outline how structuralism proposed to meet these goals, let us recall what is characteristic of the general characterization of structure in terms of structural similarity and its elaboration in the 'relation arithmetic' of *Principia Mathematica*. The model on which Russell's definition of a structure was based is the Frege–Russell definition of the cardinal numbers as similarity classes under the relation of one–one correspondence. As Frege perceived in *Grundlagen*, and as Russell was to discover some years later, the notion of one–one correspondence, being definable in wholly logical terms, is independent of spatio-temporal intuition. It follows that this must also be true of *structural similarity*, since it

[3] Russell's structuralism is discussed more fully in (Demopoulos 2003) and (Demopoulos & Friedman 1985).

rests only on the notion of one–one correspondence and the general concept of a relation. Thus, for Russell, the *philosophical* interest of structural similarity derives from the fact that, as a notion of pure logic, it owes nothing to experience or Kantian intuition.

From very early on, Russell seems to have seen his account of structure as capable of providing a framework within which it would be possible to articulate the nature of the similarity philosophers had supposed to exist between appearance and reality or, to use Kantian terminology, between the phenomenal and the noumenal worlds—a point whose significance was not lost on Russell, nor, I dare say, was the irony that a concept which owed its genesis to logicism might usefully contribute to the articulation of Kantian doctrine. What had defeated previous attempts was the want of a notion of similarity which was not so great that it would collapse the gulf that was supposed to exist between them, and was not so slight that it could not be reckoned a significant sense of similarity. Russell believed that with the discovery of the notion of structural similarity, he had solved this metaphysical and epistemological problem.

Russell's picture of how the application to Kant should go appears to have been something like this: The noumenal world, not being given to us in intuition, cannot, apparently, be required to have properties in common with the phenomenal world. This leaves us with the problem of understanding how to formulate any conception of what the noumenal world is like and of understanding how it can be knowable. But because structural similarity has a purely logical characterization, it is independent of intuition. The noumenal world thus emerges as an isomorphic copy of the phenomenal world, one which we may suppose has the requisite similarity with the world of phenomena without thereby committing ourselves to the idea that it shares any of the intuitive properties of the phenomenal world. Had it not proved possible to capture this notion of similarity by purely logical means, we would have been precluded from assuming even this degree of similarity between noumena and phenomena, and might, therefore, have been inclined toward some form of idealism regarding the world behind phenomena. The purely logical notion of structural similarity preserves us from this tendency toward idealism, since it shows how we might have knowledge about the relations that order the noumenal world, without needing to assume intuitive knowledge of those relations.[4]

It is clear that the same thought underlies our earlier quote from *Problems* regarding *shape* and *order*: it is not necessary

[4]As noted in the text, structuralism is developed at length in *The Analysis of Matter*, but this application of the view was already announced in *Introduction to Mathematical Philosophy*: 'There has been a great deal of speculation in traditional philosophy which might have been avoided if the importance of structure, and the difficulty of getting behind it, had been realized. For example, it is often said that space and time are subjective, but they have objective counterparts; or that phenomena are subjective, but are caused by things in themselves, which must have differences *inter se* corresponding with the differences in the phenomena to which they give rise. Where such hypotheses are made, it is generally supposed that we can know very little about the objective counterparts. In actual fact, however, if the hypotheses as stated were correct, the objective counterparts would form a world having the same structure as the phenomenal world, and allowing us to infer from phenomena the truth of all propositions that can be stated in abstract terms and are known to be true of phenomena. If the phenomenal world has three dimensions, so must the world behind phenomena; if the phenomenal world is Euclidean, so must the other be; and so on. In short, every proposition having a communicable significance must be true of both worlds or of neither: the only difference must lie in just that essence of individuality which always eludes words and baffles description, but which, for that very reason is irrelevant to science. Now the only purpose that philosophers have in view in condemning phenomena is in order to persuade themselves and others that the real world is very different from the world of appearance. We can all sympathize with their wish to prove such a very desirable proposition, but we cannot congratulate them on their success. It is true that many of them do not assert objective counterparts to phenomena, and these escape from the above argument. Those who do assert counterparts are, as a rule, very reticent on the subject, probably because they feel instinctively that, if pursued, it will bring about too much of a *rapprochement* between the real and the phenomenal world. If they were to pursue the topic, they could hardly avoid the conclusions which we have been suggesting. In such ways, as well as in many others, the notion of structure … is important. (Russell 1919, pp.61-2)

to suppose *acquaintance* with the relation that holds among the members of the regiment provided we can express the *structure* of that relation, as, indeed, the notion of order permits us to do. Similarly, we need to be able to express the structure of the relations that hold among events in the material world, and on this basis, one might hope to construct the whole 'spatio-temporal' framework appropriate to the world behind phenomena. By means of this framework we can proceed to describe all those material events with which we are not acquainted, but regarding which we wish to assert many propositions. But to achieve this we need not be acquainted with the relations that generate this framework.

This is an elegant application of a technical idea of mathematical logic to a philosophical problem. It is, however, subject to an important limitation. If we intend the statement that the noumenal world is isomorphic to the phenomenal one to be more than part of its *definition*—if, that is, we also intend it to be a significant claim that the noumenal and phenomenal worlds are structurally similar—then we are implicitly assuming that we have access to the relations holding among things in themselves independently of the correspondence in terms of which their similarity to phenomenal relations has been characterized. Otherwise, the observation that structural similarity allows us to preserve the comparability of the noumenal and phenomenal worlds is based on a mere stipulation, the character of the noumenal world having been *defined* in terms of the isomorphism. Since it owes nothing to intuition, structural similarity may be used to address the objection that there is *literally* nothing that can be said regarding things in themselves. But on the conception of the noumenal world to which we are led, it must be borne in mind that its similarity to the phenomenal world is the consequence of a definition.

The nature of a claim of structural similarity is such that it is a significant claim only when the relations being compared are given independently of the mapping which establishes their similarity. For this we require more than an appropriately general notion of similarity; we must, in addition, have independent knowledge of the relations between which the similarity is supposed to hold. Knowledge of the relations among things in themselves cannot be purely structural, they cannot merely be known as the images, under a suitable mapping, of the relations among phenomena, since that would make the claim of their similarity to relations among phenomena a tautology. But, things in themselves being 'in themselves,' neither can it be intuitive. The noumenal world would seem, therefore, to have retained almost all of its elusive character. Mutatis mutandis, the relations among events of the physical world cannot be defined as the isomorphic image of the relations holding among those events with which we are acquainted if we intend to preserve the non-triviality of the thesis that the two systems of relations are similar. As we will see, this difficulty is pervasive, and is one that re-emerges in a sharper form in the context of later developments.

Finally, let us note by way of conclusion that Russell's use of the distinction between acquaintance and description rests on the assumption that although knowledge by acquaintance is not required for us to have knowledge about some entity or other, knowledge of a propositional function uniquely satisfied by that entity suffices to ensure *reference* to it. This assumption is questionable if we think of knowledge of reference as something that carries with it knowledge of *what* (or of *whom*, etc.) we are referring to. In fact, Russell does *not* think that such knowledge is required for reference to be successful. Instead, he endorses the view that having such knowledge—'knowledge-wh,' as I shall call it—requires knowledge by acquaintance. But acquaintance seems to be far too strong a requirement for knowledge-wh, if the pre-analytic connection between knowledge-wh and knowledge of reference is to be preserved. It would thus seem that knowledge-wh is a type of knowledge that is not captured by the acquaintance–description division: I know that the next president of the United States will be the person who gets the most Electoral College votes, and I know that *x gets the most Electoral*

College votes in the US election of 2000 is a propositional function that, when satisfied, is uniquely satisfied, but it is at least questionable whether, knowing these things and using this expression, I succeed in *referring* to anyone.[5] And while I am not acquainted with George W. Bush, it is plausible to suppose that I know who he is, and thus satisfy one condition for sometimes successfully referring to him. In short, knowledge by description appears to be too 'thin,' and knowledge by acquaintance too 'thick,' to tell us what knowledge of reference consists in. Perhaps Russell thought that further restricting the propositional constituents with which we refer to things outside our acquaintance to constituents with which we are acquainted overcomes—rather than compounds—the difficulty with this use of knowledge by description. If so, he was mistaken and the extent of his mistake will become apparent from our discussion of the extension of his ideas to the theory of theories.

3 Ramsey's Primary and Secondary Systems

I have been careful to present Russell's elaboration of his theory of propositional understanding in such a way that its connection with a subsequent development by Ramsey will be transparent. A feature of Russell's theory that I have alluded to is the technique by which it avoids the use of a name—or more generally, of any logically simple expression—for something which is not an object of acquaintance. It is this consequence of Russell's deployment of his theory of descriptions that was imitated by Ramsey, in his posthumously published 'Theories,' when he proposed that the 'content' of a physical theory can be captured by what has come to be called its 'Ramsey sentence.' It will be recalled that the Ramsey sentence $R(\theta)$ of a theory $\theta = \theta(O_1, \ldots, O_m; T_1, \ldots, T_n)$ with observational predicates $O_1, \ldots, O_m$ and theoretical predicates

[5] See (Cartwright 1987, p. 117) for an elaboration of this point.

$T_1, \ldots, T_n$ is just the result

$$\exists X_1 \ldots \exists X_n \theta(O_1, \ldots, O_m; X_1, \ldots, X_n)$$

of existentially quantifying on the theoretical terms and replacing them by variables $X_1, \ldots, X_n$ of the appropriate arity and type (or sort, if the underlying logic of $R(\theta)$ is taken to be first order). The replacement of θ by $R(\theta)$ preserves the class of derivable consequences involving the observational vocabulary, what Ramsey called 'the system of primary propositions' of θ; it is in this sense that, for Ramsey, $R(\theta)$ can be said to capture the content of θ.[6] But of course $R(\theta)$ must necessarily diverge from θ in those consequences involving the theoretical vocabulary (Ramsey's so-called 'secondary propositions' or 'secondary system').

Now the point of Ramsey's proposal is to address the role of theoretical terms only in the *deductive structure* of θ, and then to address their role only in that part of its deductive structure that is relevant to the derivation of the primary propositions. Ramsey in effect observed that if θ contains sufficiently explicit deductions of its primary propositions, then these deductions have a representation in $R(\theta)$. But since in $R(\theta)$ the 'propositions' which are the transforms of secondary propositions contain variables wherever the original propositions contained theoretical terms, their meaning is exhausted by the contribution of the observational vocabulary they contain. As Ramsey put it,

> [w]e can say, therefore, that the incompleteness of

[6]The idea that $R(\theta)$ isolates the factual content of θ is indefensible unless some predicates of θ are exempted from Ramsification; otherwise the 'Ramsey sentence' of any consistent first order theory will be true in all models (L-true), true in every model whose domain of individuals has cardinality n for some finite cardinal n, or true in every model whose domain of individuals is denumerably infinite. This follows from the fact that if no predicates of θ are exempted from Ramsification, then the models for the language of $R(\theta)$ are all of form $\mathcal{M} = \langle M, \mathcal{P}(M^1), \ldots, \mathcal{P}(M^n)\rangle$, $\mathcal{P}(M^i)$ the power set of the i-th Cartesian product of M, and a simple application of the observation of Newman discussed below. I am indebted to Aldo Antonelli for suggesting this remark.

> the "propositions" of the secondary system [more exactly, the incompleteness of their transforms in $R(\theta)$ that results from the replacement of constants by variables,] affects our *disputes* but not our *reasoning*. (Ramsey 1960, p. 232)

And since, for Ramsey, it is only our reasoning we need to reconstruct, this incompleteness is irrelevant. The idea of a Ramsey sentence depends on nothing more contentious than this elementary observation about the formal character of logical derivation. But of course to grant this is not to concede the correctness of the view of 'secondary propositions'—as mere auxiliaries in the derivation of primary propositions—which it advances.

Regarding Ramsey's remark about disputes vs reasoning, it is obvious that if two theories T_1 and T_2 conflict in their secondary propositions, this conflict need not be preserved under 'Ramsification' but might well be 'existentially generalized away.' It is however possible to take things a step further by recalling an observation of (English 1973): it is a consequence of the Craig Interpolation Theorem that if two first order theories with disjoint T-vocabularies but coincident O-vocabularies are inconsistent with one another, then there is a sentence σ in their common O-vocabulary which 'separates' them, i.e., which is such that T_1 implies σ while T_2 implies $\neg\sigma$. Hence T_1 and T_2 cannot be compatible with the same 'data.' The proof is straightforward: If $T_1 \cup T_2$ is inconsistent, then by the Compactness Theorem there are finite subsets $\Sigma_i \subseteq T_i$ $(i = 1, 2)$ such that $\Sigma_1 \cup \Sigma_2$ is inconsistent. Let σ_i be the conjunction of the sentences in Σ_i. Then σ_1 implies $\neg\sigma_2$. By Craig Interpolation, there is a sentence σ such that σ_1 implies σ and σ implies $\neg\sigma_2$, where σ is in the common vocabulary of σ_1 and σ_2, i.e., σ is an O-sentence formulated in the common observational vocabulary of T_1 and T_2. But then T_1 and T_2 cannot both be compatible with all observations: if σ holds, then T_2 is false, and if it does not hold, T_1 is false.[7]

[7] English 's discussion is marred by a number of infelicities: contrary to what English claims, the argument on which her observation depends

We can turn this observation about first order theories into one about Ramsey sentences: there cannot be two incompatible Ramsey sentences both of which are compatible with all the true O-sentences. Provided our original theories are first order, it doesn't matter for the application of first order model theory that their Ramsey sentences are second order. The essential idea behind the use of the Ramsey sentence, as we have noted, is that only the logical category of the theoretical terms is relevant to their role in deducing the primary propositions. So long as the replacement preserves the logical category of the original vocabulary items, someone who holds that the Ramsey sentence of a theory captures its 'content' can have no objection to a uniform replacement of the theoretical vocabularies of T_1 and T_2 with new non-logical constants, making their theoretical vocabularies disjoint, and therefore ensuring that T_1 and T_2 satisfy the hypothesis of our corollary to Craig Interpolation. The identification of a theory with its Ramsey sentence therefore comes at a price: the notion that two theories might be compatible with the same data and yet conflict with one another must be given up, so that when theories are identified with their Ramsey sentences, a conflict in some secondary proposition must be reflected in a primary proposition.

The general methodological issues addressed by the notion of a Ramsey sentence do not exhaust Ramsey's interest in the elimination of 'superfluous elements.' A fragment found with 'Theories' shows Ramsey to have had an interest in particular cases that is worth noting. It is unfortunate that this fragment has not been reprinted in any of the published editions of Ramsey's papers since it shows Ramsey to have perceived with remarkable clarity the foundational problem of characterizing Newtonian space-time.[8] The text is worth

does not use the Robinson Consistency Theorem but uses, as the argument given in the text shows, a part of a well-known derivation of that theorem from the theorem of Craig.

[8]The issue received its definitive formulation in the philosophy of science literature with the publication of (Stein 1967).

quoting in full:

> In a completely satisfactory theory I think we should
>
> (a) have a complete dictionary
> (b) have no superfluous elements.
>
> (b) cannot be exactly defined: it means that we cannot get a simpler equivalent theory. But we may be able to do so by a little transformation when we cannot by simply leaving out a part as it stands. Weyl's requirements (Weyl 1922, p. 87) are *Einstimmigkeit* [unanimity] and no *überflüssen gestandteile* [superfluity of expression].
>
> Which seem to mean that every theoretical quantity can in principle be evaluated and that all ways of evaluating it lead to the same result. In principle must here mean merely that certain possible courses of experiences would determine its value.
>
> If not, of course, there is something superfluous ⟨E.⟩g. our velocity in absolute space could not be determined, and so some truth-possibilities of theoretical functions give ⟨an⟩ equivalent theory. ⟨Therefore⟩ some economy ought to be possible, but it is not clear how without a good deal of thought. That makes indeed a good exercise.
>
> What is the proper form of Newtonian Mechanics? Which gives absolute acceleration a meaning⟨;⟩ absolute velocity⟨,⟩ none.
>
> It must be a sort of geometry containing straight lines and a fixed direction. One must give an axiomatic description of such a geometry.[9]

[9] Archives of Scientific Philosophy, Hillman Library, University of Pittsburgh, ms.number: 005- 17-01, dated August 1929. Quoted with the permission of the University of Pittsburgh. My changes are enclosed between ⟨, ⟩.

The problem Ramsey has posed—the formulation of a four-dimensional affine geometry for Newtonian Mechanics—differs from the general methodological issue addressed by the Ramsey sentence in a crucial respect. The transformation of Newtonian Mechanics to *R* (Newtonian Mechanics) takes for granted the theoretical formulation of its secondary system of propositions, but the idea underlying a reformulation of Newtonian Mechanics without absolute space is that it leads to a refinement of its theoretical commitments by eliminating certain putative *primary* propositions, namely all those reporting a measurement of absolute velocity. By contrast, when we 'Ramsify' we assume that the primary propositions have been correctly circumscribed, since there is nothing in the transformation to the Ramsey sentence capable of correcting matters if this has not been done. But this is just to say that Ramsey's distinction between primary and secondary propositions really doesn't address the kind of 'superfluity' that attaches to absolute space and absolute velocity in Newtonian Mechanics, since the source of their superfluousness is the dynamical structure of the theory: in a theory with a different dynamical structure, absolute space and absolute velocity would not be superfluous, and the propositions reporting their state would properly occur in such a theory and would occur among the theory's primary propositions.

In addition to whatever other criteria they fulfill, for Ramsey the primary propositions of a properly formulated theory satisfy Russell's fundamental principle—indeed, Ramsey even goes so far as to say (Ramsey 1960, p. 213) that names of experiences are descriptions unless the experiences named are present experiences. The epistemic significance of names vs variables is therefore much the same for Ramsey as it is for Russell: names require an account of how they are understood, variables do not. Certainly, Ramsey supposes that our understanding of the vocabulary of the primary propositions (the observation or O-vocabulary) is unproblematic. And since only the O-vocabulary is regarded as unproblematic, Ramsey's account is naturally viewed as an extension to the case of theoretical predicates of Russell's analysis of the meaning of names for things falling outside our acquaintance. By eliminating theoretical predicates and replacing them with variables, Ramsey avoids any difficulty their presence might pose for an empiricist theory of propositional understanding.

Ramsey assumes that the referents of the theoretical predicates are not known by acquaintance and that therefore the theoretical predicates are not understood in accordance with the fundamental principle, but he refrains from offering a positive account of the character of our knowledge of theoretical relations along the lines of Russell's structuralism. Ramsey ignores this question and focuses on another: significantly diverging from Russell, Ramsey turns his attention to the collection of propositions which constitute the theory and then considers the effect of identifying the content of the secondary propositions with their consequences in the primary system.[10] The implicit assumption, that only the vocabulary of the primary system is fully contentful, and that our understanding of the secondary propositions consists in our understanding of logic plus the vocabulary belonging to the primary system, is in essence Russell's view. What is novel is the manner in which Ramsey bypasses the issue of the meaning of the individual terms of the theoretical vocabulary by eliminating them in favor of variables, since the technical device by which this is achieved leads to the idea (found later, as we will see, in Carnap) of expressing the content of the theory as a whole by the set of its primary propositions.[11]

[10]Compare the following passage 'the meaning of a proposition about the external world is what we should ordinarily regard as the *criterion* or *test* of its truth. This suggests that we should define propositions in the secondary system by their criteria in the primary' (Ramsey 1960, pp. 222-23).

[11]'So far … as *reasoning* is concerned, that the [transforms in $R(\theta)$ of the secondary propositions] are not complete propositions makes no difference, provided we interpret all logical combinations as taking place within the scope of a [single existential] prefix. … For we can reason about the characters in a story just as well as if they were really identified, provided we don't take part of what we say as about one story, part about another' (Ramsey 1960, p. 232).

Before proceeding to later developments, let me briefly summarize the salient points of comparison between Ramsey and Russell. First, Ramsey is prepared to preserve the idea that the theoretical or ‘T-vocabulary’ is referential, without, however, necessarily preserving the ‘order’ (in the sense of the ramified hierarchy) of the relations, since the systematic substitution of ∃-bound variables for T-vocabulary items requires only that the matrix $\theta(O_1, \ldots, O_m; X_1, \ldots, X_n)$ be satisfied by entities of the appropriate simple type. This I take it is the significance of Ramsey's remark (Ramsey 1960, p.231) that ‘it is evident that [the values of the bound variables] are to be taken purely extensionally. Their extensions may be filled with intensions or not, but this is irrelevant to what can be deduced in the primary system.’ Secondly, Ramsey does not claim to be giving explicit or implicit definitions (the concept of implicit definition is not discussed by Ramsey) of the vocabulary of the secondary system. The paper canvasses various possible general strategies for explicitly defining the secondary vocabulary in terms of the primary vocabulary but concludes that this is unlikely since the ‘secondary system has a higher multiplicity, i.e. more degrees of freedom than the primary … and such an increase of multiplicity is … [likely] a *universal* characteristic of useful theories’ (Ramsey 1960, p. 222). Correlative to this last point, by not offering explicit definitions, Ramsey's approach echoes Russell's contextual definition of problematic names. But there is no question of contextually analyzing the T-terms away in Russell's sense, since $R(\theta)$ is only ‘weakly’ equivalent to θ—equivalent over the sentences taken from the primary system. By contrast, for Russell, the transform $[S(n)]^{\mathrm{T}}$ of a sentence effected by the description theory of names is supposed to be equivalent to the original sentence $S(n)$. Finally, notice that our comparison of Ramsey and Russell is based entirely on Russell's application of his extension of his theory of descriptions to names of things falling outside our acquaintance—on aspects of the theory of propositional understanding which emerged from his theory of descriptions and description theory of names—and has not at any point appealed to Ramsey's view regarding the *nature* of our knowledge of matter and the theoretical relations of physics, beyond the minimalist claim that such knowledge as we have is not knowledge by acquaintance. It might seem therefore that Ramsey's development of the theory of propositional understanding is viable in a way that Russell's elaboration of it in terms of his structuralism was seen not to be. In the next section, we will see that this is not the case, even under a refinement of the theory introduced by Carnap.

4 Carnap's Reconstruction of the Language of Science and an Observation of Newman

The next major development in the tradition I have been reviewing is Carnap's mature reconstruction of the 'language of science,' first presented in his Santa Barbara Lecture of 1959[12] and subsequently developed in (Carnap 1963). Carnap's primary goal was to provide a reconstruction which clearly separated the factual from the non-factual assumptions of a physical theory. Carnap frequently emphasized[13] that virtually every *unreconstructed* sentence of the language of science and everyday life has both factual and non-factual aspects, so that the sharp separation of our language into factual and non-factual sentences is meaningful only relative to a *reconstruction* of that language. His proposed reconstruction is very simple. Like Ramsey, he focused on the factual content of a theory as a whole. And like Ramsey, he took this to be carried by its Ramsey sentence.

For Carnap, as for Ramsey, the decisive consideration that justifies locating the factual content of θ in $R(\theta)$ is that $R(\theta)$ has the same O-consequences as θ. Carnap, however, seeks

[12] Recently edited by Stathis Psillos and published together with a highly informative introduction as (Psillos 2000). For additional historical background, see chapter 3 of (Psillos 1999).

[13] Cf., for example, his discussion of reduction sentences and Ramsey sentences in § 24 of (Carnap 1963), which shows this to have been a feature of early formulations of his views.

to effect a general procedure for isolating the stipulational or non-factual component of θ to a proper part of the theoretical reconstruction, with the factual content of the theory being exhausted by its Ramsey sentence. On Carnap's reconstruction, the non-factual or analytic component is given by what has come to be called the 'Carnap sentence,' $C(\theta)$ $(=R(\theta) \rightarrow \theta)$, of the theory, a sentence which expresses the thought that if anything satisfies the matrix of the Ramsey sentence of the theory, then the referents of the terms of the unreconstructed theory do. Carnap argues that since $C(\theta)$ has the property that $R(C(\theta))$ is L-true, and therefore has no O-consequences except those that are L-true, it is appropriately non-factual.

There are a number of simple connections between θ, $R(\theta)$ and $C(\theta)$ which we should record. θ is obviously equivalent to the conjunction of $C(\theta)$ and $R(\theta)$. Under the assumption that the Ramsey sentence $R(\theta)$ of θ is true,

$$\theta \Leftrightarrow R(\theta) \rightarrow \theta;$$

i.e., under the 'factual' hypothesis that something satisfies θ, θ is equivalent to its Carnap sentence. More significantly,[14] under the analytically true (for Carnap, 'non-factual') assumption of the Carnap sentence, it follows that

$$\theta \Leftrightarrow R(\theta),$$

i.e., it follows that θ is *equivalent* to its Ramsey sentence—not just equivalent over the primary system, to use Ramsey's terminology, but *equivalent*. But since the Carnap sentence is an analytic, and hence, *necessary* truth, we may simply say that θ is equivalent to its Ramsey sentence *without qualification*. Intuitively, by accepting the Carnap sentence as analytic—by stipulating that $C(\theta)$ holds—we exclude as conceptually possible all those models in which the Carnap sentence fails; a model in which $R(\theta)$ holds but θ fails is simply not a possible model.

[14]Cf. (Winnie 1970, p. 294).

As simple and elegant as Carnap's proposed reconstruction is, I think it cannot be accepted as an accurate reflection of our preanalytic understanding of our theoretical knowledge about the physical world. That the attempt to do so leads to evidently unacceptable consequences is, I think, the lesson to be drawn from an old observation of (Newman 1928), an observation that lies at the basis of the formulation of the earliest and simplest of Hilary Putnam's 'model-theoretic arguments.'[15] The application of Newman's observation I will be developing—by contrast with Putnam's deployment of his argument against 'metaphysical realism'—is completely straightforward. The presentation differs from Newman's only in the use of model-theoretic terminology; conceptually, the point is entirely the same.

Suppose we are given a theory θ, all of whose observational consequences are true; it follows from this supposition that θ is empirically adequate and consistent. Suppose also that the observational consequences of θ can be characterized as a subset of the sentences generated from a given O-vocabulary. Suppose further that the interpretation of the language $\mathcal{L}(\theta)$ of θ is specified only for its O-vocabulary, and that the interpretation of the T-vocabulary is fixed only up to the logical type and arity of the T-terms; $\mathcal{L}(\theta)$ is said to be 'partially interpreted.' Notice that partial interpretation is the reflection in the theory of theories of what, in our discussion of Russell's theory of our knowledge of matter, we identified as his tacit assumption (iv): the logically primitive predicate expressions we are entitled to suppose we understand are restricted in their extension to objects of possible acquaintance. Without this assumption, the empiricist motivation for distinguishing between the O- and T-vocabularies—and, therefore, Ramsey's primary and secondary systems—would be lost.[16] Notice also that it is no objection to the distinc-

[15]First presented in (Putnam 1977).

[16]The well-known paper (Lewis 1970) is often represented as a part of the Carnap–Ramsey program we have been reviewing. This assimilation is a mistake. Aside from Lewis's adoption of the formal appara-

tion to observe that it cannot be satisfactorily drawn within the *un*reconstructed vocabulary of the language of science. The point of the reconstruction of theories we are exploring is to show that it is possible to achieve the theoretical knowledge we take ourselves to have by showing how—within a reconstruction in which the distinction between the observational and the theoretical does its intended work—we can give a faithful representation of what we take ourselves to know. Provided we can isolate the observable part of any intended model of θ, it is always possible to introduce into the reconstructed language of θ properly observational predicates, predicates defined in terms of the restrictions of the interpretation of the predicates of the unreconstructed vocabulary to the subset of observable elements of the domain. (Indeed, as we will see later, this possibility establishes a connecting link between Carnap–Ramsey and 'constructive empiricism.')

It is clear from the foregoing that for Carnap θ is effectively identified with the matrix of its Ramsey sentence and that this is entirely in keeping with his view, for Carnap says of the Ramsey sentence of θ, that while it

> does indeed refer to theoretical entities by the use of abstract variables … it should be noted that

tus of the Ramsey and Carnap sentences, both his positive contribution and the philosophical concerns he address are orthogonal to Carnap–Ramsey. Lewis is himself explicit on the central point, namely that his O-vocabulary is not restricted in the way it must be for the Carnap–Ramsey reconstruction to have the epistemological interest claimed for it. A close study of Lewis's paper will show that its contribution is wholly logical: Lewis assumes that the theories to which his analysis applies are such that their O-terms implicitly define the T-terms in the sense that every automorphism of a model for the language of the theory that preserves the relations denoted by the O-terms will also preserve the relations denoted by the T-terms. Lewis's principal contribution consists in embedding this assumption and the formal features of the account of Ramsey and Carnap into a framework whose underlying logic allows for the possibility of denotationless terms. While this is not without an interest of its own, it lacks the philosophical motivation that prompts the Carnap–Ramsey reconstruction. (Thanks to Philip Percival for asking after the relevance of Lewis's paper.)

> these entities are ... purely logico-mathematical entities, e.g. natural numbers, classes of such, classes of classes, etc. Nevertheless, $R(\theta)$ is obviously a factual sentence. It says that the observable events in the world are such that there are numbers, classes of such, etc. which are correlated with the events in a prescribed way and which have among themselves certain relations; and this assertion is clearly a factual statement about the world. (Carnap 1963, p. 963)

That is to say, Carnap shares with Ramsey the idea that the 'factual content' of the theory consists in its consequences in the language of the primary system; the theoretical claims taken by themselves are 'purely logico-mathematical' in character. It has been objected[17] that passing to the Ramsey sentence of a theory that is advanced as merely an idealization, such as the theory of ideal gases, we lose the intuitive content of the original theory, since the Ramsey sentence represents it as advancing the false claim that, for example, *there is* an ideal gas, contrary to our pre-analytic understanding of the theory of ideal gases. However, as our quote from Carnap shows, the understanding of the distinction between theories involving idealization and theories not involving idealization on which this pre-analytic intuition depends is precisely what his use of the Ramsey sentence rejects. It cannot be emphasized too strongly that for Carnap all that needs to be preserved is reference to entities of the appropriate logical category, and it is a matter of indifference whether these entities are 'concrete' or 'ideal' (abstract). Although the case of idealization points to a difficulty in the equation of the factual content of a theory with the content of its Ramsey sentence, the difficulty goes much deeper than the failure of this equation to capture our pre-analytic understanding of theories of ideal systems. This is what we will now show.

Since θ is consistent, there is an 'abstract' model $\mathcal{M}$ of the

[17] See (English 1973).

T-sentences of θ, although nothing of philosophical interest would be lost if we were forced to proceed *directly* from the assumption that θ has such an 'abstract' model and were unable to infer this (e.g. because of the higher-order character of the language in which θ is formulated) from the mere consistency of θ. Without any significant loss of generality or philosophical interest, we may choose an $\mathcal{M}$ that is a model of the same cardinality as θ's intended domain. Let W denote not the domain of the abstract model $\mathcal{M}$, but the domain we take θ to make assertions about—the 'intended domain' of θ. By hypothesis, W has the same cardinality as M. It is therefore possible to extend the partial interpretation to the theoretical vocabulary of θ by letting each predicate of its theoretical vocabulary denote the image in W of its interpretation in $\mathcal{M}$ under any one-one correspondence between M and W. For example, suppose T is a binary theoretical relation of θ. Then the interpretation $T^{\mathcal{W}}$ of T in $\mathcal{W}$ is defined as the image under φ, φ one-one from M onto W, of its interpretation $T^{\mathcal{M}}$ in $\mathcal{M}$. Since by construction $\langle a, b\rangle$ is in $T^{\mathcal{M}}$ if and only if $\langle \varphi a, \varphi b\rangle$ is in $T^{\mathcal{W}}$, φ is an isomorphism; and therefore, if $\mathcal{M}$ is a model of θ, so is $\mathcal{W}$.

Call the interpretation of θ's T-vocabulary in $\mathcal{W}$ that we have just described '$\mathcal{I}$.' Any theory of knowledge and reference that is incapable of distinguishing truth from truth under $\mathcal{I}$ is committed to the implication that θ is true if θ is true under $\mathcal{I}$. But modulo our assumption about cardinality, that θ is true under $\mathcal{I}$ is a matter of model theory. $\mathcal{I}$ is arbitrary; the construction which employs it is clearly unacceptable, since it trivializes the question whether θ is true. Any account of our theoretical knowledge that cannot exclude $\mathcal{I}$ as an interpretation of the language which adequately captures the interpretation under which we suppose that θ is true, cannot account for our naive confidence in the belief that our theories, if true, contain significant theoretical truths about the world. By equating truth with truth under $\mathcal{I}$ we rob our knowledge of the truth of our theoretical claims of its a posteriori character: modulo a single assumption about cardinality, the theoretical statements of an empirically adequate theory come out true as a matter of metalogic. But we take the truth of θ over its intended domain to be a significant truth, not one that is ensured by what is virtually a purely logical argument. Since there is nothing in Carnap's reconstruction to exclude the identification of truth with truth under $\mathcal{I}$, an essential feature of the truth of our theories has been lost, and Ramsey's and Carnap's reconstructions cannot therefore be judged successful.[18]

The conclusion we have just reached was partly anticipated by John Winnie (Winnie 1967).[19] Winnie's emphasis is, however, different from ours, since for him the salient point is that if we are given a 'physical' interpretation under which θ comes out true, it is virtually always possible to find another, as it happens abstract, arithmetical and unintended, interpretation that also makes θ true. The problem Winnie emphasizes is therefore one of finding conditions that will

[18]Notice that the observation on which the 'model-theoretic argument' depends is quite elementary. In particular, the argument does *not* appeal to any major metalogical result, let alone anything so sophisticated as the Löwenheim–Skolem Theorem. One can formulate things so that the argument appears to require the model-existence lemma, but this is misleading since the semantic consistency of θ would certainly be taken for granted by Carnap–Ramsey and therefore hardly needs to be derived. Notice also that, as presented here, the argument is not a permutation argument; such arguments are used by Putnam in other contexts (see note 19 below), but they are not essential to our concerns here.

[19] The existence of an arithmetical model is the content of Winnie's second theorem and the main focus of his paper. His first theorem shows that, modulo a trivial restriction (noted below), if $\mathcal{W}$ is a model of θ, so is $\mathcal{W}^*$, where $\mathcal{W}^*$ is like $\mathcal{W}$ except for the interpretation of some T-predicate. Winnie's discussion assumes that the domain of any model of θ is the disjoint union of subdomains W_U and W_O of (respectively) unobservable and observable entities, with the interpretation of the T-predicates restricted to W_U. The trivial restriction is that if a T-predicate T is monadic, there is a u in W_U such that u is not in the interpretation $T^{\mathcal{W}}$ of T, and if T is n-adic, there is a u in W_U such that u is not a component of some n-tuple in $T^{\mathcal{W}}$. Putnam (Putnam 1981, p. 217) appeals to a permutation argument of this sort, without, however, correctly identifying the conditions under which the argument is valid. For a counterexample to Putnam's claim see (Keenan 2002, Section 1).

exclude such unintended interpretations without compromising the assumptions of the framework within which the reconstruction is expressed. However, this way of presenting the difficulty is misleading. The Carnap–Ramsey reconstruction is unacceptable because it implies that the existence of an *abstract* model for θ suffices for its truth over θ's intended domain. What the reconstruction misses is a preanalytic intuition that governs our conception of truth, and this is missed even when we set aside any difficulty in fixing θ's intended domain.

5 Extension to Constructive Empiricism

There is a common misperception of the point of Newman's observation, one for which Putnam is largely responsible. The misperception is that the interest of the preceding 'model-theoretic argument' must stand or fall according to how successfully it refutes 'metaphysical realism.' Perhaps there is an interesting position of this character which the argument refutes. But to show this we would have to engage in a rather involved investigation into the nature of realism, and should we fail to find a plausible version of metaphysical realism to serve as a suitable target for the argument, we might be led to suppose that the observation on which it is based fails to show anything of philosophical interest.[20] Nothing could be further from the truth. In addition to its relevance to the Carnap–Ramsey reconstruction just reviewed, there is an 'epistemology of science' to which Newman's observation applies virtually directly. Bas van Fraassen's 'constructive empiricism' is essentially characterized by a central tenet and a pair of definitions. The *central tenet* is that it is always more rational to accept a theory as empirically adequate than to believe it true. The *definitions* are: (*i*) that to *accept a theory as empirically adequate* is to hold that 'the phenomena' form a substructure of a model belonging to the class of models that

[20]See, for example, (Chambers 2000).

is the theory; and (*ii*) *to believe a theory true* is to hold that it contains the world among its models.[21] What distinguishes constructive empiricism from its less circumspect and non-empiricist opposition is its agnosticism regarding truth and the fact that the distinction between what is and is not observable is not drawn on the basis of vocabulary, but concerns the demarcation of substructures of the models which comprise the theory. Constructive empiricism does not deny that theories—i.e., the models theories comprise—contain unobservables in their intended domains; it is merely agnostic regarding a theory's claims about them. What it holds in common with Carnap–Ramsey—when formulations are transposed to the framework preferred by the semantic view of theories—is the notion that to assert a theory as true is to say that *there is a model* belonging to the theory (recall that for van Fraassen a theory is identified with a class of models) that 'corresponds' to the world:

> My view is that physical theories do indeed describe much more than what is observable, but that what matters is empirical adequacy, and not the truth or falsity of how they go beyond observable phenomena .... To present a theory is to specify a family of structures, its *models*; and secondly, to specify certain parts of those models (the empirical substructures) as candidates for the direct representation of observable phenomena. The structures which can be described in experimental and measurement reports we can call *appearances*; the theory is empirically adequate if it has some model such that all appearances are isomorphic to empirical substructures of that model. (van Fraassen 1980, p. 64)

But so long as a theory consists of empirically adequate models with domains of the right cardinality, constructive empiricism is committed to the view that a theory is true

[21] See (Ladyman 2000) and (Alspector-Kelly to appear).

provided that it is empirically adequate. The issue is not constructive empiricism vs realism but whether the framework within which constructive empiricism is expressed—the semantic view of theories—successfully recovers the intuitive sense we attach to ascriptions of truth to our theoretical claims. Constructive empiricism supposes that it has captured this sense, and proceeds to oppose the belief that our theories are true. The difficulty is that on its own account of the truth of theoretical claims that go beyond the phenomena there is virtually nothing to choose between truth and empirical adequacy. This is contrary to the idea—evident to preanalytic intuition and, apparently, to constructive empiricists—that it is a substantial philosophical commitment to believe a theory true rather than to accept it as merely empirically adequate. What is in dispute is not whether we should believe our theories true—although this is of course an important issue—but whether constructive empiricists have captured what such a belief consists in, since, using our earlier construction, we can always ensure that any theory that saves the phenomena contains the world among the models it comprises: By hypothesis, one of the theory's empirically adequate and partially abstract models is in one-one correspondence with the world or 'intended domain' of observable and unobservable entities. In exact analogy with our argument against Carnap–Ramsey, *define* the theoretical relations on this intended domain so that, for example, $T^{\mathcal{W}}$ holds of $\langle a, b\rangle$ if and only if $\langle \varphi^{-1}a, \varphi^{-1}b\rangle$ belongs to $T^{\mathcal{M}}$, where φ is one-one from M onto W and is the identity map on $M \cap W$—the observable part of M and of W. This transforms a theory that merely saves the phenomena into a true one, since it ensures that the world is a member of the class of models that is the theory. But then what has happened to the central tenet of constructive empiricism, according to which it is always more rational to accept a theory as empirically adequate than it is to believe it to be true? This appears to be based on nothing more than an agnosticism regarding the cardinality of the intended domain. Our deployment of Newman's observation thus shows that the combination 'constructive empiricism + semantic view of theories' yields an inherently unstable position.

6 Putnam's Model-Theoretic Argument and the Semantic View of Theories

It might seem as if the considerations we have reviewed derive their force against Carnap because of his commitment to what van Fraassen (van Fraassen 1989, Ch. viii, Section 6 and Ch. ix, Section 3) has called the 'syntactic' view of theories, in contrast to van Fraassen's own 'semantic' view of theories. Recall that, on the semantic view, we explicitly assume that

$$\theta \Leftrightarrow \theta \text{ is true} \Leftrightarrow \theta \text{ has a class of } \textit{intended} \text{ models,}$$

so that on this view the truth of θ is always understood relative to an 'intended' model. Indeed, on the semantic view a theory is *identified* with a class of intended models, and the truth of a theory consists in the world being one among the class of *intended* models which *is* the theory. Perhaps the restriction to intended models, on which the semantic view insists, makes it not vulnerable to the difficulties to which Carnap's syntactic account is subject.

There are however *two* senses of 'intended model' (or 'intended interpretation') at play here. The first sense is related to the distinction between what is sometimes called 'real' second order logic and the logical systems considered by Leon Henkin in his proofs of completeness. Such systems allow for general or Henkin models and this leads to a distinction between theories which hinges on the scope of the domain of the property and relation *variables* of a theory. This is the sense of 'intended model' that proponents of the semantic view emphasize when expounding their position, since unless a theory is identified with a class of intended models in this sense—thus invoking in an essential way a notion of theory that is not purely formal (but is what is sometimes called 'semi-formal')—the semantic approach's distinctive emphasis on *mathematical* as opposed to *metamathematical* methodology cannot be sustained. The semantic approach correctly observes that the use of axiomatics in foundational studies in the exact sciences is rather different from its use in metalogical investigations. Roughly speaking, in the former case we are concerned to characterize some property or other—say, the property of being a Euclidean space—and we proceed to do so by capturing ('up to isomorphism') a class of structures which exhibit the desired property. In the latter case, our interest in axiomatization is motivated by an altogether different set of considerations, such as, for example, our interest in the recursive enumerability of a particular set of truths.

Interesting as this methodological observation is, the sense of 'intended model' that needs to be addressed in order to avoid Putnam's argument is what is involved when, in describing the meaning of some fragment of our language, we distinguish one interpretation of the non-logical *constants* of this fragment as the intended one. This problem is at best only partially addressed by the issues peculiar to the interpretation of the *variables* of the underlying logic; while this fixes the domain from which the properties and relations are drawn, it leaves unsettled which relations on the domain are the subject-matter of the theory. Putnam's observation, that we can't take the intended model in this second sense to be singled out by $R(\theta)$, also applies to the semantic view's identification of a theory with a class of intended models, since this identification concerns only the first sense of 'intended model.' To address the second sense of 'intended model,' involving as it does the constant terms, would be to engage in an analysis of the meaning of the language of θ, contrary to the promise of the semantic view that it allows us to dispense with such questions altogether. Invoking the semantic view appears therefore to have brought us no closer to a satisfactory account of our theoretical knowledge.

In a recent paper, van Fraassen addresses Putnam's model-theoretic argument at some length, without however explicitly observing its bearing on the formulation of the constructive empiricist theory of theories. The core of his analysis is that Putnam has exposed a new 'pragmatic paradox,' a paradox that emerges when we question the interpretation of our own language. Van Fraassen's idea is that while we can coherently ask of a language we do *not* understand whether a predicate of that language refers to (say) green things, we cannot coherently ask of a language we *do* understand whether 'green' refers to green things, since it is implicit in the idea that 'green' is a predicate of *our* language that we know it to refer to green things.[22] The syntactic view of theories falls victim to Putnam's argument because it seeks to resolve the problem of fixing an interpretation of $\mathcal{L}(\theta)$ by purely formal means, something which the model-theoretic argument shows particularly clearly to be a mistake. By contrast, although the semantic view does not address the question of how the interpretation of $\mathcal{L}(\theta)$ is fixed, this is not a defect because the correctness of the intended interpretation is a presupposition to which the semantic view is entitled—it is a presupposition of the pragmatic background within which the models that *are* the theory are given and within which the semantic view is expressed.

Here I think it is important to be clear on the *kind* of epistemological question Putnam's argument raises regarding the intended interpretation of $\mathcal{L}(\theta)$. We may distinguish a skeptical interpretation of the question, 'How do we know that "green" refers to green things?' from an entirely different and non-skeptical interpretation of this question, one according to which we presume that we know that 'green' refers to green things and ask for the correct representation of our knowledge of this fact.[23] Clearly, under the latter interpretation, this is a question we can ask about *any* language—

[22]Putnam makes a similar observation in his original formulation of the argument, using the observation to motivate his 'internal realism.' For van Fraassen, the observation shows that we have left metaphysics and have entered the realm of pragmatics.

[23]See (Frisch 1999) for a related observation.

including languages we ourselves understand—without raising any hint of paradox, pragmatic or otherwise. Even if it is not legitimate to question *whether* the reference of one's language is correctly fixed in accordance with its customary interpretation, it is surely legitimate to ask *how* it is fixed. Carnap–Ramsey attempts to address this question by arguing for the adequacy of a reconstruction in which the theoretical vocabulary is eliminated; if such a reconstruction were successful, the question of how the reference of the theoretical vocabulary is fixed could be put to one side. What is peculiar about the Carnap–Ramsey reconstruction is that it articulates its empiricist commitments by isolating the terms of the theoretical vocabulary as those for which an explanation of reference is especially pressing. Even if we believe that Carnap–Ramsey is insufficiently critical of its account of how the reference of the observational vocabulary is fixed, it is nevertheless intelligible that it should be puzzled by how the reference of the theoretical vocabulary might be fixed: the empiricist commitment it shares with Russell's theory of knowledge and propositional understanding yields a model for addressing the reference of the observational vocabulary that has no clear application to the theoretical case. The problem thus becomes one of explaining—or explaining away—the reference of the theoretical vocabulary; hence the Carnap–Ramsey strategy just reviewed. But the semantic view simply fails altogether to address the problem.

7 The Problem Clarified and Resolved

To be significant, the claim that θ's theoretical sentences are true must be understood relative to an interpretation of the language of θ that is capable of being given independently of the truth of the theoretical sentences themselves. Preanalytically, we assume that the intended interpretation of the theoretical vocabulary is such an interpretation and that truth with respect to it is not something that can be settled by a purely logical argument. While the rational reconstructions we have been considering are perfectly adequate accounts of the deductive structure of our theories, they are unable to accommodate this feature of our ascriptions of truth. This is the central lesson of Newman's observation.

In order to understand what has gone wrong and to better see both the nature of the problem we have uncovered and its solution, it is instructive to look at Russell's formulation of what he calls the problem of 'interpretation.' The problem and its significance are explained in the Introduction to *The Analysis of Matter* as follows:

> It frequently happens that we have a deductive mathematical system, starting from hypotheses concerning undefined objects, and that we have reason to believe that there are objects fulfilling these hypotheses, although, initially, we are unable to point out any such objects with certainty. Usually, in such cases, although many different sets of objects are abstractly available as fulfilling the hypotheses, there is one such set which is much more important than the others. ... The substitution of such a set for the undefined objects is 'interpretation.' This process is essential in discovering the philosophical import of physics. (Russell 1927, pp. 4-5)]

For Russell, the point-instants of the theory of space-time pose a problem exactly similar to that posed by the numbers in Peano's axiomatization of arithmetic. Recall that so far as the Peano axioms are concerned, any ω-sequence forms the basis of a suitable model of the axioms. But among ω-sequences, there is one that is distinguished, namely the one which consists of 'the' cardinal numbers, since, as Russell says, this fulfills the requirement 'that our numbers should have a *definite* meaning, not merely that they should have certain formal properties. This definite meaning is defined by the logical theory of arithmetic' (Russell 1919, p. 10). The 'logical theory of arithmetic' associates '0' with the class of all null classes, '1' with the class of all singletons, '2' with the class of all couples, etc. Although any association with the members of a progression will succeed in giving a 'definite meaning' to the numeral names, Russell, like Frege before him, argued that among the various possible definite meanings, the one indicated is distinguished by the fact that it captures our use of the numeral names in our judgements of cardinality. The Frege–Russell cardinals are perhaps the simplest example of a successful application of the method of logical construction to a problem of 'interpretation' in Russell's sense. We would today express this by saying that the set-theoretic construction consisting of the Frege–Russell cardinals form the basis of a *representation* of any model of the Peano axioms. The abstractness of the number-theoretic axioms consists in the fact that they fail to distinguish, among all possible ω-sequences, the one which is associated with their foremost application. This is what the logical theory achieves.

The way to think of Russell's construction of point-instants is to see it as an attempt to accomplish for the theory of space-time what the definition of the Frege–Russell cardinals achieved for number theory. In each case, the axiomatically primitive notions of number and point-instant are to be replaced by something else—classes of equinumerous classes and maximal copunctual classes of events, respectively—in order to display the canonical applications of the theories in which these notions occur.[24] In the arithmetical case, as we

[24]To fully understand the program of logical construction advanced in (Russell 1927), it is necessary to look in detail at the notion of copunctuality and the 'logical constructions' the book advances. A complete study would be a large undertaking, but the basic idea may perhaps be at least indicated. In (Russell 1927) a class of events is said to be *copunctual* when every five-tuple or quintet of events in the class has a 'common overlap.' When a quintet of events has a common overlap, Russell says that the five events stand in the relation of *copunctuality* (Russell 1927, p. 299). The terminology is potentially confusing since the predicate, 'x is copunctual' refers to a *property of classes of events*, while *copunctuality* is a

have seen, the canonical application of the theory was the use of numbers as cardinal numbers. Under the influence of Eddington,[25] Russell took the canonical application of the theory of space-time to be its use in *measurement*, a view to which they were led by the fundamental role of length and time in the measurement of physical magnitudes. Since Russell's construction of point-instants in terms of events is compatible with the assumption that events comprise only finite volumes, the representation of point-instants by classes of events was motivated by the observation that although there is no bound on the precision we may achieve in any actual measurement, we are always restricted to finite quantities.

There is, however, an important difference between the

five-place *relation between events*. The use of quintets rather than some other number of events arises for technical reasons having to do with the dimensionality of the space whose points (point-instants) are being characterized as logical constructions—set-theoretic structures, as we would today say—out of events. Russell's point-instants are defined as maximal copunctual classes of events—maximal, that is, with respect to class inclusion. Assuming that there are copunctual classes of events, the success of the proof of the existence of space-time points, which occupies Chapter xxviii of (Russell 1927), turns on showing that every such copunctual class can be extended to a maximally copunctual class. Russell's proof uses Well Ordering, applied to the domain of all events, and his theorem has an evident similarity to the Ultrafilter (or Maximal Dual Ideal) Theorem for Boolean Algebras, with the property of being a copunctual class playing a role analogous to that played by the 'finite intersection property' in the context of the representation theory of Boolean algebras. Indeed, the analogy can be developed further to illuminate the difference between Russell's construction and the earlier, but more restricted, construction by Whitehead in terms of 'enclosure series.' Whitehead's construction of spatial points consists in identifying a point with a class of nested volumes ('nested' by the relation of spatial inclusion). Transposed to the spatial case for comparison, Russell's construction requires only the existence of a common spatial overlap among the volumes belonging to the class of volumes which *are* the point. Russell's notion generalizes Whitehead's enclosure series in the same sense in which the notion of filter generalizes that of a nested sequence or chain.

[25]Russell's understanding of General Relativity seems to have been largely derived from (Eddington 1924). This is not to say that Russell's reading of Eddington was uncritical; see, e.g., (Russell 1927, pp. 90-92) for an assessment of Eddington's operationalism.

arithmetical and spatio-temporal cases, one which arises from the fact that they involve applications of theories of very different character. In the context of Russell's logicism, the central question the arithmetical case raises is: 'Given a domain of individuals of the right cardinality, can we recover the structure of the numbers as a theorem of *Principia*?' As George Boolos (Boolos 1994) observed, it is a remarkable and insufficiently appreciated fact that we can. The answer is not obvious since Russell is *not* assuming that the numbers occur among the elements of the domain of individuals but at a higher type. A preliminary definition: Russell calls a class of individuals *inductive* if it belongs to every class which contains the null class of individuals and is closed under the addition of singletons (or, equivalently, if it is in one-one correspondence with the integers less than n for some integer n; this terminology is evidently motivated by the analogy with the principle of mathematical induction and the definition of the finite numbers). A class is *non-inductive* if it is not inductive. Whitehead and Russell's Axiom of Infinity states that the class of individuals or entities of Type 0 is non-inductive. On the basis of this assumption, it is possible to prove (Whitehead & Russell 1912, *124.57)—without the Axiom of Choice, and therefore one might well argue, by employing only logical modes of reasoning—that the Frege-Russell cardinals, which are entities of Type 2 in the simple type hierarchy, are 'reflexive' or *Dedekind*-infinite, and thus form the domain of a model of the Peano axioms.

Now for Russell the analysis of matter just *is* the extension of the method of logical construction to physics in general, and to the theory of space-time, in particular. Restricting our attention to the space-time case, here the successful execution of Russell's program and a satisfactory solution to his formulation of the problem of interpretation requires that every abstract model of the theory should have an isomorphic representation by one constructed in terms of maximal copunctual classes of events, where, in analogy with the use of the Axiom of Infinity in the number-theoretic case, events are presumed to comprise a countable collection of concrete individuals (for Russell, events are the 'ultimate constituents of the material world'). The program of construction requires, quite properly, and again in parallel with the number-theoretic case, that it be provable that the class of events gives rise to an isomorphic representation of any model of the theory of space-time. Thus formulated, the program of logical construction is a familiar part of the nature and methodology of representation theorems, a part which Russell understood perfectly well.

It is important not to mistake the successful execution of the program of logical construction with the vindication of the central epistemological contention of Russell's structuralism: from the fact that the representation of any model of space-time is purely structure-preserving, it might well seem to follow that the knowledge expressed by the original theory—in this case, the theory of space-time—is, in the sense appropriate to Russell's theory of knowledge, 'purely structural.' As Russell understands them, the analysis of matter and (more specifically) the problem of the interpretation of the theory of space-time, do not in any way require non-structural knowledge of spatio-temporal relations. This is not an oversight, but an essential feature of the notion of analysis in which Russell is engaged. Non-structural knowledge is not required because the task which the program of interpretation sets itself is the purely mathematical one of constructing an isomorphic representation of the space-time of one or another physical theory, a representation defined over point-instants, appropriately constructed as maximal copunctual classes of events of finite extent.

There are many obstacles to successfully deploying the method of interpretation or logical construction in aid of structuralism.[26] But the central difficulty arises from the fact that

[26]For example, there is a difficulty already at the level of point-instants, since only some of the events used in their construction are the loci of percepts. The construction requires that all of the events that go to make up a point-instant are in the field of the relation of copunctuality. Thus, even

for Russell a successful interpretation of spatio-temporal concepts is not essentially different from the analysis and interpretation of arithmetical concepts with which *Principia* is occupied. Both projects proceed relative to a non-logical assumption regarding the nature and cardinality of the domain over which the construction is effected, and in both cases the problem is to show, by purely logical modes of reasoning if possible, that a particular mathematical structure—that of the natural number system or of space-time—can be represented by a logical construction. Such a notion of interpretation addresses one way of understanding the problem of the 'applicability' of a mathematical framework, one that is resolved for Russell once the fundamental objects of the framework are constructed—in the case of space-time, when point-instants are constructed from 'percepts,' or more generally, from events of finite extent—and appropriate relations are defined over them.

Suppose we allow Russell the assumptions he requires for his construction of point-instants to succeed, and grant that *overlapping* and *copunctuality* are relations which are both perceptible and hold among events which are not percepts. Then even leaving to one side the conflict this poses for his theory of knowledge—for the structure/quality division of what we are capable of knowing regarding the material world—there remains an important difficulty with Russell's notion of interpretation, a difficulty which attaches to the adequacy of his philosophy of physics and theory of theoretical knowledge. Russell's approach makes perfect sense when the theory of space-time is conceived as the a priori or 'quasi-a priori' background for spatio-temporal measurement. But from our post-Einsteinian perspective there is more to the

the specification of the domain over which the spatio-temporal structure is to be defined implicitly presupposes the existence of non-structural knowledge of the relations among events which are not percepts. When this assumption is dropped, it is not clear whether the definition of the manifold of point-instants can be carried out without the extension of a copunctual class to a maximally copunctual class involving the addition of events which are not percepts.

theory of space-time than understanding its role in measurement. To accommodate this additional element, our analysis of the interpretation of spatio-temporal concepts must be capable of revealing the theory of space-time as an a posteriori theory of the spatio-temporal structure of the world. *Russell's notion of interpretation as logical construction fails to address this task.* And while it is certainly true that knowledge of space-time structure rests on an interpretive claim, the nature of this interpretive claim is not illuminated by Russell's notion of interpretation. To see this, notice that there are at least three notions of interpretation that are relevant to space-time theories, of which Russell might be understood to countenance only the first two.

1. There is the official notion, the one to which Russell's structuralism entitles him. Interpretation in this sense seeks to construct the domain of objects over which the theory is normally interpreted as quantifying. This is what the construction of point-instants seeks to secure in the case of theories of space-time. But not having access to the *relations* on this domain, providing an 'interpretation' reduces to the purely mathematical problem of finding some family of relations definable over the constructed objects which constitute the basis for a model of the theory. By its very nature, such an interpretation cannot succeed in illuminating the epistemic status of the theoretical claims of a theory, since there will always be some family of relations of the appropriate structure. This is in essence just Newman's observation.

2. There is a second sense of 'interpretation' that arises if we ignore the constraints of structuralism and allow that we have access to the relations of the theory in addition to having specified the domain of objects among which they hold.[27] The problem of interpretation then

[27]See (Anderson 1989) for an authoritative discussion of Russell's investigations along these lines.

consists of two tasks: the purely logical one of showing the sufficiency of a set of postulates regarding the relations of the model—Do the postulates capture the basic truths of the original theory?—and the epistemological task of motivating the 'naturalness' of the properties the postulates impose. Although interpretation in this sense is not intrinsically incapable of illuminating the epistemic status of theoretical claims—indeed it is hardly different from articulating a well-motivated axiomatization of the theory—its success turns entirely on how it is elaborated in particular cases, on the persuasiveness and sufficiency of the individual axiomatization on offer.

3. But there is an altogether different notion of 'interpretation,' one that neither of the previous notions captures. It arises when we expect interpretive claims to clarify the nature and epistemic status of our theoretical commitments in a sense that we can perhaps make clear by briefly reviewing the analysis that lies at the basis of our knowledge of the space-time structure of special relativity and its divergence from Newtonian space-time.[28]

Generally speaking, we wish to discover those criteria of application that our use of a theoretical predicate or relation-expression presupposes so that our assertions regarding the structure in which the associated relation occurs are seen to have the epistemic status that preanalytically we suppose them to have. It is a presupposition of our successful reference to theoretical relations that we know *what* the relations are that our theoretical predicates refer to. Even if we take for granted our knowledge of what relation a particular predicate refers to, this falls short of assuming everything of interest that might be asked regarding our knowledge of the reference

[28]The discussion which follows is greatly indebted to (DiSalle 2002), to which the reader is referred for more detail, applications to other aspects of the theory of space-time, and an account of the historical background to these issues.

of the predicates of our language. We have seen that whatever such knowledge consists in, it is not recoverable within the simple empiricist model that Carnap and Ramsey inherited from Russell. The problem is not addressed by a psychological investigation into the origin of our ideas, but requires a conceptual analysis of our theoretical knowledge in general, and of our knowledge of physics in particular, one that clarifies and reveals those assumptions that are implicit in our basic judgements involving the theoretical vocabulary. Assuming that theoretical predicates *do* refer to their intended referents, we wish to indicate what criteria control the *application* of such predicates, and we wish to indicate what the *conditions of adequacy* for such criteria of application might be. The main desideratum for a successful analysis along these lines—and the successful account of our understanding of theoretical predicates such an analysis would yield—is that it should imply that our theoretical claims, when true, are significant truths about the world. Neither the Carnap–Ramsey reconstruction nor the constructive empiricist alternative to it meets this condition.

The nature of the problem and the basic idea underlying our proposed solution can be made clear if we take as the predicate whose criterion of application needs to be uncovered the predicate, 'x is simultaneous with y.' This predicate is fundamental to both the Newtonian and Einsteinian cases, since even spatial measurements implicitly assume a comparison of distances *at a time*, and this requires that we should have settled on a criterion of application for simultaneity. Since simultaneity is an empirical relation among events, one which holds or fails to hold as a matter of fact, we must be able to specify a criterion in accordance with which we can say that pairs of events—including distant events—are or are not in the relation of simultaneity. To grant this is not to say that 'meaning is verification': it must be part of any view of physics that a presupposition of our use of 'x is simultaneous with y,' both in our theorizing and in the evaluation of our theorizing about space-time, is that it requires an empirically-based criterion for applying spatial and temporal predicates. We wish to know those criteria of application that enable us to extend our ordinary judgements involving simultaneity to cases which they may not originally have been designed to cover; and we wish to know whether, in the process, the criteria of application that govern spatio-temporal predicates come to be subject to any new constraints. An 'interpretation' of a space-time theory seeks to address these tasks. Let us consider more specifically what the special relativistic analysis of simultaneity contributes to our understanding of these matters.

Among the criteria of application we actually employ, it is clear that some implicitly assume a process of signaling—as, for example, when we count as simultaneous two distant events which are *seen* to coincide with some local event, such as the position of the hands of a clock. This is a criterion that is shared not only by the theoretical frameworks—special relativity and Newtonian Mechanics—we are discussing, but is also found among our commonsense criteria of application for simultaneity, given the assumption that the signaling criteria employed by our theories are merely refinements of what we deploy in pre-scientific contexts. But because it admits infinitely fast signals, the Newtonian theory allows for a *velocity-independent* criterion of application for 'x is simultaneous with y.' Within the Newtonian framework, this criterion has the further justification of being in an appropriate sense 'absolute,' since, being independent of any finitely transmitted signal, it is also independent of the relative velocity of the frame of reference, and is therefore formulable without reference to the peculiarities of the circumstances of its application.

There is, however, another criterion of application that is absolute in this same sense even though it is based on the use of *finite* signals and is therefore velocity-dependent. This is the criterion based on light-signaling that emerges from Maxwell's theory, since according to this theory, the velocity of light is the same for all 'observers'—all inertial frames. On

Maxwell's theory, a signaling criterion which employs light signals is therefore as observer-independent—as absolute—a criterion of application for 'x is simultaneous with y' as is the Newtonian criterion. But unlike the Newtonian criterion, light signaling also supports our practical criteria of application for simultaneity, insofar as they are implicitly based on some sort of finite signaling procedure. It is therefore a criterion that is much closer to what we typically rely upon in our actual judgements regarding spatial and temporal relations; but when it is embedded in the context of Maxwell's theory, it has the added advantage of being as *absolute* a criterion of application—as independent of the frame of reference—as the Newtonian one.

What is shown by the analysis of simultaneity just sketched is that there is a suitably absolute criterion, that it is based on the common practice of signaling, that it conforms with our most successful theoretical understanding of light transmission, and that under these circumstances it should be acceptable even from a Newtonian perspective. Contrary to a standard understanding of Einstein's methodology, the basis for this choice of criterion of application cannot, therefore, be regarded as founded on nothing more than a free decision. Our 'choice' is clearly very highly constrained and is a reflection of our pre-analytic practice; what is striking is that it is also a reflection of a methodology that is largely shared by both the Newtonian and the Maxwellian frameworks. But adopting the light signaling criterion implicit in Maxwell's theory has the unexpected consequence that the relation of simultaneity that it governs is 'relative'—it has the consequence, that is, that 'observers' in relative motion to one another will disagree on which events are simultaneous with each other. Most importantly, for our purposes, it has the consequence that, relative to this criterion of application, the space-time which the relation of simultaneity generates is Minkowskian rather than Newtonian. And on this reconstruction, the truth of this structural claim is correctly represented as a significant a posteriori truth about the physical world.

8 Acknowledgements

I am particularly indebted to Michael Friedman, Robert DiSalle and Peter Clark for conversations on the topics of this paper and for their advice on its final form and organization. I also wish to thank David DeVidi, Paul Humphreys, Tim Kenyon, James Ladyman, Gregory Lavers, Adele Mercier, Christopher Pincock and Edward Stabler for comments; the Philosophy Departments of Bristol, Queen's, Glasgow and USC, and the Logic and Philosophy of Science Department of the University of California, Irvine for the opportunity to present my ideas to them, and the Social Sciences and Humanities Research Council of Canada for financial support. This paper is dedicated to the memory of my friend, Graham Solomon, who died on November 1, 2001.

Chapter
SIX

DO WE HAVE THE RIGHT LIMITATIVE THEOREMS?

A.P. Hazen

My original intention was to give this as a talk to the University of Melbourne Philosophy Department Colloquium in August, 2000. As often happens, I found the hour was not long enough: trying to ensure that everything was clear to the non-logicians in the audience, I got through about half before it was time to go to lunch. Emulating Tully, I went off and wrote out the oration I wished I had delivered. [1]

1 History

The decades around 1900 saw a concentration of studies of the axiomatic method of an intensity unmatched since Aristotle. Mathematicians and logicians in Germany (Dedekind, Hilbert, Frege), Italy (Peano's school) and the United States (cf. (Scanlan 1991)) formulated axiomatic descriptions of a

[1]Which I now include in a volume dedicated to Graham Solomon. While I never met Graham Solomon, I corresponded with him by e-mail over several years. The impression I got was of someone whose love of philosophy was sincere and disinterested: someone as happy to help me with my studies as to make a name for himself.

D. Devidi and T. Kenyon (eds.), A Logical Approach to Philosophy, 128-150.

variety of mathematical systems and studied the general theory of axiom systems. Two distinct goals were identified for an axiomatization. One was *descriptive*. An axiomatization should describe a system in enough detail to specify it uniquely: the intended system should be the *only* one compatible with the axioms. This goal is a natural one, but takes some explanation. In one sense one can specify something uniquely simply by naming it: which system are you interested in? the *natural numbers* (for example). This is trivial, and not what interests axiomatizers: the specification must be via the *content* of the axiomatic description, and not simply by the convention of a name. This notion is made precise by distinguishing the (non-logical) *vocabulary* of a set of axioms from its (logical) *form*. Even if a reader is ignorant of the meaning of the descriptive terms used in the axioms (point, line, ..., number ...), the structure described by use of these terms should be characteristic enough to allow the precise identification of the subject matter. Now there is a problem: as a general fact about language, *no* description can pick out a unique set of objects independently of the interpretation of its descriptive vocabulary. Suppose the system of objects you wish to describe is that of the natural numbers, organized by the operations of addition and multiplication. If I am allowed to reinterpret your descriptive terms, I can reinterpret your *number* as referring to even numbers, your $x + y$ as referring to addition, and your $x \times y$ as referring to half the product of x with y. I will have interpreted your description as describing a *different* system of objects from the one you intended (mine excludes all the odd numbers!), but everything you can say in trying to specify your system will also hold of mine: you say, for example, that $2 \times 3 = 6$, but I agree, taking your 2 to refer to 4, your 3 to refer to 6, your 6 to refer to 12, because half of 4×6 is 12!

So the correct goal has to be a bit more circumspect. An axiomatic description ought to describe the *structure* of the intended system uniquely: there will, trivially, be other systems satisfying the description, but they must share a structure with—to use the technical term, be *isomorphic* to—the intended system. The example just given will serve. The system of *even numbers* (organized by plus and “half times”) is isomorphic to the system of *natural numbers* (organized by *plus* and *times*): there is a one–one correspondence between the objects of the two systems (an object in the system of evens corresponding to half of itself in the system of all naturals), and this correspondence—this is the defining feature of isomorphism—“preserves structure.” That is, if objects in one system are related by one of the organizing operations or relations of that system, then the objects in the other system corresponding to them will be related in the same way by the operation or relation of the other. As in our example: if three natural numbers are related by one being the product of the other two, then the corresponding objects in the system of evens will similarly be related by one being half the product of the others.

A set of axioms is said to be *categorical* if it fulfills this descriptive goal. To define it formally, an axiomatization is categorical if and only if any two systems both described by it must be isomorphic. (For a clear and philosophically literate discussion of this notion and its development, cf. (Corcoran 1980, Corcoran 1981).)

There were major successes. Veblen, Huntington, and Hilbert all gave demonstrably categorical axiomatizations of Euclidean geometry. Dedekind gave what are now known as the *Peano Postulates* for the system of natural numbers, and proved their categoricity. Cantor and Dedekind analyzed the continuum—the system of real numbers in their relation of greater to less—in ways that yielded categorical axiomatic descriptions.

1.1 Notion of Set Problematic

These categorical axiomatizations, however, were all *Higher Order*. The axioms referred, not only to the objects in the system, but to *sets* of these objects, and in proving the categoricity of the axiomatization the notion of *set* had to be treated as a logical one: two systems described by the axioms (*models* of them, to introduce one more piece of jargon) could only be guaranteed isomorphic if generalizations about *sets* of objects were interpreted, in both systems, as referring to *all* the subsets of the set of objects in the system. Now, the notion of *set* soon came to seem problematic. On the one hand, the set-theoretic paradoxes seemed to cast doubt on the coherence of the notion. On the other, there was a general epistemological consideration: we seem to have much less of an intuitive grasp of the general notion of *set* than we do of such notions as *natural number* or (Euclidean) *point*. (The latter was, though less clearly stated, probably the more important worry: the set-theoretic paradoxes do not arise when—as in the categorical axiomatizations cited—only subsets of an independently given domain of objects are considered.) It came to be felt that the system of sets was a system of mathematical objects that should itself be characterized axiomatically, not something that should be assumed as part of the logical "machinery" of axiomatization. (This line of thought leads to what are called *structuralist* approaches to set theory: cf. discussions (Parsons 1990, Lewis 1990, Lewis 1993). Ironically, the ZF axiomatization of set theory, *understood in a Higher Order fashion*, comes tantalizingly close to categoricity: cf. (McGee 1997). ... Of course, to avoid confusion, subsets of the system *sets* are usually called something else: classes, or *definite Eigenschaften*.)

1.2 First Order Axiomatizations

A refinement of the axiomatic goal gradually came to be accepted. Axiom systems should describe systems without *assuming* the notion of set. In re-interpreting the language of an axiomatic system, to determine how adequately the axioms describe the structure of the intended system, only the narrowly logical vocabulary of quantifiers, connectives, and identity was to be assumed fixed. This version of the axiomatic program is what is now called *First Order axiomatization*: axioms are thought of as sentences in a First Order logical language, and are considered to describe only those structural features of the system of objects which are common to all *models*—in the sense of the model theory of First Order Logic—of the axioms. The goal of categoricity is now much harder to attain. A First Order axiom can give a categorical description of a finite system: start the sentence $\exists x \exists y \exists z \dots$ (as many existentially quantified variables as there are objects in the system) and continue $((x \neq y \wedge x \neq z \wedge \cdots \wedge y \neq z \wedge \dots)$ (stipulating that the variables stand for distinct objects) $\wedge \forall w(w = x \vee x = z \vee \dots)$ (there are no more objects than there are existentially quantified variables), and finish with a conjunction of (positive and negated) atomic formulas in the existentially quantified variables giving—by exhaustive list—the extensions of the various predicates and operators used in describing the system. Interesting mathematical systems, however, are infinite ….

1.3 Löwenheim–Skolem

At this point we run up against the first of the great 20th century *limitative theorems*: the chastening series of metatheorems that have led modern logicians and mathematicians to accept that there are severe limits on the degree to which the goals of axiomatization can be achieved. Cantor had drawn distinctions in the previously unsurveyed wilderness of the infinite, defining different "sizes" of infinity. He showed, for example, that there were exactly as many—in a natural sense of *as many* applicable to infinite as well as to finite collections—natural numbers as there are rational numbers. Since a (non-negative) rational number is just a fraction with natural numbers as numerator and denominator, the rationals are in one–one correspondence with *pairs* of natural numbers. Pairs of natural numbers, however, can be "coded" by single numbers (given the "code number" $((m + n) \times (m + n)) + n$, the two "encoded" numbers m

and n can be calculated). Thus, in two steps, rational numbers can be one–one correlated with certain natural numbers. In contrast, in the same natural sense of "size," he showed that there are *more* real numbers than there are natural numbers: the naturals are included among the reals, so there are *at least as many* reals as naturals, but there is no possible one–one correlation between the reals and the naturals, so there *aren't* as many naturals as reals. By the Löwenheim–Skolem Theorem, any First Order axiomatization describing an infinite system of *any infinite size* can be reinterpreted to describe one with only as many objects as there are natural numbers. No First Order axiomatization, then, can *categorically* describe a system whose size is one of Cantor's higher infinities. (Systems of different sizes are not isomorphic: if there is *no* one–one correlation of the objects in one system with those in the other, then *a fortiori* there is no one-one correspondence that preserves structure.) Some—Skolem, for instance—have felt that this result raises serious philosophical problems about the interpretation of axiomatic set theory: cf. (Hart 1970, Benacerraf 1985). By itself, it would not necessarily have occasioned a major crisis in the program of axiomatics. Cantor's higher infinities were seen as far removed from ordinary mathematics (it was only much later that assumptions about the higher infinities were shown to have interesting consequences in "core" mathematics: cf. surveys in (Harrington et al. 1985)). The Löwenheim–Skolem Theorem left open the possibility that all the goals of axiomatization could be achieved if attention was restricted to systems of the size of the natural number series. In technical terms, since the counterexamples to categoricity it provides involve changing the *cardinality* (size) of the system described when reinterpreting the axioms, it left open the possibility that an interesting mathematical system could be given an ω*-categorical* First Order axiomatization: one all of whose models of the size of the natural numbers are isomorphic.

And, in fact, there are some: the system of the rational numbers organized by their ordering relation has an

ω-categorical First Order description. (In the First Order language with $\leqslant$ as its only non-logical predicate, take axioms saying

1. the domain is linearly ordered,
2. it has no first or last member, and
3. between any two objects in the domain there is a third.

The axiom system so formed is ω-categorical; its denumerable models are isomorphic to the rationals under their natural order. The result goes back to Cantor; proving it is a standard exercise of elementary model theory.) This is, however, a very simple system. It is *now* a familiar part of logical lore that no such success is to be anticipated for such richer systems as that of the natural numbers under addition and multiplication, but it took more than the Löwenheim–Skolem Theorem to make this into part of "common sense."

1.4 Goal of Deductive Completeness

So far the discussion has focussed on the *descriptive* goal of axiomatization, the axiomatic characterization of the structure of interesting systems. There is another, equally important, goal. Axioms, from Euclid's time on, have been seen as starting points for proofs. A good axiomatization of a branch of mathematics is one that allows the mathematician to deduce the theorems of that branch *logically* from the axioms. Corresponding to the goal of categoricity, then, we have the goal of *deductive completeness*. A set of axioms—sentences, in the examples we will be most concerned with, in a First Order language—is *deductively complete* if and only if *every* sentence in the language is either logically derivable from axioms in the set or logically refutable (i.e., its negation is logically derivable) from them. If the axioms describe a given system of objects, deductive completeness means that all the truths (*in the language*) about that system can be established by deriving them, logically, from the axioms.

Again, there were noteworthy successes in the pursuit of this goal. By Gödel's (1930) *Completeness Theorem* for First Order Logic, our earlier example of an ω-categorical axiomatization is also an example of a deductively complete axiomatization. Presburger, in 1929, gave a deductively complete axiomatization of the system of natural numbers as organized by the operation of addition ("Presburger Arithmetic"). Tarski, sometime before the Second World War, found a deductively complete First Order systematization of a large part of Euclidean Geometry. (*Most* of the theorems covered in a traditional high school geometry course can be translated into the language of Tarski's system. For an accessible discussion, cf. (Tarski 1959).)

Again, and famously, however, there are limitations. The greatest and best-known of the limitative theorems, Gödel's *Incompleteness Theorem* (Gödel 1931), shows that the successes just mentioned were about the best possible. To be of any epistemic use, the axioms of an axiomatization have to be recognizable as such. This condition is satisfied if we simply have a finite *list* of axioms, but many useful axiomatic systems have infinitely many: systems of axiomatic set theory are typically formulated with axiom *schemes* of comprehension, systems of number theory *schemes* of induction, and Tarski's axioms for geometry just alluded to include an axiom *scheme* expressing the continuity of space. So a more liberal criterion of usability would be that the axioms should be given in the form of two finite lists: one of particular sentences taken as axiomatic, and the second a list of *schemes*, patterns such that any of the infinitely many sentences formed by substitution of appropriate subformulas in them count as axioms. An even more liberal condition would be: it must be possible to write a computer program for determining whether or not a given sentence is an axiom. What Gödel showed was that, under even the most liberal of these criteria, no set of true axioms for "Elementary Number Theory" can be deductively complete.

Presburger Arithmetic, mentioned above as a success story, can be formulated as a First Order theory: the quantified variables are to be thought of as ranging over the natural numbers, and the non-logical vocabulary could include numerals for 0 and 1, predicates for the identity and *less than* relations, and symbols for the addition and successor (+1) functions. Presburger's result was that the system of natural numbers, organized by the relations and functions expressed in this language, has a deductively complete description. Gödel's is that if we consider the slightly richer system obtained by considering the multiplication function as well (so: add one more dyadic function symbol to the formal language), this is no longer possible. It had been clear since Whitehead and Russell's *Principia Mathematica* (1910-1913), if not since (Zermelo 1909), that the natural numbers and the addition and multiplication functions over them could be defined within apparently consistent systems of axiomatic set theory. (In technical jargon: Elementary Number Theory is *interpretable* in the systems of set theory.) So Gödel's result implies that there can be no deductively complete description of the systems of sets described by Whitehead and Russell or by Zermelo. As Tarski pointed out, the theory of addition and multiplication of real numbers can be given a geometrical interpretation in the language of his geometrical axioms, with the result that if Tarski's geometry is supplemented with an additional predicate (expressing, for example, the notion that the distance between points x and y is an *integral* multiple of that between points z and w), it too will cease to be completely axiomatizable.

It has become increasingly apparent with experience that Gödel's result is robust and absolute: it is not dependent on accidental features of the formalizations chosen, but holds for all reasonable alternatives. (Gödel, apparently, planned to follow his "On undecidable propositions …part I" with a "part II" arguing this. He didn't have to because the logical community quickly accepted the generality of his results.) Largely as a result, it is now generally accepted that the two goals of axiomatization can only be attained in connection with severely limited systems of mathematical objects.

2 Tennant

Neil Tennant has recently argued (Tennant 2000) that this acceptance should have come much earlier, that reflection on fairly basic properties of the concepts involved should have shown, even before Gödel's famous theorems, that the two goals of categoricity and deductive completeness were not jointly achievable in important cases. His central argument is a familiar one, a standard item on the syllabus of an introductory course on *model theory*.

By Gödel's Completeness Theorem—not his famous limitative Incompleteness Theorem, but the "positive" result he proved the year before—any sentence which is a *consequence* of a set of First Order axioms is *derivable* from them in the standard formal systems for First Order Logic. But a derivation is a *finite* array of formulas. Thus only finitely many axioms can figure in any one derivation. We thus arrive at a result, the *Compactness Theorem* for First Order Logic, which is purely semantic, and makes no mention of the formal proof procedure Gödel showed complete:

THEOREM 6.1 (COMPACTNESS)
If a set of First Order sentences has a given First Order sentence as a consequence, then the given sentence is a consequence of some finite *number of sentences from the set.*

Surprisingly, this gives us a negative result: Elementary Number Theory cannot be given a categorical axiomatization. Look again at the First Order language of Elementary Number Theory. Using the constant (name) for 0 and the symbol for the successor function, we can construct a term, or *numeral*, designating each natural number: 1 is the successor of 0, 2 is the successor of the successor of 0, 3 is the successor of the successor of the successor of 0, and so on. Even very weak sets of axioms formulated in this language imply that all these numbers are distinct from each other. It is also provable, in very weak axiomatic systems, that these numbers come in the right order: the successor of the successor of 0 is *less than* the successor of the successor of the successor of 0, and so on. Now add a new individual constant (name), *a*, to the language. Choose your favorite set of axioms for Elementary Number Theory, and add infinitely many new axioms: *a* is not identical to 0, *a* is not identical to the successor of 0, *a* is not identical to the successor of the successor of 0, and so on: for each natural number, in other words, add an axiom saying that the object named by "*a*" is not identical to that number. By the Compactness Theorem, no number theoretic falsehood is derivable from this enlarged set of axioms (on the assumption, of course, that your original axioms were correct!). For suppose the enlarged axiom set *did* imply a falsehood. By compactness, some finite subset of the axiom set—some finite number of your axioms together with a finite number of "disclaimers" about the identity of object *a*—would also imply this falsehood. But for any such finite set, it is possible to choose a natural number—say, the smallest one not mentioned in one of the finitely many disclaimers included—such that, if we interpret "*a*" as a name for that number, every sentence in the finite set is true. Thus no falsehood can be deduced from the finite set.

Now, by Gödel's Completeness Theorem, any *consistent* set of First Order sentences has a model. (If no model makes all the sentences true, then they logically imply an arbitrary contradiction, but to say the set is consistent is to say that no contradiction can be derived formally from it.) So there is a model making every sentence of the enlarged axiom set true. This model, *a fortiori*, is described by your original set of axioms. It is not, however, isomorphic to the system of "genuine" natural numbers. (Every genuine natural number is reached in some finite number of successor "jumps" from 0. The new model, since it must contain an object denoted by each numeral, and order these objects appropriately, will contain objects corresponding to each genuine number. It will also contain an object named by "*a*" which is not one of these, and is not reached in finitely many jumps from the object corresponding to 0.) So your favorite set of axioms for describing the system of natural numbers in a First Order language has at least two *non-isomorphic* models: the system of the genuine numbers and the new model with a "fake" number denoted by "*a*." It is thus, by definition, not categorical.

Tennant abstracts from this proof. As I have stated it, it uses the Compactness Theorem for First Order Logic, but the argument appeals to no special features of this particular logical framework. The argument applies to any framework in which every logical consequence of an axiom set is implied by some finite set of axioms. This, however, is a property of *any* framework with a usable and complete formal proof procedure. This is not a particularly mathematical fact about logical frameworks, but rather a general *epistemological* requirement: if we are to be able to use a formal proof procedure to learn what follows from the axioms, the formal derivations will have to be *finite* arrays of symbols. But this elementary piece of epistemological reflection hardly required the genius of Gödel! So, Tennant argues, the conceptual basis for the argument was available to logicians some time before Gödel's stunning theorems.

But wait! The Compactness Theorem *can* be proven without mentioning Gödel's Completeness Theorem, but it usually isn't. The natural way of presenting the subject makes the Compactness of First Order Logic a corollary to its Completeness. What's more, this is one of the cases in which History took place in the Rational order: the Compactness Theorem was first announced by Gödel as a corollary to the Completeness Theorem in his 1930 paper. So the argument, taken as an argument for the non-categoricity of First Order axiomatizations of the properties of the system of natural numbers, was *not* available until Gödel proved Completeness!

At this point Tennant abstracts again. As he presents the argument, there is only one assumption: the system of objects to be described is to be (like that of the natural numbers) denumerably infinite, and each object in the system (again like the natural numbers) should have a name (or some other designation) in the language of the proposed axiomatization. It establishes, in his version, a *disjunctive* conclusion: *either* the framework lacks the compactness property (in which case the axiomatization is not deductively complete) or it has the compactness property (and so cannot be categorical). In this form at least, he contends, the argument was available to logicians before Gödel's results, and so it should have been obvious that the twin goals of axiomatization were not both attainable.

Tennant's question, then is this: why *wasn't* this obvious? Why was it only after Gödel that people realized mathematics was not fully axiomatizable?

3 Replies to Tennant

I would like to suggest three answers to this question. The first is perhaps the closest to the actual psychological and historical explanation, but this is uninteresting. There is also an *epistemological* dimension to Tennant's question: why wasn't the joint achievability of the goals of axiomatization *obvious*?

3.1 Answer the First

Why didn't logicians find Tennant's proof earlier? Because they weren't looking for it! Frege would probably have had no difficulty in following Tennant's argument had it been presented to him, nor would Russell (who was by no means as blind to the distinction between formalized systems and their metatheory as has been claimed). Discovering it for themselves was another matter. We are familiar with compactness: for over half a century it has been a star performer on the stage of mathematical logic, in the spotlight in proofs of important theorems. (Cf. (Boolos & Jeffrey 1989, chapter 26) to see it at work in the proof of a famous result in combinatorics. This chapter is also in the 1980 second edition, but not in the first.) Because we are familiar with it, we are also fully aware of its epistemological aspect, and able to see it as an important general feature of linguistic frameworks for formal axiomatics. We are prepared to see Tennant's argument. Logicians before Gödel had never thought about compactness. They were not attuned to its possible uses. Even after the concept had been defined (in Gödel's Completeness Theorem paper), it took several years before logicians discovered that it could be used to prove interesting results. Indeed, Gödel himself, reviewing in 1934 a paper in which Skolem had proved the existence of non-isomorphic models for First Order theories of natural numbers, commented that the existence of such models was a consequence of his *Incompleteness Theorem*, but overlooked the easier argument given above (for discussion cf. Vaught's introduction, (Vaught 1986), to Gödel's review (which Tennant also cites)).

And, anyway, Tennant's weak disjunctive conclusion is not the sort of elegant result mathematical logicians try for!

3.2 Answer the Second

The two goals are not of equal interest. Even in formulating the first, descriptive, goal of categoricity, reference is made to *models*: abstract, in interesting cases infinite, systems of objects. Gödel, with his strong sympathy for philosophical *realism*, was perhaps unusually prepared to take such *platonistic* concepts seriously. More pragmatically minded logicians at the time might have been tempted to leave categoricity questions to the metaphysicians: "Give me an axiom set from which I can actually derive theorems, and leave the platonic entities in Plato's heaven!" The most active school investigating axiom systems in the 1920s was Hilbert's, and they were primarily interested in *consistency* questions. Metamathematics, for the Hilbertian, was to be pursued by *finitistic* methods, and categoricity, which refers to infinite models in its very definition, cannot even be expressed in a finitistic language. (In the introduction to the 1929 dissertation (Gödel 1930) in which he first gave his Completeness Theorem, Gödel stressed that, in proving it, he had made essential use of the (non-finitistic) "principle of the excluded middle for infinite collections." He justified this by arguing that, since the *problem* of completeness, unlike that of the consistency of various strong axiomatic systems, had not arisen in the context of "the controversy regarding the foundations of mathematics," there was no need to impose the restrictions of finitist metamathematics on the methods used in its solution.) Only later in the 1930s did the use of (infinitistic) model-theoretic concepts in the investigation of logic (in, e.g., Tarski's work on truth and logical consequence) become fashionable.

So another reason why logicians in the 1920s weren't *looking* for Tennant's proof is that it gives no information about the more important goal of deductive completeness. To see this, consider again the first version of the argument given above, in which—assuming the Completeness (and hence compactness) of First Order Logic—it proved the non-categoricity of your favorite set of axioms for Elementary Number Theory. It did not, in doing so, prove Gödel's more famous result, that your favorite axioms are not deductively complete. To see this, note that the same argument shows that Presburger Arithmetic—which is deductively complete—is not categorical. Or that even if you had taken, as your set of "axioms," the set of all true sentences in the (First Order) language of Elementary Number Theory, your axiomatization would *still* not have been categorical. (Of course, since the set of all true sentences in the language is not decidable, it is not a usable axiom set, even by our most liberal criterion.)

Note finally that your favorite set of axioms (assuming it is at least as strong as the systems of "Robinson Arithmetic," **Q**, or the even weaker system **R**, described in (Tarski, Mostowski & Robinson 1953)) has a property almost as satisfying as categoricity. The axioms are intended as a description of a particular system of objects, the intended or *standard* model consisting of the (genuine) natural numbers under the relations and functions of Elementary Number Theory. Now, every object in this system is designated by a term of the language. (We made use of this fact in the proof of non-categoricity.) Suppose we consider, not models in general, but only systems of objects given by terms of the language. (In other words: consider models with no *anonymous* objects.) Call such a system of objects *fully named*. Your axioms determine the extensions of the primitive relations, and the graphs of the primitive functions, of Elementary Number Theory over the objects of any fully named model. Any two fully named models, therefore, are isomorphic. Obviously, this is a much weaker property than categoricity. It is less obvious, however, that the difference between full categoricity and "categoricity for fully named models" would have seemed of great moment to a logician still hoping for deductive completeness.

3.3 Answer the Third

The third reason why logicians before Gödel weren't looking for Tennant's theorem is that it is false.

4 Counterexample

Consider *Simple Type Theory*, as formulated in, e.g., (Quine 1953) or section 36 of (Quine 1963), but with identity a primitive rather than defined predicate. The language of this system can be considered a form of First Order Logic, in that the logical machinery is limited to the identity predicate, the truth-functional connectives, and the quantifiers. It differs from standard, text-book, First Order languages in that, instead of a single sort of variables, it has an infinite series of distinct "alphabets" of variables: one used for quantifying over individuals, one for quantifying over sets of individuals, one for sets of sets of individuals, and so on. (For a possible confusion, see § 5, comment 1.) We assume the usual logic of the connectives, the usual logic of identity (with the further stipulation that an identity formula whose two terms are of different sorts is either false or not well-formed) and the usual logic of the quantifiers (for *each* sort of variables). As proper axioms we include Axioms of Extensionality ("Two sets with the same members are one set") and Axiom Schemata of Comprehension for every type other than that of individuals. This is a system whose essential features were familiar to logicians and students of formal axiomatics from the time *Principia Mathematica* was published in 1910–1913, though only about 1930 was it given a formulation as "clean" as that just described. *Principia* had taken it—with an additional assumption, the infinity of individuals, appealed to as needed—as the foundational framework for mathematics. (Gödel 1931) gave a clean axiomatization, added axioms in effect saying that the individuals are the natural numbers, and allowed the resulting system to stand in for that of *Principia* in establishing his Incompleteness Theorems. Only after 1930 was it replaced, as a primary object of logical study, by systems formulated in single-sorted First Order Logic: systems in which there is only a single alphabet ("sort") of variables.

Now add an axiom saying that there is some specific *finite* number of individuals—1, or 7, or your favorite number. (Formulation of such an axiom in First Order Logic with identity is, of course, a standard exercise in introductory logic courses.) The resulting system is categorical! The trick is that it is a many-sorted First Order language, and the variables of each sort range (when the system is interpreted) over *finitely* many objects. First Order axioms *can* describe a *finite* structure up to isomorphism, and our theory just describes the result of lumping infinitely many finite structures together. … In detail: our axiom stipulating that there are exactly N individuals (objects in the domain that the first sort of variable ranges over) fixes the structure of this part of a model—any model of our theory must have N objects in its first domain, so the first domains of any two models are in one–one correspondence. But now our Comprehension and Extensionality axioms allow us to derive a statement to the effect that there are precisely 2^N objects in the domain of the second sort of variables, and that these play the role of the subsets of the the first domain (via the interpretation of the membership predicate). This structure, though larger, is still finite, and our axioms describe it categorically: in any two models of the theory, the parts of the models consisting of the first and second domains and the interpretation of the membership predicate relating first-sort objects to second-sort will be isomorphic. But now the Comprehension and Extensionality axioms for higher types come into play. There must be exactly $2^{(2^N)}$ objects in the third domain, playing the roles of the subsets of the second domain, and so on. (For detailed proofs, see (Gentzen 1936, Asser & Schröter 1958).) Any two models of our theory will be isomorphic. Each will have an infinite series of domains, corresponding to the different sorts of variables in the language. Corresponding domains in the two models will contain the same, finite, number of objects: there is, therefore, a one–one correspondence between the objects in the union of the domains of one model and those in the union of the domains of the other, with corresponding objects being in corresponding domains. Further, this correspondence can be defined (inductively on the series of domains) in such a way that two objects from one model will stand in the membership relation of that model if and only if the corresponding objects in the other model stand in the membership relation of the other model.

Since it is categorical, it is (by the completeness of First-Order Logic, which holds for many-sorted as well as single-sorted variants) complete. Since it is complete, it is decidable.

Add names to the language for the finitely many objects in the domain for the first sort of variables (the "individuals"). The system now satisfies all the hypotheses of Tennant's argument: it describes a denumerably infinite system of objects, each of which is uniquely specifiable in the language of the system, and it has a complete (and hence compact) proof procedure. The lacuna in Tennant's proof has to do with the new name, *a*, added to the language in defining the "non-standard" model. Tennant tacitly assumes that axiomatic systems will be formulated in a language like modern *single-sorted* First Order Logic, in which there is a grammatical category for terms capable of referring to any of the infinitely many objects in the system. It is because of this assumption that he is able to claim that infinitely many new axioms are needed to say that the *designatum* of *a* is not identical to any of the objects in the original system, which is required if compactness is to establish the existence of a "non-standard" model. In our system there is no such category: each sort of variable ranges over only finitely many objects. Whatever sort is chosen for *a*, therefore, only finitely many "axioms" will be needed to distinguish its denotation from all the objects in the range of the variables of that sort, and the addition of these "axioms" would lead to formal contradiction.

The system described is non-trivial (though of course limited) in its mathematical expressiveness. As Russell pointed out, with only finitely many individuals you don't get all the natural numbers at any one type, but any natural number you want will exist at some type or other (so all numerical equations get the "right" truth value if you go to a high enough type). (Russell makes this point at the end of (Russell 1908); the wording is changed but the same point made in the introduction to *Principia Mathematica*.) We can go beyond simple equations: all Σ^0_1 sentences of arithmetic have translations in the language of this system, and for any true one, an appropriate translation is provable. (This is an extensive class of number-theoretic sentences, familiar to mathematical logicians, in which the only unrestricted quantification over numbers is effectively existential.)

So is my counterexample to Tennant's theorem an *interesting* one, or just a curiosity? I think we are now, post-Gödel, in a position to say it is a mere curiosity. A brief examination will reveal that the expressive power of the system is really very severely restricted.

What sort of variables one should use in translating a Σ^0_1

sentence (in order to get a translation with the right truth value) varies from sentence to sentence, and there is no general algorithm for choosing the right translation of a given sentence. Open Σ^0_1 formulas cannot in general be expressed in it. These weaknesses are not hard to spot, if one knows what to look for. We do, but only because Gödel's Incompleteness Theorem tells us that (correct) theories (in languages with classical negation operators—cf. (Myhill 1950)), capable of expressing open Σ^0_1 formulas, are not deductively complete.

5 Comments

1. *First Order vs. Higher Order.* There is a terminological tangle here which has occasioned more than its share of philosophical confusion. Simple Type Theory is often called a formulation of Higher Order Logic, so what is the distinction between a Higher Order Logic and a many-sorted First Order Theory? The language, considered syntactically, can be the same. When mathematical logicians speak of Higher Order Logic, however, they are thinking of this language as semantically interpreted in a way that assumes set-theoretic notions: it is stipulated that the second sort of variables (e.g.) ranges over *all* the subsets of the domain the first sort range over, and so on. We are considering the language, and the formal axiomatic system formulated in it, from a First Order standpoint: the only restrictions we put on what can count as a model are the usual ones defining what a model of a First Order Theory is: each "alphabet" of variables must range over some (non-empty) domain, but instead of requiring that the domains for later sorts be the power sets of the immediately preceding domain, we let them be anything which will allow the satisfaction of the axioms. In the jargon of mathematical logic, we are allowing "non-standard" or "Henkin" models: this is what it means to treat the language of Higher Order

Logic as First Order. If the domain of "individuals" (the domain for the first sort of variables) is *infinite*, the axiomatic formulation of Higher Order Logic is not categorical: Henkin models need not be isomorphic to the standard model. The point of our counterexample in this section is that, with finitely many individuals, categoricity *is* obtainable, even allowing the variety of interpretations allowed in discussing the semantics of First Order theories.

2. *Single-sorted vs. Many-sorted.* Now for the fun methodological bit. By what (Hailperin 1957) calls the Herbrand–Schmidt Theorem, any many-sorted theory can be translated into a single sorted one. The standard way of doing this involves a new, monadic, predicate to demarcate the part of the domain the variables of each old sort ranged over, but I think it might be a bit more elegant here (since we're assuming the types are disjoint) to have a single dyadic *cotypicality* predicate. Axioms:

   (a) Cotypicality is an equivalence relation.

   (b) Things which have cotypical members are cotypical.

   (c) Things which are members of cotypical things are cotypical.

   (d) Things are not cotypical with their members.

   (e) Comprehension, appropriately hedged: For all x it is the case that there is a y such that for all z cotypical with x, z is a member of y if and only if ....

   Define individuals as those things which are not cotypical with anything having members.

   (f) Extensionality for non-individuals.

   (g) Same statement as before about the number of individuals.

The many-sorted theory is interpretable in the single-sorted. (For each n, "is of type n" is definable as in § 37 of (Quine 1963).) *But it is utterly different in its methodological properties!* It is not categorical. (This can be shown by means of the standard trick for getting non-standard models by compactness, as described at the beginning of section 3—the non-standard things will belong to non-standard types: no nonstandard "set" is cotypical with any standard one.) It is not decidable. (This involves some fussy technicalities—choosing bounds large enough to allow all the terms in a sentence to denote, etc.—but in effect all Σ^0_1 formulas of arithmetic are expressible, and all and only the true ones are provable, so by Gödel's Incompleteness Theorem ....) So it is not complete. So single-sorted and many-sorted First Order Logic are more interestingly different than people sometimes realize.

3. *An interesting property of theories.* Look at the standard model of the single-sorted version a bit more closely. No object in its domain is related to more than finitely many others by either of the relations—membership or cotypicality—expressed by the primitives. By induction on the construction of formulas, it is easy to show that, for any formula containing $m + n$ free variables, any m-tuple will be related by the relation it expresses to *either* finitely many *or* to *all but* finitely many n-tuples of objects in the domain. This is a remarkable property for an interesting interpreted theory to have. (For one of its consequences, see (Gaifman 1982).) In a model of arithmetic, however, every number forms unique sums with all other numbers: the ternary relation of addition, therefore, relates it to infinitely many pairs of numbers but also *fails* to relate it to infinitely many others. Similarly, in a model of (untyped) set theory, every set is a member of infinitely many others and a non-member of infinitely many others. It follows that neither arithmetic

nor any usual system of set theory can be interpreted, as a whole and in the standard sense, in this version of type theory.

Chapter
SEVEN

EMPIRICAL NEGATION IN INTUITIONISTIC LOGIC

Graham Solomon and David DeVidi

1 Prologue

There are many reasons one might adopt intuitionistic logic for particular purposes. It is most famously advocated, of course, by the inuitionists in the philosophy of mathematics who give the logic its name, particularly Arend Heyting and others influenced by L.E.J. Brouwer. It is also well known to be the logic advocated by Michael Dummett, on roughly-speaking Wittgensteinian meaning theoretic grounds, as a sort of metaphysically neutral core to logic. It is purported to be innocent of principles of reasoning (such as the principle of bivalence) which depend on assumptions which are only legitimate if realism is correct, and so as the logic which is appropriate for a particular sort of discourse if anti-realism is correct for that discourse.[1] It is advocated by some defenders of category theoretic foundations in the philosophy

[1]The details seems to vary, but essentially this view is advocated by Dummett in many places. See, especially, (Dummett 1991). The most accessible presentation of Dummett's views on these issues is (Dummett 1993).

D. Devidi and T. Kenyon (eds.), A Logical Approach to Philosophy, 151-168.

of mathematics, because it turns out that the principles valid in all toposes form a sort of higher order intuitionistic logic.[2] Adopting intuitionistic logic rather than classical logic has been advocated as a way to defuse the paradoxes of vagueness, most famously by Hilary Putnam (Putnam 1983). See also (Wright 1987), (Wright 1992*a*), (Read & Wright 1985), (DeVidi to appear).) These are only a few of the better known examples (to us).

It is well known, though, that the move from classical to intuitionistic logic involves certain costs. One such cost is that having available *only* the logical constants of intuitionistic logic leaves one unable to express things which we manifestly can and do express all the time in ordinary language—hence intuitionistic logic alone cannot be the "logic of ordinary discourse." More importantly, though, having available only the operators of intuitionistic logic sometimes leaves one unable to express certain ideas which seem crucial *to the very subject matter under analysis.* Such cases call into question not merely the suitability of intuitionistic logic to be a "logic of everything," a role which a lot of people feel no single system of logic could play, but its usefulness even for the more specific purposes for which it is touted by its advocates. Our purpose in this paper is to address one of these limitations which has been the subject of some discussion in the philosophical literature, the inability of intuitionistic logic to express "empirical" negations, and in particular its inability to express *never*.

Let's consider a couple of examples of the sort of case we have in mind. The best known reason for advocating intuitionistic logic is that one has an antecedent commitment to some or other version of constructive mathematics. One then

[2]Perhaps the most influential advocate of this view is Bill Lawvere, one of the inventors of topos theory. However, his philosophical prose tends to be somewhat opaque. (cf. Lawvere (Lawvere 1976).) A more accessible argument along these lines is (Bell 1986). Note that the philosophical starting point for these philosophers is very different from that of the intuitionists, or of Dummett.

adopts, for familiar reasons, some strong views about the close relationship between proofs and mathematical truth. But one doesn't say that a mathematical statement is true if and only if it is *proved*, for one doesn't want to end up saying that π wasn't transcendental before 1882. Rather one says that a mathematical statement is true if and only if it is *provable*. But then being able to say what we ought to be able to say about the truth values of provable statements before they are proved requires that we postulate some other negation in addition to the usual intuitionistic negation. Consider the following passage from Michael Dummett's *Elements of Intuitionism*:

> Our reluctance to say that pi was not transcendental before 1882, or, more significantly, to construe mathematical statements as significantly tensed, is not merely a lingering effect of platonistic misconceptions; it is, rather, because to speak in this way would be to admit into mathematical statements a non-intuitionistic form of negation, as will be apparent if one attempts to assign a truth-value to "pi is not algebraic," considered as a statement made in 1881. This is not because the "not" which occurs in "…is not true" or "…was not true" is non-constructive: we may reasonably view it as decidable whether or not a given statement has been proved at a given time. But, though constructive, this is an empirical type of negation, not the negation that occurs in statements of intuitionistic mathematics. The latter relates to the impossibility of ever carrying out a construction of some fixed type, the former to the outcome, at variable times, of some fixed observation or inquiry. (Dummett 1977, P. 337)

Dummett returns to the problem, in a more general setting, in his paper "Realism and Anti-Realism" (Dummett 1993, p. 473),

> Negation …is highly problematic. In mathematics, given the [constructive] meaning of "if …then," it is trivial to explain "Not A" as meaning "If A, then 0=1"; by contrast, a satisfactory explanation of "not," as applied to empirical statements for which bivalence is not, in general, taken as holding, is very difficult to arrive at. …

Another reason one might advocate acceptance of intuitionistic logic is as part of a defense of the coherence of epistemically constrained notions of truth. An easy way to see how one could be driven to this is by considering the so-called paradox of knowability. For if a view of truth compels acceptance of the validity of the scheme $P \rightarrow \Diamond \mathrm{K}P$ (which we'll call (AR)) then given classical logic and other minimal assumptions about knowledge and possibility, one can show $P \rightarrow \mathrm{K}P$.[3] While one might want to hold that every truth is knowable, one probably doesn't want to admit that every truth is known. Presuming that one doesn't want to abandon the mentioned minimal assumptions about knowledge or the modalities, one might be tempted instead to abandon classical logic in favour of intuitionistic logic, for in intuitionistic logic one can show only $P \rightarrow \neg\neg\mathrm{K}P$, and not $P \rightarrow \mathrm{K}P$. And, since $\neg P$ in intuitionistic logic is provably equivalent to $P \rightarrow \bot$ and so expresses, in some suitable sense, that P is *impossible*, the former amounts to something like "if P is true, it is

[3]The assumptions about knowledge are that it is *factive*, i.e. that $\mathrm{K}P \rightarrow P$ is valid, and that it *distributes over conjunction*, i.e. that $\mathrm{K}(P \wedge Q) \rightarrow \mathrm{K}P \wedge \mathrm{K}Q$ is valid. (Indeed, a slightly weaker condition than factivity will do, but we leave this complication aside since it is usually taken for granted that knowledge is factive.) The assumptions about possibility is that $\neg\Box P \rightarrow \Diamond\neg P$, and that if P is provable, so too is $\Box P$ (i.e., we have the inference rule of necessitation). Then the proof can proceed as follows: assume for reductio that $\mathrm{K}(P \wedge \neg\mathrm{K}P)$. Then by distributivity we have $\mathrm{K}P \wedge \mathrm{K}\neg\mathrm{K}P$, so by simplification and factivity on the right conjunct we have a contradiction. So we can prove $\neg\mathrm{K}(P \wedge \neg\mathrm{K}P)$. By necessitation and the noted relationship between boxed negations and negated diamonds, we have $\neg\Diamond\mathrm{K}(P \wedge \neg\mathrm{K}P)$. Since $(P \wedge \neg\mathrm{K}P) \rightarrow \Diamond\mathrm{K}(P \wedge \neg\mathrm{K}P)$ is an instance of the (AR) scheme, we have $\neg(P \wedge \neg\mathrm{K}P)$ by modus tollens, which is equivalent to $P \rightarrow \neg\neg\mathrm{K}P$. In classical logic, this is in turn equivalent to $P \rightarrow \mathrm{K}P$.

impossible that it be impossible to know that P," something advocates of epistemically constrained accounts of truth are likely to be inclined to accept.

However, from the validity of $P \to \neg\neg \mathrm{K}P$ the impossibility of $P \wedge \neg \mathrm{K}P$ follows directly, and so $\neg(P \wedge \neg \mathrm{K}P)$ is provable. This might look like a very serious problem, for if $\neg$ is read classically $P \wedge \neg \mathrm{K}P$ says merely that P is true but not known, and that is something that surely ought not to turn out to be absurd—surely this would be as paradoxical as having to accept the validity of $P \to \mathrm{K}P$. The problem is not so serious as that when one remembers to read $\neg$ as the negation of intuitionistic logic, for in that case $\neg \mathrm{K}P$ says not merely that P is unknown, but that it is *unknowable*, and so $P \wedge \neg \mathrm{K}P$ is something an advocate of epistemically constrained truth *ought* to regard as absurd. But this raises another, presumably less serious problem. For given just the usual tools of intuitionistic logic and the K operator, it seems that we *cannot say* that P is *true but unknown*, nor that P is *true but will never be known*—for never being known doesn't imply unknowability. After all, a truth can remain unknown because, for instance, nobody was interested in the question of whether it was true or not at the time when the relevant information was available, rather than due to any "in principle" constraint on its becoming known. Attempts to assert the existence of statements satisfying such descriptions using these tools seem to leave us asserting absurdities instead. But presumably all participants in the debate must grant that asserting the existence of such statements is not absurd, or else the "paradox" of knowability would not have struck them as paradoxical in the first place.

Our goal is to show how one can add a second negation, compatible with intuitionistic negation but distinct from it, so that one may express the sorts of claims illustrated above. We think, in fact, that it is probably wrongheaded to try to isolate a single formal notion which could play the role of "empirical negation." We suspect that there are many distinct and useful notions of negation, most of which could suitably be described as “empirical.” However, we think it worthwhile to illustrate how one such negation can work. We therefore concentrate on a negation which captures a notion of “never,” with the idea that it can serve as an illustrative example.

Before proceeding with a brief sketch of some of the formal details, we should pause to see what is wrong with one obvious proposal. Why not combine intuitionistic negation with classical negation, allowing classical negation to play the role of “empirical negation”? The short answer is this: if P has both a classical and an intuitionistic negation, then these two negations coincide. For the classical negation of P is the unique proposition $\sim P$ such that $P \wedge \sim P$ is a contradiction while $P \vee \sim P$ is a tautology. The intuitionistic negation of P is the unique proposition $\neg P$ such that $P \wedge \neg P$ is a contradiction and $P \vee \neg P$ is as close to logically true as possible (i.e., it is $\bigvee\{x \mid x \wedge P = 0\}$). Clearly if a classical $\sim P$ is available, it will also be the intuitionistic negation of P. So if every proposition has both a classical and an intuitionistic negation, we really have one sort of negation and not two.

2 The Proposal

The goal is to introduce an additional negation operator to standard intuitionistic logic. For simplicity, we restrict attention to the propositional case. To that end, we begin with a standard propositional language with the operators $\wedge$, $\vee$, $\rightarrow$ and $\neg$, which we assume (redundantly but conveniently) to contain a falsity symbol $\bot$, and add to it the unary sentential operator $\sim$ (and, presently, a binary operator $\Rightarrow$). We denote the set of atomic formulas by **Atoms**, and define the language $\mathcal{L}$ by the usual recursive clauses for the binary connectives $\rightarrow$, $\wedge$, and $\vee$ (and, presently, $\Rightarrow$) and the unary connectives $\neg$ and $\sim$. We use P, Q and R as variables ranging over sentences of $\mathcal{L}$, while we use B and C as variables ranging over atoms.

The intended interpretation of $\sim P$ is that it will be true at some state of information if and only if there is no *actual* state of information "subsequent to" x in which P is true. We will return to the question of what "subsequent to" ought to mean here. First, we introduce the following formal machinery.

2.1 Basic Semantics

A *frame* is a triple $\langle W, \leqslant, \mathbf{A}\rangle$, where $\langle W, \leqslant\rangle$ is a Kripke frame for intuitionistic logic, i.e., W is a set and $\leqslant$ is a preorder (i.e., a reflexive and transitive relation) on W, and $\mathbf{A} \subseteq W$. The idea is that W is a set of states of information, while $\mathbf{A}$ is the set of states of information which are (ever) *actualized*. If F is a frame, an *interpretation* on F is a map from W to the power set of **Atoms** subject to the *persistence condition*: if $w \leqslant x$, then $I(w) \subseteq I(x)$. For convenience we often write $x \geqslant y$ for $y \leqslant x$, $x < y$ for $x \leqslant y \wedge x \neq y$, etc.

We write $w \Vdash P$ to abbreviate "P is *true* (or *forced*) in w." We define this notion for a particular interpretation I on a frame F as follows:

- If B is atomic, $w \Vdash B$ if and only if $B \in I(w)$.
- $w \Vdash P \vee Q$ if and only if $w \Vdash P$ or $w \Vdash Q$.
- $w \Vdash P \wedge Q$ if and only if $w \Vdash P$ and $w \Vdash Q$.
- $w \Vdash P \to Q$ if and only if $\forall x \geqslant w$, if $x \Vdash P$ then $x \Vdash Q$.
- $w \Vdash \neg P$ if and only if $\forall x \geqslant w$, $x \nVdash P$.
- $w \Vdash {\sim} P$ if and only if $\forall x[(x \geqslant w$ and $x \in \mathbf{A})$ implies $x \nVdash P]$.

These clauses are the usual ones for standard Kripke semantics for intuitionistic logic, except, of course, for the clause for $\sim$. We say that P is *valid under* I if P is forced at every $w \in W$ under I. We say that P is *valid in* F if P is valid under every I on F. We say that P is *valid* if P is valid in every frame F.

Proposition 1
If $w \Vdash P$, then $x \Vdash P$ for all $x \geqslant w$.

This is easily proved by induction on the complexity of P. This is well known to hold in intuitionistic logic, so we need only check the clause for $\sim P$. But if $w \Vdash \sim P$, and there is an $x \geqslant w$ such that $x \nVdash \sim P$, there is a $y \geqslant x$ such that $y \in A$ and $y \Vdash P$. But in that case, by the transitivity of $\leqslant$, $w \leqslant y$, which contradicts the claim that $w \Vdash \sim P$.

This sort of negation has some interesting features. We give simple informal arguments which could easily made into rigorous proofs.

Proposition 2
We note a few interesting facts about this negation.

1. *$P \vee \sim P$ is not valid.* For if $x \nVdash P$ but $y > x$, $y \in \mathbf{A}$ and $y \Vdash P$, then $x \nVdash \sim P$, either.

2. *Similarly, we can show that the other classically valid but intuitionistically invalid schemes remain invalid when one replaces $\neg$ by $\sim$.* To see that this is true in general, consider frames where $\mathbf{A} = W$, in which case $\neg$ and $\sim$ will be equivalent.

3. *$\neg P \rightarrow \sim P$ is valid.* For if $x \Vdash \neg P$, then there does not exists any $y \geqslant x$ such that $y \Vdash P$. A fortiori there is no $y \geqslant x$ which is both a member of $\mathbf{A}$ and which forces P.

4. *$\sim P \rightarrow \neg P$ is not valid.* For it is possible that $x \Vdash \sim P$ while there is some $y \geqslant x$ such that $y \notin \mathbf{A}$ and $y \Vdash P$, hence $x \nVdash \neg P$.

5. *If $x \notin \mathbf{A}$, then it is possible that $x \Vdash P \wedge \sim P$ for some P.* For possibly $x \Vdash P$, but there is no $y \geqslant x$ such that $y \in \mathbf{A}$. Thus $\neg(P \wedge \sim P)$ is *not valid*. However, $\sim(P \wedge \sim P)$ *is valid*, because $P \wedge \sim P$ cannot be true at any x which is a member of $\mathbf{A}$.

6. *Neither $\neg \sim P \rightarrow P$ nor $\sim \neg P \rightarrow P$ is valid.* For $\neg \sim P$ requires that every "later" stage have some actual later

stage in which P, something which can be fulfilled without P holding in the current state of information. Similarly, $\sim\neg P$ requires that every actual later stage have a possible, not necessarily actual, later state in which P. This once again needn't involve the present truth of P. Hence (given 2 above) there is no interesting version of double negation elimination in this system.

7. *These negations do not form a "split negation" in the sense of that term used in substructural logics.* For $P \wedge \sim P$ implies both P and $\sim P$, but does not imply $\neg(P \wedge \sim P)$. (Cf. (Restall 2000), p. 62.)

The philosophical payoff of the addition of this sort of negation is supposed to be in the extra expressive capacity it give to the language. For instance, if we adopt principles governing a knowledge operator which render $P \rightarrow \neg\neg \mathrm{K}P$ provable, $P \wedge \neg \mathrm{K}P$ will be absurd. However, this is quite compatible with the consistency of $P \wedge \sim \mathrm{K}P$—roughly, $P \wedge \sim \mathrm{K}P$ can be forced in a state of information x if for all $y \geqslant x$ such that $y \Vdash \mathrm{K}P$, $y \notin \mathbf{A}$. That is, using the heuristic reading we have given to the formal semantics, there are extensions of x in which P is known, but none of those are ever actualized. Hence in x P is true and knowable, but turns out never to be known.

2.2 Why Stop with Negation?

While the primary focus in this paper is a sort of negation, the formal framework we have proposed seems to us to raise an obvious question. Intuitionistic logic and classical logic share the same notions of conjunction and disjunction. In Kripke semantics this is reflected in the fact that whether a conjunction or a disjunction is forced at w depends only on whether its immediate logical components are forced at w or not. The existence of other states of information is only relevant to the interpretation of negations and conditionals. We've introduced a new negation by restricting the class of states of information which need to be considered. So the obvious question is whether one gets an interesting conditional by making the same restriction.

Let's use $\Rightarrow$ to designate this new sort of conditional, and $P \Rightarrow Q$ the sentence which asserts that P implies Q in the relevant sense, whatever that sense turns out to be. The clause for interpreting such sentences will then be:

- $w \Vdash P \Rightarrow Q$ if and only if $\forall x[(x \geqslant w$ and $x \in \mathbf{A}$ and $x \Vdash P)$ implies $x \Vdash Q]$.

Were the name not already appropriated for another candidate, we might suggest the name "material implication" for this notion. For what it states is that in all actual states of information extending the current one, if P is forced then so, too, is Q.

PROPOSITION 3

1. *First,* $\sim P \equiv P \Rightarrow \bot$.

2. $P \Rightarrow Q$ *implies neither* $\sim P \vee Q$ *nor* $\neg P \vee Q$*, though each of these implies* $P \Rightarrow Q$.

3. $P \rightarrow Q$ *implies* $P \Rightarrow Q$*, but not conversely.*

4. $P \Rightarrow (Q \Rightarrow P)$ *is valid, but* $P \wedge \sim P \Rightarrow Q$ *is not.* The first is clear. For the second, note that those states of information, already mentioned, where $P \wedge \sim P$ is true are not inconsistent, hence there will be Q not true there (for instance, we can set $Q = \bot$). Hence this arrow validates one of the schemes most obnoxious to relevance logicians, while invalidating the other.[4]

5. *If* $\mathbf{A}$ *is linearly ordered, then* $(P \Rightarrow Q) \vee (Q \Rightarrow P)$ *is valid. However, if we do not restrict* $\mathbf{A}$ *in this (or some closely related) way, this will not be valid. Similarly, if we impose linearity, we get both* $\neg(P \wedge Q) \Rightarrow \neg P \vee \neg Q$ *and* $\sim(P \wedge Q) \Rightarrow \sim P \vee \sim Q$.

[4] I thank Jeff Pelletier for this observation.

The philosophical worth of this implication is less clear than that of the corresponding negation. Cases in which $x \Vdash P \Rightarrow Q$ and yet $x \nVdash P \rightarrow Q$ are ones in which there are possible states of information extending x which are counterexamples to the statement $P \rightarrow Q$ (i.e., in which P is forced and Q is not), but where none of these counterexamples is ever actualized. But even if this turns out to have no profound philosophical significance in itself, the possibility of adding such a notion as $\Rightarrow$ to intuitionistic logic does bolster the suggestion that there might be many potentially useful operators compatible with intuitionistic logic.

2.3 Further Conditions on A?

Since **A** is supposed to represent the *actual states of information*, one might wonder whether it is appropriate to allow, as we have so far, that **A** be an *arbitrary* subset of W. Is there something about the concept of actuality which would license restrictions on what sort of subset can serve as **A**?

Someone might wonder, for instance, whether it would be reasonable to restrict our choices of **A** to rule out the existence of two actual states of information x and y such that, for some P, x forces P while y forces $\neg P$. The presence of such states of information in W is part of the usual Kripke semantics for intuitionistic logic. And, indeed, it must be so. For if this were not a possibility, then "weakened excluded middle," $\neg P \vee \neg\neg P$ would be a valid scheme, since the only sort of counterexample to weakened excluded middle is a state x for which there are distinct y and z with $x \leqslant y$ and $x \leqslant z$ with y forcing P and z forcing $\neg P$. Since weakened excluded middle is equivalent in intuitionistic logic to the De Morgan law $\neg(P \wedge Q) \rightarrow \neg P \vee \neg Q$, this law would also be valid. But it is hard to see how there could be two *actual* states of information, one which forces P and one which rules P out, in light of the persistence requirement. Building in such a requirement would increase the stock of validities. In particular, it would validate the schemes $\sim P \vee \sim\sim P$ and $\sim P \vee \sim \neg P$, two peculiar sorts of excluded middle, and this version of De Morgan's law: $\sim(P \wedge Q) \to \sim P \vee \sim Q$.

One way to enforce this prohibition of distinct actual states of information forcing a statement and its negation would be to require that **A** be a *chain*, i.e. that for every $x, y \in \mathbf{A}$, either $x < y$, $y < x$ or $x = y$. Does this make any difference? In the case where we require that $\leqslant$ be a linear ordering of W (and not merely of **A**), this does make a difference. In particular, *Dummett's scheme*, $(P \to Q) \vee (Q \to P)$ becomes valid. For if P is never forced at any $y \geqslant x$, then $P \to Q$ is vacuously forced at x. Similarly, of course, for the case where Q is never forced. So suppose there are $y, z \geqslant x$ such that $y \Vdash P$ and $x \Vdash Q$, and take each to be the earliest such. Then, thanks to persistence, if $z \leqslant y$, $x \Vdash P \to Q$, while if $y \leqslant z$, $x \Vdash Q \to P$.[5] A parallel argument shows that if we require that **A** be linearly ordered, then $(P \Rightarrow Q) \vee (Q \Rightarrow P)$ is valid. Note, though, that this condition is stronger than the previous one: a counterexample to this scheme would require actual states x, y, z with $x \leqslant y$ and $x \leqslant z$ with P forced in y while Q is not and Q forced in z while P is not. In such cases we needn't have $\neg Q$ (nor $\sim Q$) forced in y nor $\neg P$ (nor $\sim P$) forced in z, so such a case needn't be a counterexample to $\sim P \vee \sim\neg P$ (or $\sim P \vee \sim\sim P$).

2.4 Another Negation, Briefly

From the point of view of the present semantics, it's possible to regard both $\neg$ and $\sim$ as universal quantifications of a more basic sort of negation. That is, both $\neg P$ and $\sim P$ are generalizations of the claim that P fails to be forced in a particular state of information. We might introduce the unary operator ? into our language to capture this notion,[6] and interpret it by the clause

[5] Of course, Dummett's scheme implies weakened excluded middle and the De Morgan's law $\neg(P \wedge Q) \to \neg P \vee \neg Q$. However, the opposite implication does not hold: there are cases in which weakened excluded middle and De Morgan's law hold, but Dummett's scheme does not.

- $w \Vdash ?P$ if and only if $w \nVdash P$.

One effect of introducing this operator is that it will mean that Proposition 1 will no longer hold in general. Moreover, while $\neg P \vdash ?P$ will be valid, $\sim P \vdash ?P$ will not, since a non-actual x might force $\sim P$ even though $x \Vdash P$.

Note that it is a mistake to read $?P$ as "P is false." Instead, it would be closer to read it as "P is not (yet) established as true," i.e., as the claim that P is either false or is an open question. The usefulness of such a negation in the present context is that it allows us to express something which is crucial: namely $P \wedge ?\mathrm{K}P$ can be forced at x. In such a case P is forced at x, but is not (yet) known. This, perhaps, is something like the "non-intuitionistic negation" Dummett needs to make sense of the status of the claim "pi is transcendental" before 1882.

Notice that this negation, in spite of its invalidating of the persistence condition, is, as is $\sim$, conservative over intuitionistic logic. This is clear because the interpretations of all formulas which include only the operators of intuitionistic logic remain exactly as there were in standard Kripke semantics. The examples considered in this paper should make clear that there are many options for varieties of negation which are compatible with intuitionistic logic.

3 Understanding A and $\leqslant$

The point of introducing $\sim$ is to allow us to express certain claims that one ought to be able to express, but which are inexpressible if one has available only the usual intuitionistic operators. In particular, $P \wedge \sim \mathrm{K}P$, we suggest, is a candidate to express the claim that P is true but will never be known. Understanding the actual truth of such a sentence involves taking the ordering $\leqslant$ which is used in the Kripke semantics

[6] The notation is borrowed from chapter 5 of (Bell et al. 2001), though here note that ? is taken to be an operator within the object language rather than as a special symbol in the metalanguage.

for intuitionistic logic to interpret the logical operators, and restricting it to the set of states of information which *are actualized*. A natural question to raise at this point is how $\leqslant$, so restricted, is related to the *temporal* ordering of these states.

That something needs to be said on this matter is especially clear if one includes such predicates as K among those to which the given semantics is supposed to apply. For one certainly must grant that it often happens that there is knowledge at one time which is not possessed at a later time: Eloise and Abelard might have carried their pet names for one another to the grave, so where "Eloise calls Abelard is 'Honeybun'" was known at one time it is no longer known. But this seems to violate the persistence requirement for knowledge statements, if $\leqslant$ is understood so that $x \leqslant y$ implies that x is temporally before y.

One approach to this seeming problem is to distance K from knowledge in the usual sense that implies the need for an agent to do the knowing at the time at which the knowing is done—so that, in our example, it might turn out that the claim is question remains known even though the knowers of it are no longer around. The hope would be to produce a sense of "known" that conforms to the persistence condition. One question for this project is how this sense of "known" is related to the usual one. This is a problem of no small consequence if, for instance, one intends to defend the idea that truth is epistemically constrained, since it's presumably human epistemic states that one wants to restrict truth to if the view is to have much philosophical interest. But the case is perhaps not hopeless: perhaps we can distill a sense of knowledge which doesn't require an agent from hoary examples like "no individual knows how to build a skyscraper (since no one person has *all* the manifold skills required), but skyscrapers nevertheless get built, so how to build skyscrapers is known." It's less clear how that sort of knowledge might continue to exist if the bomb were to wipe out all sentient life from the universe, and it's not obvious that distributed knowledge is really going to help with the Eloise and Abelard case. In any case, it is presumably worries of this sort which lead many authors discussing Kripke semantics for intuitionistic logic to appeal to *idealized* agents, who, it seems, must be immortal to avoid such problems.

A more promising approach is to disentangle *temporal* order from $\leqslant$. Indeed, this seems called for in any case if one wants to defend a reasonable epistemically constrained notion of truth. Presumably it is a condition on a reasonable account of truth that there be a true statement of the form "n pedestrians crossed Weber Street at King on the 15th of March 2000" even though nobody counted them on that day and there is now no way to recover the information. For in this case, or anyway in cases of the same general type, there is no future state of information in which that true statement is known. But an advocate of epistemically constrained truth must hold it to be knowable, and so there must be a y which is $\geqslant$ the present state of information in which the statement is known. But in the case of the imagined sort, this state must be *earlier*, temporally, than the present state. Hence $y \geqslant x$ does not imply that y is later than x.

This casts some doubt on the suggestion that one might want to require that **A** be linearly ordered by $\leqslant$. For it is not clear why a situation of the following sort might not arise: the current actual state of information is one where some information from an earlier state has been lost. But at a later stage there will be other information which is gained which was not available in the earlier state. Hence there are two distinct actual state of information extending the present state, neither of which is an extension of the other. So the states in question are non-comparable, and **A** is therefore not linearly ordered.[7] However, as noted above, we needn't require that

[7]That is, if x is the current actual state of information, there may be a statement B not forced at x which is forced at y even though y is in the past due to, say, the degrading of information. If, also, there is a future z with C forced in z, B not forced in z, and C not forced in y, but these are the only atomic changes in information between the three states, we

A be linearly ordered to ensure there can be no P such that both it and $\neg P$ are forced at actual states of information.

4 Concluding remarks: Why isn't this More Complicated?

This proposal strikes us as a pretty straightforward one. At any rate, from the technical point of view it is neither particularly difficult nor especially subtle. So why isn't it in common use? Of course, this sort of "why not" question is an invitation to the worst sort of speculation. We will now accept this invitation.

One reason is made explicit by one of the most influential authors considering the range of questions we are touching upon in this paper. Timothy Williamson, in a paper called "Never say Never," uses the following piece of reasoning as a linchpin of his argument that intuitionists cannot have available the means to say that something is true but never known:

> There is reason to believe that any alternative negation must be at least as strong as $\neg$. For if $\sim$ is to count intuitionistically as any sort of negation at all, $\sim A$ should at least be inconsistent with A in the ordinary intuitionistic sense. A warrant for $A \wedge \sim A$ should be impossible. That is, we should have $\neg(A \wedge \sim A)$. By the intuitionistically valid schema $\neg(A \wedge B) \rightarrow (B \rightarrow \neg A)$, this yields $\sim A \rightarrow \neg A$. ((Williamson 1994), p. 139)

Notice that $\sim$ does not satisfy $\sim A \rightarrow \neg A$. If Williamson's argument is correct, then it would seem to follow that $\sim$ doesn't "count intuitionistically as any sort of negation at all." How should we reply?

First, we note that Williamson is directing his discussion at the question of whether *intuitionists*, by which he means

have $x \leqslant y$ and $x \leqslant z$, but neither $y \leqslant z$ nor $z \leqslant y$.

people who advocate philosophical views sufficiently Brouwerian in spirit, can have available the means to express the claim that P is true but never known. This is a distinct question from the one of whether such means are compatible with *intuitionistic logic* which, as we have noted, can be adopted for a variety of reasons besides some sort of constructivism.

How, then, does Williamson's comment fit with our proposal? Note that on our proposal $\neg(P \wedge {\sim}P)$ is *not valid*. Williamson's claim is that it ought to be valid, because if $\sim$ is a genuine negation $P \wedge {\sim}P$ ought to be a claim which it is impossible to warrant. We note two things about this. First, when one moves from the constructivist to the Kripke semantic reading of intuitionistic $\neg$, what it means for an $\neg$ sentence to be valid changes. In particular, its meaning is no longer directly tied to the possibility of a warrant in the way necessary for Williamson's argument to go through. Which brings us to the second point. If we make a (simplifying?) equation of being forced at x and being warranted at x, then there is a sense in which $A \wedge {\sim}A$ *is* impossible to warrant according to the present proposal—it could never *actually* be warranted. There are possible states of information in which both P and ${\sim}P$ are forced, but these necessarily are non-actual ones. The non-actuality of such states of information dissolves the appearance of paradox because ${\sim}P$ says, in effect, that P is not forced in any *actual* state of information. So, perhaps, in the relevant sense of "impossible to warrant" the conjunction in question is impossible to warrant, but the the intuitionistic negation of that claim doesn't follow. So it doesn't follow that $\sim$ must be at least as strong as $\neg$. Which is a good thing, because our proposal for $\sim$ is *weaker* than $\neg$.

What should we conclude? Either: (1) That Williamson is right in his characterization of what is acceptable for an *intuitionist*, and so our proposed $\sim$ is not a genuine negation for such philosophers, but that it remains a candidate negation for others who adopt intuitionistic logic for some other reason; or (2) Williamson's argument is not right, even for intuitionists, because he doesn't sufficiently consider the possibility that the relevant notion of "impossible to warrant" might change if it is *empirical* rather than logical negation that is in question. Whichever disjunct is correct, there seems to be plenty of wiggle room for philosophers who are worried about the charge that moving to intuitionistic logic makes it impossible to express things which are manifestly expressible in ordinary discourse.

On the other hand, even if it is not true that the move to intuitionistic logic requires one to give up the possibility of expressing things which, for good philosophical reasons, we need to be able to express, much philosophical work remains to be done by anyone motivated to advocate intuitionistic logic for the sorts of reasons discussed here. For instance, even if all due caution is paid to the relationship between $\leqslant$ and the temporal ordering and so on, one might worry that it doesn't make sense to have the persistence condition apply to statements involving K*P*. If knowledge of *P* requires that belief in *P* be warranted, and the common view that we can be warranted in a belief on the basis of good information but that further good information can dissolve that warrant—that is, if warrants of at least some sorts are defeasible, and can be defeated even though we were not mistaken in any of the warranting beliefs nor in taking them to warrant belief in *P*—then at least *prima facie* we ought to allow that K*P* might be true in some state of information, but that an extension of that state might mean we no longer know *P*. But these we regard as separate issues, to be taken up elsewhere.

Chapter
EIGHT

NEGATION'S HOLIDAY: ASPECTIVAL DIALETHEISM

JC Beall

What does the Liar teach us about English? According to 'orthodox' dialetheism, as espoused by Graham Priest (using his *LP*-based logic), the Liar teaches us that the negation of some true English sentence is true (and, hence, that English is underwritten by a paraconsistent logic). That lesson, in addition to being very simple, avoids the familiar expressive problems that confront its ('consistency') rivals. I am inclined to accept dialetheism, although not the version advanced by Priest. Liar-like sentences are true and false; however, they are also sentences in which negation is on holiday, in a sense to be explained. *Negation*, I suggest, exhibits a 'double-aspect'—behaving 'classically,' for the most part, but very non-classically (indeed, 'free-floating') when involved in paradoxical constructions. Some (many) philosophers think that the very meaning of 'falsity' rules out dialetheism; the double-aspect hypothesis has a nice explanation of such thinking, and, indeed, acknowledges a sense in which it is correct. In addition, the double-aspect view avoids recent objections (by Field and Shapiro) against 'orthodox' dialetheism. In this paper I present a novel version of di-

D. Devidi and T. Kenyon (eds.), A Logical Approach to Philosophy, 169-192.

aletheism—'double-aspect dialetheism,' for want of a better name. For space considerations, I assume familiarity with dialetheism and its various virtues, including the 'semantic self-sufficiency' that it affords.

1 Falsity and the Liar

What is falsity? The standard answer is that falsity is truth of negation, that falsity reduces to truth and negation; the idea is that 'is true' is our only primitive semantic predicate, that truth is our only primitive semantic 'truth value.'

In this paper, I want to take that idea seriously (that truth is our only primitive semantic value); but I also want to take dialetheism seriously—the idea that some truths have true negations. If we take both of those ideas seriously we get not *LP* (the logic usually associated with dialetheism) but, rather, a different paraconsistent logic that, as I will explain, reflects a 'double aspect' theory of negation.

The structure of the paper is as follows. § 2 rehearses (relational) classical semantics, wherein truth is the only primitive truth value. In § 3 I (briefly) discuss the familiar dialetheic lesson of the Liar, which calls for dropping the 'exclusion' constraint of classical semantics.[1] § 4, in turn, presents the semantics (and theory of negation) that results from dropping the exclusion constraint of classical semantics. The theory of negation in § 4, as will be clear, is hard to accept on its own, and, for reasons discussed in § 5, I suggest that it not be accepted on its own. My suggestion, as presented in § 6, is that negation enjoys a 'double aspect,' behaving classically in most (familiar) cases but behaving non-classically when negation is

[1] I assume familiarity with the virtues of a dialetheic response to paradox, especially with respect to familiar 'revenge' and 'expressive' problems that confront all its many rivals. A basic review of dialetheism is given by Sainsbury (Sainsbury 1995) and a thorough review is given by Priest (Priest 1987). Beall and van Fraassen (Beall & van Fraassen 2003) provide an elementary review of *FDE* and *LP*, which are typical, indeed 'orthodox,' logics associated with dialetheism.

involved in paradoxical sentences. The 'double aspect' theory of negation may be modeled via a non-monotonic (adaptive) logic that I present in § 6. Finally, § 7 mentions a few advantages that the 'double aspect' theory has over *LP*, especially concerning the notion of 'true only' or 'non-dialetheia.'

2 Relational Classical Semantics

As above, we want to take the 'reduction of falsity' seriously; accordingly, we want to recognize no primitive semantic values (truth values) beyond truth. In turn, we want to define falsity in terms of our primitive semantic value (truth) and negation. The task can be done by using 'partial functions' for our valuations.[2] For present purposes, I will use the more general notion of a relation—leaving its functional character (if it has one) to be a matter of additional constraints.

2.1 Syntax

I'll concentrate on the syntax of classical propositional logic with p_i ($i \in \mathbb{N}$) being any atomic sentence (propositional parameter) and connectives $\sim$, $\wedge$, and $\vee$. Let S comprise all sentences.

2.2 Valuations

A *relational valuation* (henceforth, valuation) on S is a relation $\mathcal{R} \subseteq S \times \{1\}$.[3] Define *exhaustive* and *exclusive* valuations thus:

[2]In mathematics, a partial function on $\mathbb{N}$ is a function that is undefined for some elements of $\mathbb{N}$. If we define the *domain* of a function f to comprise all elements for which f is defined (i.e., gives a value in its co-domain), then, by analogy with mathematics, a partial function f from $\mathcal{X}$ into $\mathcal{Y}$ is such that the *domain* of f is a proper subset of $\mathcal{X}$. That is what I mean by 'partial function' here.

[3]Of course, this amounts to taking $\mathcal{R}$ to simply be a subset of S. I stick with the product notation to make clear the kinship of this approach to those that take both truth and falsity as primitive truth values and in turn define $\sim A$ to be true if and only if A is false.

- An interpretation $\mathcal{R}$ is *exhaustive* iff for any A, either $\langle A, 1\rangle \in \mathcal{R}$ or $\langle {\sim}A, 1\rangle \in \mathcal{R}$.
- An interpretation $\mathcal{R}$ is *exclusive* iff no A is such that $\langle A, 1\rangle \in \mathcal{R}$ and $\langle {\sim}A, 1\rangle \in \mathcal{R}$.

Further terminology will be useful:

- A is *true in* $\mathcal{R}$ iff $\langle A, 1\rangle \in \mathcal{R}$.
- A is *false in* $\mathcal{R}$ iff $\langle {\sim}A, 1\rangle \in \mathcal{R}$.
- $\mathcal{R}$ *satisfies* Γ iff every element of Γ is true in $\mathcal{R}$.
- $\mathcal{R}$ *satisfies* A iff $\mathcal{R}$ satisfies $\{A\}$.
- $\mathcal{R}$ is a *model* of Γ iff $\mathcal{R}$ satisfies Γ.
- $\mathcal{R}$ is a *model* of A iff $\mathcal{R}$ satisfies A.

For simplicity I will sometimes write '$\mathcal{R} \vDash A$' to abbreviate 'A is true in $\mathcal{R}$,' and hence '$\mathcal{R} \vDash {\sim}A$' for '$A$ is false in $\mathcal{R}$.'

2.3 Admissible Valuations and Classical Consequence

Classical semantics defines an *admissible valuation* to be any valuation $\mathcal{R}$ such that

- $\mathcal{R}$ is exhaustive. (Call this *Exhaustion*.)
- $\mathcal{R}$ is exclusive. (Call this *Exclusion*.)
- $\mathcal{R} \vDash A \wedge B$ iff $\mathcal{R} \vDash A$ and $\mathcal{R} \vDash B$.
- $\mathcal{R} \vDash A \vee B$ iff $\mathcal{R} \vDash A$ or $\mathcal{R} \vDash B$.

Classical consequence $\Vdash_c$, in turn, is defined as usual:

- $\Gamma \Vdash_c A$ iff every model of Γ is a model of A.

Logical truth (valid sentence) may be defined as truth in every (admissible) valuation: $\mathcal{R} \vDash A$. Given classical semantics, an equivalent account is available: A is logically true iff $\varnothing \Vdash_c A$.

3 The Dialetheic Response to Paradox

Classical semantics, as above, allows us to treat truth as our only primitive semantic value ('truth value'); and classical logic, at least on the surface, seems to get things right, at least with respect to conjunction, disjunction, and even negation. On the other hand, various phenomena call Exhaustion and Exclusion into question. For example, vagueness seems to question the Exhaustion constraint, while the Liar seems to question the Exclusion constraint. For present purposes I will ignore the former issue; I will concentrate on Exclusion.[4]

What does the Liar teach us about English? According to dialetheism, the lesson is that some truths have true negations; in particular, 'this sentence is not true' is true and false—both it and its negation are true. Exclusion, according to the dialetheic lesson (as I advance it), needs to be rejected; there are some (admittedly peculiar) sentences that are both true and false.

While it strikes many philosophers as radical (in some pejorative sense), I think that the dialetheic lesson is quite natural. One immediate advantage of dialetheism is that it allows us to retain the foundational feature of (naive) truth: namely, the intersubstitutivity of A and $\mathsf{T}\underline{A}$, where T is our truth predicate (or, at least, our simple disquotational truth predicate). Non-dialetheic responses to the Liar are either forced to give up such intersubstitutivity or they achieve as much by reducing expressive power—requiring a richer meta-language, and so on.

The general problem, in short, runs thus: One's aim is a theory of how truth (or 'true') behaves in English—its logical behavior. In giving one's theory one attempts to avoid inconsistency by introducing some crucial semantic notion

[4]I'm inclined to think that there are 'gappy' sentences; however, I also think that we consistently assert that A is gappy—that A is neither true nor false—by invoking a gap-closing negation, in which case the issue of Exclusion arises again. For present purposes, I leave gaps aside. (Note that, as is familiar, gaps alone do not resolve the Liar phenomena.)

(context, stages, gaps, stability, or the like). On pain of inconsistency, one is forced—the 'revenge' problem—to say that the crucial semantic notion(s) is not expressible in the 'object language.'[5] The trouble is evident: What language was used to express the theory itself? The answer, of course, is *English*—a language that can express such notions. One's efforts to avoid the claim that English is inconsistent force one into denying what seems very difficult to deny: that English can express the crucial semantic notions.

Some philosophers, of course, may be happy to 'just live' with the noted (expressibility) tensions; I am not suggesting that such tension serves as a knockdown argument against non-dialetheic responses to the Liar. By my lights, the dialetheic response is simpler, and ultimately more natural than denying the apparent expressive power of English.

The advantages of a dialetheic response to paradox have been well-documented by Graham Priest (Priest 1987), and I will not review them further here. One point is worth emphasizing: that the sorts of sentence that are true and false are peculiar ones that arise merely out of grammatical necessity. There is no reason to think that anything but such peculiar, circularity-ridden sentences are true and false. Indeed, there is no reason to think that classical logic, and in particular its theory of negation, is incorrect *except* for a 'few' peculiar sentences that enter the language due only to grammatical necessity. I will return to that point. For now, I turn to further aspects of the dialetheic lesson (at least as I see it).

4 Dropping Exclusion: The Logic *P*

Let *P* (for paradox) be the logic that results from dropping only Exclusion from the classical semantics in § 2 while adding

[5]In the case of popular contextual theories, the expressibility problem is that one cannot quantify over contexts, or stages, or etc., at least not to the extent that one can in the 'metalanguage.' Representative theories and their respective expressibility problems are discussed in (among other places) (Martin 1984, McGee 1991, Priest 1987, Simmons 1993).

Double Negation:

- $\mathcal{R} \vDash A$ iff $\mathcal{R} \vDash {\sim}{\sim}A$.

The resulting logic is paraconsistent: arbitrary B does not follow from arbitrary A and ${\sim}A$. (Consider any valuation that relates both A and ${\sim}A$ to 1, but does not relate B to 1.) Accordingly, we may recognize, without the pain of triviality, sentences (e.g., Liars) that are true and false.

Let $\Vdash_p$ be the consequence relation of P, defined as per the classical case (preservation of truth). One feature of $\Vdash_p$ should be noted:

- If $\Gamma \Vdash_p A$ then $\Gamma \Vdash_c A$.

The class of admissible P-valuations properly includes the class of classically admissible valuations; hence, if truth is preserved over all admissible P-valuations, then it's preserved over the restricted classical class of such valuations. The converse does not hold; for example

$${\sim}A, A \vee B \Vdash_c B$$

but

$${\sim}A, A \vee B \nVdash_p B.$$

That said, it is easy to see that the given converse holds for the *negation-free* fragment. The chief difference between classical logic and P concerns negation; and the difference is significant.

Before highlighting some differences between classical and P-negation, the issue of Double Negation should be addressed. One might think that there is something ad hoc about adding Double Negation once Exclusion has been dropped. After all, Double Negation results from the joint work of Exhaustion and Exclusion; and since Exclusion was dropped, so too ought Double Negation be dropped. While I am sympathetic with such thinking, I think that it need not be accepted. Dialetheism, as I advance it, takes the Liar to be true and false—a true sentence A such that ${\sim}A$ is also true. Such a sentence calls for dropping Exclusion, which significantly affects negation; however, there seems to be no reason—none motivated by the Liar—that directly challenges Double Negation. Until such motivation emerges, I suggest retaining Double Negation.[6]

4.1 *P*-theory of Negation

Negation, according to *P*, exhibits many unfamiliar features. Given the motivation behind *P*, the motivation to accommodate apparently true and false sentences, one feature is quite natural and expected, specifically, that $\sim(A \wedge \sim A)$—the 'law of non-contradiction'—is not logically true. That much, as I said, is expected.[7] What may be startling are other features of negation:

- $\sim A \nvDash_p \sim(A \wedge B)$
- $\sim(A \vee B) \nvDash_p \sim A \wedge \sim B$
- $\sim A \wedge \sim B \nvDash_p \sim(A \vee B)$
- $\sim A \vee \sim B \nvDash_p \sim(A \wedge B)$

While the 'failure' of such familiar principles may be startling, it shouldn't be terribly surprising, at least on reflection. After all, it is precisely Exclusion that, when coupled with the behavior of conjunction and disjunction, ensures the classical behavior of negation. Taking truth as our only primitive semantic ('truth') value, we do not get the familiar De Morgan 'laws' without Exclusion. So, as said, the 'failure' of such 'laws' should not be terribly surprising, at least given the prior rejection of Exclusion.

[6] I should note that dropping Double Negation does not affect the chief 'double-aspect' view that I propose; it simply makes *P*-negation even less constrained than it is with Double Negation. (The extent to which it is 'unconstrained' is discussed below.)

[7] Priest and Sylvan (Priest & Routley 1989) disagree; I discuss some of their remarks in § 5.

5 But is *P*-negation Negation?

Surprising or not, the theory of negation reflected in *P* is very unfamiliar, so much so that philosophers will question whether the given negation is 'really' negation. Indeed, Graham Priest and Richard Sylvan (Priest & Routley 1989) have leveled just such objections.[8] Priest and Sylvan launch a number of arguments against the *P*-theory of negation; I will briefly discuss three such arguments.

5.1 The Law of Non-Contradiction

The first objection concerns the 'law' of non-contradiction (LNC), at least understood as $\sim(A \wedge \sim A)$, which is logically true in Priest's *LP*.[9] Priest and Sylvan begin by searching for a reason to reject LNC:

> [P]resumably any case against [LNC] will hinge on the undesirability of secondary contradictions. Conceivably, we might invoke the razor that contradictions should not be multiplied beyond necessity. However, even if this is correct (and is it?) it does not get us very far until we know what 'necessity' is. (Priest & Routley 1989, p 164)

Suppose, as I have, that some A is such that $A \wedge \sim A$ is true.[10] A 'secondary contradiction' immediately emerges if LNC is

[8]Priest and Sylvan addressed many of their objections directly against a different paraconsistent logic, namely, Da Costa's C_ω, which is slightly stronger than *P*.

[9]It is worth noting that Priest (Priest 1987) actually rejects 'another' version of non-contradiction; he rejects $\sim(\mathsf{T}\underline{A} \wedge \sim\mathsf{T}\underline{A})$. That Priest rejects the T-ful version of LNC (as it were) is perhaps telling; for in order to reject the T-ful version while accepting the T-free version (as it were), Priest must reject the intersubstitutivity of truth. On the theory that I advance, there is no need to reject the fundamental intersubstitutivity of truth.

[10]Incidentally, Priest and Sylvan call $A \wedge \sim A$ a 'true contradiction.' That tag is fine, so long as 'contradiction' is understood purely in terms of logical form, namely, as any sentence of the form $A \wedge \sim A$. Some phil-

accepted, namely, $\sim(A \wedge \sim A)$. As Priest and Sylvan note, the 'razor' against such secondary contradictions might be a reason to reject LNC, but that response, as Priest and Sylvan note, requires more work, particularly concerning 'necessity.'

Fortunately, the work on 'necessity' need not be done; the reason to reject LNC has little to do with secondary contradictions. The reason for rejecting LNC is that (so far) we have no reason to accept it, at least given our recognition of sentences that are true and false—sentences that are true and have a true negation. After all, the reason that we imposed Exclusion in our (classical) semantics is that we had no reason to think that some truths had true negations. Then came the surprising Liar, which, coupled with naive truth theory, suggested that some (admittedly peculiar) sentences are both true and false. In turn, our grounds for accepting Exclusion—and, hence, LNC—vanished when we recognized true sentences of the form $A \wedge \sim A$. Accordingly, secondary contradictions have little to do with rejection of LNC.

5.2 Mere Sub-Contraries

Along the vein of LNC Priest and Sylvan launch a further argument:[11]

> Traditionally A and B are sub-contraries if $A \vee B$ is a logical truth. A and B are contradictories if $A \vee B$ is a logical truth *and* $A \wedge B$ is logically false. It is the second condition which therefore distinguishes contradictories from sub-contraries. Now in [P] we have that $A \vee \sim A$ is a logical truth. But $A \wedge \sim A$ is not logically false. Thus A and $\sim A$ are sub-contraries, not contradictories. Consequently

osophers use 'contradiction' to mean *an explosive sentence*, a sentence A from which every sentence (logically) follows; in the explosive sense of 'contradiction' no dialetheist thinks that there are true contradictions (except trivialists, who believe that everything is true, but they are non-actual, as far as I know).

[11]In the following quote I replace 'Da Costa' by 'P.'

> [*P*-] negation is not negation, since negation is a contradiction forming functor, not a sub-contrary forming functor. (Priest & Routley 1989, p 165)

This argument, like the others, is not strong. Granted, *traditionally* A and B are contradictories iff $A \vee B$ and $\sim(A \wedge B)$ are logically true. But that is simply the *traditional* theory of negation on which, *nota bene*, $A \wedge \sim A$ is *explosive*![12] Appeal to the traditional theory of negation (and, in turn, *contradiction*) does not give reason to accept LNC once Exclusion has been dropped. That negation, according to P, is a device that forms only sub-contraries is not in itself an argument against P-negation; what is needed is reason to think that, once Exclusion is dropped, we ought still accept the traditional view that negation is a 'contradiction-forming' device.[13]

5.3 Traditional Properties of 'Real Negation'

Setting LNC aside, Priest and Sylvan advance one final argument against P-negation. In effect, the argument is that P-negation affords too few traditional inferential features, including the ones mentioned in § 4.1. Priest and Sylvan argue that such 'failures'

> show that [*P*-] negation has virtually none of the inferential properties traditionally associated with

[12]I am not saying that any theory according to which $A \vee \sim A$ and $\sim(A \wedge \sim A)$ are logically true is an explosive theory; that is plainly wrong (as Priest's *LP* shows). The point above is that *traditionally* contradictories are explosive.

[13]Note that I avoid the term 'functor' here. The reason is that if we take truth to be our only primitive semantic value, it is misleading to say that $\sim$ is interpreted as a 'truth function.' Of course, $\sim$ is taken to be a functor—defined by 'truth tables'—so as to avoid the circularity of using 'not' in one's truth conditions for $\sim$-sentences; however, I think it is fairly clear that there is no avoiding such circularity, at least if one thinks that there is no language 'richer' than English in which to give such truth conditions. The relational semantics of § 2, which are due essentially to Tarski, does not hide such circularity.

> negation .... This is a further piece of evidence suggesting that [*P*-] negation is not really negation. We have now mustered strong evidence to this effect and the case seems pretty conclusive. (Priest & Routley 1989, p. 165)

How strong is this argument? Contrary to Priest and Sylvan, the arguments from LNC (and, in turn, sub-contraries) are not strong, and so that part of the 'evidence' needs to be set aside.

What about the argument from inferential properties? That argument, I think, does not provide reason to accept LNC (or backtrack on the original rejection of Exclusion). After all, the inferential properties at issue (the ones traditionally associated with negation) are simply those that, for the most part, result from imposing Exclusion; and we have been given no reason to 'take back' our initial rejection of Exclusion.

By my lights, then, none of Priest's and Sylvan's arguments provide strong reason to reject the *P*-theory of negation. Is my proposal, then, that the *P*-theory of negation, *on its own*, does adequately characterize negation? Not quite.

5.4 The Double-Aspect Hypothesis

The situation, as I see it, is as follows. A dialetheic response to the Liar is a natural and simple move; and with that move it is natural to reject Exclusion. Given that truth is our only primitive 'truth value' a rejection of Exclusion yields *P*. The trouble, as Priest and Sylvan note, is that the *P*-theory of negation yields very unfamiliar inferential properties—very abnormal features.

Given the noted abnormalities, one is tempted to think, as Priest and Sylvan suggest, that *P*-negation—and, generally, any 'negation' that doesn't satisfy the traditional features—just *isn't* negation. But that temptation, I think, ignores an interesting option: that negation has a *double aspect*. The idea is that, on one hand, negation *normally* behaves according to Exclusion, normally exhibits 'classical behavior'; on the other hand, when it is involved in a paradoxical construction, negation exhibits freedom from Exclusion—it is in many ways 'free-floating,' as witnessed in *P*. To be sure, its 'free-floating' behavior still strikes us as odd, but that is simply because such behavior is *rare* and *abnormal*. Indeed, the motivation for rejecting Exclusion came from a very abnormal source—paradoxical sentences that, except for grammatical necessity (or, perhaps, the odd contingent twist of affairs), play little role in our normal inferential practices involving negation. And it is important to recall that our 'intuitions' about 'normal negation-behavior' are formed by familiar cases; they are not formed by the odd Liar-like sentences.[14]

My suggestion, then, is that negation enjoys a double-aspect. Normally, negation behaves classically; however, when it is involved in 'paradoxical set-ups' negation exhibits the 'free-floating' behavior reflected in *P*. The task of the next section is to make the double-aspect theory precise.

6 Double-Aspect Negation: The Logic *AP*

The double-aspect theory of negation may be formalized via a non-monotonic logic, specifically, what Diderik Batens (Batens 2000) calls an *adaptive logic*. I dub the target logic '*AP*' (for *aspectival P*).[15] I will present the core ideas, and then the target philosophical interpretation.

Aspectival Negation: *AP*

Let $\mathcal{R}$ be any admissible *P*-valuation (as per § 4) and p any atomic sentence (propositional parameter). Define $\mathcal{R}^\star$, an *inconsistency measure of* $\mathcal{R}$, as follows:

[14]This is in large part why the paradoxical sentences are so surprising upon discovery: they show us exceptions to the norm, as it were.

[15]I learned of Batens' work after formulating *AP*. As far as I can see, the logic that I dub '*P*' is slightly stronger than what Batens calls '*CLuN*,' and my target logic, namely *AP*, is slightly stronger than what Batens calls '*ACLuN2*.' While I do not want to needlessly proliferate names, I will none the less stick to my original tags.

- $\mathcal{R}^\star = \{p : \mathcal{R} \vDash p \text{ and } \mathcal{R} \vDash {\sim}p\}$

Next, define the following relation on admissible *P*-valuations:

- $\mathcal{R}_i \prec \mathcal{R}_j$ iff $\mathcal{R}_i^\star \subset \mathcal{R}_j^\star$

We say that $\mathcal{R}_i$ is *less inconsistent than* $\mathcal{R}_j$ iff $\mathcal{R}_i \prec \mathcal{R}_j$. In turn, define a *minimally inconsistent model* (MI-model) of Γ thus:

- $\mathcal{R} \vDash_{mi} \Gamma$ iff $\mathcal{R} \vDash \Gamma$ and if $\mathcal{R}_i \prec \mathcal{R}$ then $\mathcal{R}_i \nvDash \Gamma$

Intuitively, $\mathcal{R}$ is an MI-model of Γ just if it is a model of Γ and any (admissible) valuation less inconsistent than $\mathcal{R}$ fails to be a model of Γ. The idea, in effect, is that MI-models 'seek' the least inconsistent way to model (satisfy) a set of sentences; in particular, any classical model of Γ will be an MI-model of Γ. (Recall that the admissible *P*-valuations include the class of classically admissible valuations.)

The logic *AP* results from defining consequence over minimally inconsistent models:

- $\Gamma \Vdash_{ap} A$ iff any MI-model of Γ is a model of A.

Logical truth (valid sentence) may be defined as truth in every (admissible) valuation: $\mathcal{R} \vDash A$, for all admissable $\mathcal{R}$.[16]

6.1 Non-monotonicity

That *AP* is non-monotonic is plain:

$${\sim}p, p \vee q \Vdash_{ap} q$$

but

$$p, {\sim}p, p \vee q \nVdash_{ap} q$$

[16]Suppose that one defines *logical truth* thus: A is logically true in *AP* iff $\varnothing \Vdash_{ap} A$. That does not work. On that account, A may be 'logically true' without being true in every admissible valuation. Consider ${\sim}(p \wedge {\sim}p)$, which is not true in every admissible $\mathcal{R}$, but which does follow from $\varnothing$ in *AP*. (Every admissible $\mathcal{R}$ is a model of $\varnothing$, and any classical model is an MI-model. So, that $\varnothing \Vdash_{ap} {\sim}(A \wedge {\sim}A)$ is not surprising but, as said, it is insufficient for logical truth.)

What is important to note is that the consequences of Γ in *AP* are precisely the 'standard,' classical consequences of Γ *if* Γ is consistent; otherwise, the consequences reflect the free-floating *P*-behavior of negation, which results only when p is both true and false.

6.2 The Philosophical Import

The philosophical import of *AP* is the double-aspect hypothesis: that negation is aspectival. Recall the lesson of the Liar. According to dialetheists, the lesson is simply that some sentences are true and false. That, by my lights, is correct, as far as it goes; however, that lesson is incomplete. The lesson is not only that there are sentences that are true and false; the lesson is that *negation* is aspectival. *AP* is intended to record the aspectival nature of negation.

The idea is that *except* for the odd paradoxical sentences—sentences that arise out of mere grammatical necessity (or the rare twist of contingent affairs)—negation behaves precisely as philosophers have always taken it to behave, namely, classically. When it is involved (by grammatical necessity or the rare twist of contingent affairs) in a paradoxical construction—a sentence that, due to the overall workings of the language, has no way of being true without thereby also being false—negation exhibits free-floating behavior. Truth remains our only primitive semantic value; and falsity remains a derivative notion, defined as always in terms of (aspectival) negation and truth.

What is attractive about the double-aspect hypothesis is that it respects many 'intuitions' about negation. Consider, for example, the standard intuition that LNC is a logical truth. That intuition, like most classical intuitions, is founded on the *normal* behavior of negation. The Liar, of course, is abnormal; it is linguistic residue of grammar. But when the Liar is taken seriously, one soon finds—on pain of expressive difficulties—that the Liar challenges our classical intuitions. The double-aspect hypothesis is that those classical intuitions needn't be rejected so much as slightly modified: one needs to recognize that negation enjoys a double life, as it were. LNC (contrary to Priest and Sylvan) is not valid simpliciter; however, it is valid over the restricted class of sentences on which our classical intuitions are built. The same considerations apply to all classical intuitions about negation: they are correct over the vast range of (non-paradoxical) sentences on which they were formed. But the vast range of sentences on which our classical intuitions were formed does not exhaust the entire range of sentences. When paradoxical constructions are at hand, negation behaves oddly; it exhibits behavior that we rightly think to be strange—it *is* strange, at least in the sense of being abnormal, deviating from the normal sentences on which our intuitions are built.

Another virtue of the double-aspect hypothesis is that it explains why philosophers cringe—or wield an incredulous stare—at the dialetheic response to paradox. Many philosophers simply declare (sometimes with a fist-thump) that no sentence can be both true and false; they simply declare that *the very meaning of falsity* rules out dialetheism. The double-aspect theory affords a nice explanation of such declarations. After all, 'is false' is *normally* associated with the *normal* behavior of negation! And in that *normal* aspect of negation the given declaration is perfectly correct: no sentence can be true and false, where 'is false' is restricted to truth of negation *as negation normally behaves*. Once the two aspects of negation are distinguished, dialetheists may (and should) join in the given declaration.

In the end, then, *AP* is intended to reflect the aspectival nature of negation (and, derivatively, falsity). We retain a single, primitive 'true value' while, for reasons having to do with paradox, rejecting Exclusion. But the rejection of Exclusion, according to the double-aspect hypothesis, is not so much a rejection of our 'exclusive intuitions' as it is a modification of our negation-theory. Exclusion still holds over the vast range of sentences with which we normally reason; it fails only for odd constructions that cry out for a free-floating negation—

and, hence, a free-floating notion of falsity. *AP* is intended to reflect just such play between 'normal' and 'free-floating' falsity.

7 Further Virtues of Double-Aspect Dialetheism

I have already mentioned some of the explanatory virtues of the double-aspect hypothesis, particularly concerning classical intuitions about negation and falsity. My aim in this section is to (very briefly) indicate a few advantages that double-aspect dialetheism has over its 'orthodox' rival—where its orthodox rival is Priest's *LP*-based version.

7.1 Field's Criticism

Hartry Field (Field 2003) recognizes the immediate advantage that a dialetheic response to paradox affords, namely, retaining the fundamental intersubstitutivity of truth. Let T and F be our truth and falsity predicates. Given the intersubstitutivity of truth and the mere derivative 'nature' of falsity, $\sim\mathsf{T}\underline{A}$ is equivalent to $\mathsf{T}\underline{\sim A}$, which is equivalent to $\mathsf{F}\underline{A}$.

Dialetheism is the view that some sentences are true and false. That is the way 'dialetheism' has usually been defined; and that is precisely the view that dialetheists hold. Field's criticism is that, given the intersubstitutivity of truth (and the resulting equivalences with falsity), dialetheists are stuck in a very odd situation. Field advances two criticisms, both aimed at oddities involved with typical ways of characterizing dialetheism. The first criticism runs thus:

> [A] problem with defining dialetheism as the doctrine (D) *that certain sentences are both true and false* is that while a dialetheist should certainly assert
>
> i) $\mathsf{T}\underline{A} \wedge \mathsf{F}\underline{A}$

> for certain *A* (e.g. the Liar sentence), he should deny this [i.e., assert the negation of it] as well. For the dialetheist asserts both $\mathsf{T}\underline{A}$ and $\mathsf{F}\underline{A}$. But from $\mathsf{F}\underline{A}$ we get [via the noted equivalences] $\sim\mathsf{T}\underline{A}$ ...; so
>
> ii) $\mathsf{F}\underline{A} \wedge \sim\mathsf{T}\underline{A}$
>
> which surely entails the negation of (i).
>
> ... Of course, it is a consequence of dialetheism that some sentences are both true and false, and there's no particular problem in the fact that the particular sentence (D) is among them. But what is odd is to take as the doctrine that defines dialetheism something that the dialetheist holds to be false as well as true. (Field 2003)

Does (ii) 'surely entail' the negation of (i)? On the orthodox (single-aspect) version of dialetheism, underwritten by *LP*,[17] the answer is *yes*. On the double-aspect view, underwritten by *AP*, the answer is *no*.

Field's faulty presupposition, then, is that negation has a single aspect. The standard definition of 'dialetheism,' Field's (D), is unproblematic given the double-aspect character of negation and, in turn, falsity. Indeed, it is precisely (and only) when some *A* is both true and false that the given entailment—which involves normal De Morgan properties—fails. A double-aspect dialetheist, unlike Priest, need not hold that her definitive doctrine is false.

Field's second criticism is the observation that

> it is misleading to characterize the dialetheist's attitude towards, say, the Liar sentence as the view (i) *that it is both true and false*, when one could equally well have characterized it as the view (iv) *that it is neither true nor false*, or as the view (ii)

[17] Or, for that matter, underwritten by Priest's own adaptive logic 'minimally inconsistent *LP*' (Priest 1991).

> *that it is false and not true*, or the view (iii) *that it is true and not false*. (Field 2003)

Once again, Field is presupposing that negation behaves normally; for the various listed ways of characterizing the dialetheist's attitude are 'equally good' only if normal De Morgan principles hold. Targeted, as it is, against orthodox dialetheism Field's observation is important; however, the observation is off the mark against double-aspect dialetheism.

7.2 Shapiro's Challenge

Dialetheism, as I mentioned, is a very attractive response to paradox because (among other things) it avoids the expressive problems that perennially confront its rivals. Stewart Shapiro (Shapiro 2004) challenges that (alleged) virtue of dialetheism.

Shapiro agrees that dialetheism avoids the usual expressive problems confronted by non-dialetheic rivals; he agrees that it does not need to invoke an 'essentially richer metalanguage' or forbid 'quantification over contexts' or so on. (After all, the motivation for such familiar restrictions is to avoid having to recognize that some sentences are true and false.) But while dialetheism avoids such typical problems, Shapiro contends that it confronts at least an analogous expressibility problem; specifically, the dialetheist has no way of expressing the apparently important notion of a *non-dialetheia*—a sentence that is not both true and false.

Shapiro's paper is rich in its discussion and I will not attempt to address all of his arguments here; indeed, I will simply isolate one of his arguments and indicate how double-aspect dialetheism avoids the criticism.[18] Shapiro's challenge

[18]Of course, Shapiro's criticisms, like Field's, are directed at orthodox dialetheism. Like Field, he recognizes the apparent virtues of dialetheism but finds its orthodox version wanting. Part of my aim in this paper is to put double-aspect dialetheism on the table, especially since, by my lights, it is a much more natural version of dialetheism, one that respects and explains 'classical intuitions' while affording the fruits of dialetheism, in general.

runs thus:[19]

> There are a number of (perhaps non-equivalent) ways to indicate that a given sentence A is a dialetheia. In the 'object language' (so to speak), one can just assert $A \wedge {\sim}A$. Or one can say that A is true and false ($\mathsf{T}\underline{A} \wedge \mathsf{T}\underline{{\sim}A}$) or that A is true and not true ($\mathsf{T}\underline{A} \wedge {\sim}\mathsf{T}\underline{A}$). But how can one say that A is a non-dialetheia? It will not do to simply say ${\sim}(A \wedge {\sim}A)$. For this last is a logical truth in Priest's semantics [i.e., 'orthodox dialetheism']. It holds no matter what sentence A is. Priest points out in several places that if A is a dialetheia, in the sense that $A \wedge {\sim}A$ is true, then ${\sim}(A \wedge {\sim}A)$ is another dialetheia. That is, we have both $A \wedge {\sim}A$ and ${\sim}(A \wedge {\sim}A)$. So if 'A is a non-dialetheia' is defined as '${\sim}(A \wedge {\sim}A)$,' then every sentence is a non-dialetheia, including every dialetheia. (Shapiro 2004)

That Shapiro's challenge, at least as formulated above, does not apply to double-aspect dialetheism is plain. As Shapiro makes clear, the problem with defining 'A is a non-dialetheia' as the falsity of $A \wedge {\sim}A$, and so the truth of ${\sim}(A \wedge {\sim}A)$, is that every sentence is thereby a non-dialetheia, including every dialetheia. But that problem (as Shapiro goes on to note) is not a problem if LNC is invalid, and it is invalid in *AP*.

There is more to Shapiro's challenge. Let A be a dialetheia. Then, by definition, $A \wedge {\sim}A$ is true. But, at least given *AP*, it is 'logically possible' that ${\sim}(A \wedge {\sim}A)$ is also true—at least where 'logical possibility' is defined in terms of *AP*-models. So, some dialetheia might also be a non-dialetheia, and hence the given definition of 'non-dialetheia' is not exclusive, in the sense that A might be both a dialetheia and a non-dialetheia.[20]

[19] I change Shapiro's symbolism for uniformity's sake. Nothing hinges on the change.

[20] Note that there is an *AP* 'trivial model,' the model in which every sentence is true (and, hence, also false). That feature is shared with

Is the non-exclusivity of 'non-dialetheia' a problem? I think that it is not a problem, at least for the double-aspect view. After all, the 'non' in 'non-dialetheia' will itself be aspectival, in as much as it is negation, which is aspectival. Accordingly, if (as the grammar always seems to ensure) there are some constructions in which 'non' in 'non-dialetheia' forces abnormal (paradoxical) behavior, then the overall intersection of 'is a dialetheia' and 'is a non-dialetheia' will be non-empty.[21] That said, the double-aspect hypothesis does not (as far as I can see) *force* the claim that 'non-dialetheia' is actually non-exclusive, that there are dialetheia that are also non-dialetheia. To be sure, in as much as *AP*-models are taken to represent 'logical possibility,' there is (thereby) the logical possibility of such non-exclusivity; however, as far as I can see, there is no need to take the given models in that way.[22]

For present purposes, the important point is that double-aspect dialetheism, unlike its orthodox single-aspect version (represented in *LP* or its adaptive counterpart), seems to avoid Shapiro's challenge. Indeed, given *AP*, one can successfully express (perhaps, at times, with the help of pragmatics) Shapiro's target claims, and do so in the natural way that

LP (and, more generally, *FDE*). I have some reservations about that, especially since the only motivation for recognizing 'gluts' arises from a 'small' class of very peculiar, abnormal sentences. But I will leave that issue for a larger project. (One route I've explored is to divide the atomic sentences into disjoint classes, and then define admissible valuations in such a way that only one of the two classes can be 'glutty.' Intuitively, the idea is that our target language—the language modeled by the formal language—already affords such a distinction (e.g., paradoxical sentences and otherwise), even though there is no decidable method for distinguishing the classes. But, again, I leave that issue for a larger project.)

[21] Note the 'non' in 'non-empty,' and the implicit 'not' in 'exclusive'! In normal cases, all such 'not's are unravelled classically, but the double-aspect view recognizes that, due to grammatical residue, there may be abnormal, free-floating behavior too.

[22] Few philosophers will say that it is logically possible that grass is red but not colored, despite the entrenchment of Tarskian (classical) logic. (There are classical Tarskian models in which grass is red but not colored, at least on the standard 'translation' or regimentation of English into classical logic.)

Shapiro suggests:

- *that A is true only* may be expressed as $A \wedge \sim(A \wedge \sim A)$.
- *that A is false only* may be expressed as $\sim A \wedge \sim(A \wedge \sim A)$.

Pragmatics, as mentioned, may be needed at times, but that is not a problem; pragmatics will inevitably be invoked in any final analysis of English. As Shapiro argues, the situation with respect to 'orthodox' dialetheism appears (at least prima facie) to be different: pragmatics will not help, given the validity of LNC in *LP*. At least on that score, double-aspect dialetheism is preferable.[23]

8 Closing Remarks and Further Directions

In this paper I have advanced what I call a double-aspect theory of negation, and in particular the theory of negation reflected in *AP*. While I have not argued the point (but, instead, relied on other literature), dialetheism appears to be a natural lesson to draw from the Liar paradox (and its kin). One obstacle to dialetheism has always been the 'gut-feeling' that the very meaning of *falsity* rules out the dialetheic lesson. That reaction, I have suggested, ignores the hypothesis that negation is ultimately aspectival, that negation generally exhibits classical behavior but, given odd (grammatically necessitated) sentences, sometimes behaves non-classically—indeed, behaves in a very free-floating fashion.

The double-aspect hypothesis respects the strong, classically minded intuitions that many philosophers have about negation; it respects such intuitions by *preserving* them—at least with respect to the normal, non-paradoxical cases,

[23] Joachim Bromand (Bromand 2002) raises what, in effect, is the same expressibility challenge that Shapiro raises. Bromand argues that orthodox dialetheism, underwritten by *LP*, cannot express that *A* is a non-dialetheia. As should be clear, double-aspect dialetheism does not fall prey to Bromand's version of the objection, which turns on the 3-valued semantics of *LP*.

which are precisely the ones on which such intuitions are founded. By my lights, the advantages of a dialetheic response to paradox—in particular, the preservation of expressive appearances—are strong; and the virtue of double-aspect dialetheism is that it retains such advantages while none the less retaining the insights of classical logic. Moreover, as indicated (albeit briefly) in § 7, double-aspect dialetheism, unlike its orthodox (single-aspect) counterpart, seems to avoid many of the oddities that have troubled those who have taken dialetheism seriously.

My aim in this paper has been to introduce the double-aspect view of negation (or double-aspect dialetheism). My hope is that the aim has been achieved and that the view is interesting and promising enough to foster further consideration. Before closing, however, I will mention an issue on which I have said very little: *conditionals*.

What (if any) conditional is at play in *AP*? At the very least, there is a 'material-like' conditional.[24] Suppose, for example, that $A \Rightarrow B$ is defined in the 'original' classical semantics: $\mathcal{R} \Vdash A \Rightarrow B$ iff either $\mathcal{R} \Vdash {\sim}A$ or $\mathcal{R} \Vdash B$. In the classical semantics, $\Rightarrow$ will be just the regular (classical) hook—the material conditional. But what happens when, after confronting the Liar, Exclusion is dropped? As discussed, negation becomes 'free-floating,' and so $\Rightarrow$, being defined in terms of $\sim$, will itself exhibit abnormal behavior. $A \Rightarrow A$ remains valid, but abnormality will none the less abound. For example, modus ponens will fail, as will many other traditional inferential properties; however, such abnormalities of $\Rightarrow$ piggy-back on the double-aspect of $\sim$, the upshot being that $\Rightarrow$ will behave classically for the vast range of sentences on which our traditional intuitions about 'if' are grounded.[25]

[24]Some do not take the 'material conditional' to be a genuine conditional. In other work, I hope to bolster *AP* with intensionality, perhaps thereby alleviating some of the concerns about the material conditional.

[25]Strengthening $\Rightarrow$ by adding modality to *AP* should yield similar results. I leave that project for another venue. (If, contrary to my current thinking, an aspectival approach to the conditional proves not to be viable, there are suitable conditionals to do the trick. See (Priest 1987).)

I close by clearing away one worry. One might think that the failure of modus ponens immediately undermines the entire project, since, one might think, with the failure of modus ponens one no longer derives the Liar paradox, and hence no longer derives that some (admittedly peculiar) sentence is true and false, and hence that there is no reason to think that negation has a double-aspect. That would be a devastating objection were it correct; however, it is not correct. What is interesting (and in some ways quite astonishing) is the sheer persistence of the Liar; the paradox persists even in the very weak logic P. Recall that $A \Rightarrow A$ and, in turn, $A \longleftrightarrow A$ remain valid in AP, at least on the given 'material-like' definition. Given intersubstitutivity of $\mathsf{T}\langle A\rangle$ and A, we have $\mathsf{T}\langle A\rangle \longleftrightarrow A$. As above, $A \Rightarrow B$, by definition, is $\sim A \vee B$, in which case $\mathsf{T}\langle A\rangle \longleftrightarrow A$ is

$$(\sim\mathsf{T}\langle A\rangle \vee A) \wedge (\sim A \vee \mathsf{T}\langle A\rangle)$$

Consider the Liar λ, which is $\sim\mathsf{T}\langle\lambda\rangle$. The λ-instance of $\mathsf{T}\langle A\rangle \longleftrightarrow A$ is $\mathsf{T}\langle\lambda\rangle \longleftrightarrow \sim\mathsf{T}\langle\lambda\rangle$, which by definition is

$$(\sim\mathsf{T}\langle\lambda\rangle \vee \sim\mathsf{T}\langle\lambda\rangle) \wedge (\sim\sim\mathsf{T}\langle\lambda\rangle \vee \mathsf{T}\langle\lambda\rangle)$$

Even in P (and, hence, in AP), the left conjunct yields $\sim\mathsf{T}\langle\lambda\rangle$ while the right conjunct yields $\mathsf{T}\langle\lambda\rangle$. Hence, even in P, and so without modus ponens, the original paradox arises: λ is a sentence that is both true and false; it is a sentence in which negation is on holiday.[26]

[26]I am grateful to Otávio Bueno, Mark Colyvan, Dave DeVidi, Hartry Field, Michael Glanzberg, Tim Kenyon, Phil Kremer, Daniel Nolan, Graham Priest, Stephen Read, Stewart Shapiro, and Greg Restall for comments or discussion. This paper was read at the philosophical logic workshop at the University of Waterloo (2003), held in memory of Graham Solomon. (I should note that, since writing this paper, I have moved towards a slightly more 'conservative' position, one that treats negation in a non-aspectival fashion (Beall unpublished). That said, I find the aspectival approach worth pursuing, and hope to further pursue it in the future—it may well be right, in the end.)

Chapter
NINE

MONISM: THE ONE TRUE LOGIC

Stephen Read

Logical pluralism is the claim that different accounts of validity can be equally correct. Beall and Restall have recently defended this position. Validity is a matter of truth-preservation over cases, they say: the conclusion should be true in every case in which the premises are true. Each logic specifies a class of cases, but differs over which cases should be considered. I show that this account of logic is incoherent. Validity indeed is truth-preservation, provided this is properly understood. Once understood, there is one true logic, relevance logic. The source of Beall and Restall's error is a recent habit of using a classical metalanguage to analyse non-classical logics generally, including relevance logic.

1 Logical Pluralism

JC Beall and Greg Restall have recently defended a position they call "logical pluralism", that "there is more than one sense in which arguments may be *deductively valid*, that these senses are equally good, and equally deserving of the name *deductive validity*" (Beall & Restall 2000, § 1). Their argument for logical pluralism is this:

D. Devidi and T. Kenyon (eds.), A Logical Approach to Philosophy, 193-209.

1. the meaning of the term 'valid' is given by (V):

   > (V) A conclusion, A, *follows from* premises, Σ, if and only if any case in which each premise in Σ is true is also a case in which A is true.

2. A logic specifies the cases which are mentioned in (V)

3. There are at least two different such specifications. (Beall & Restall 2000, pp. 476–7)

In fact, they describe three such specifications, namely, worlds (yielding classical logic), constructions (yielding constructive logic) and situations (yielding relevant logic) (*op. cit.*). Thus there are at least three logics, all equally good. All three tell us when truth is preserved, as (V) shows: classical logic tells us when logic is preserved in complete and consistent situations, that is, worlds; constructive logic tells us when truth is preserved in possibly incomplete (better, indeterminate or undecidable) situations, that is, constructions; and relevance logic tells us when truth is preserved in possibly inconsistent (and incomplete) situations. Indeed, Beall and Restall later introduce us to a fourth possibility, truth-preservation in all situations (possibly incomplete, inconsistent and indeterminate), which they rather confusingly also call "relevant consequence" (Beall & Restall 2001, § 4 fn. 17).

The position described here as logical pluralism is in fact incoherent. To see this, we need to look more closely at (V).

2 Priest's Challenge

Graham Priest challenges Beall and Restall as follows: suppose there really are two equally good accounts of deductive validity, K_1 and K_2, that β follows from α according to K_1 but not K_2, and we know that α is true. Is β true? (Priest 2001). Cf. (Beall & Restall 2001, § 6). Does the truth of β follow (deductively) from the information presented? Beall and Restall do not mean that β is true according to K_1 but not true according to K_2. K_1 and K_2 are accounts of validity, not of truth. As Priest notes, Beall and Restall deny that they are relativists about truth. So the question, 'Is β true?' is a determinate one. It follows K_1-ly that β is true, but not K_2-ly. Should we, or should we not conclude that β is true? The answer seems clear: K_1 trumps K_2. After all, K_2 does not tell us that β is false; it simply fails to tell us whether it is true. The information in the case is insufficient to determine, according to K_2, whether β is true. But according to K_1, the information supplied does tell us that β is true. So if K_1 and K_2 are both good accounts of derivability, K_1 tells us what we want to know: β is true.

It follows that in a very real sense, K_1 and K_2 are not equally good. K_1 answers a crucial question which K_2 does not. For Priest's question is the central question of logic. As Beall and Restall say, "the chief aim of logic is to account for [logical] consequence," that is, to tell us when "a conclusion A … logically follow[s] from premises Σ" (Beall & Restall 2000, pp. 475–6). In none of Beall and Restall's examples do logics seriously disagree, that is, does one logic say that A follows from Σ and the other that $\ulcorner{\sim}A\urcorner$ follows (unless, of course, Σ is inconsistent and they both say that both follow). And their pluralism is not unbounded. Although they admit classical, constructive and relevant accounts of validity to be equally good, they do not countenance any and every account of consequence to be logic (Beall & Restall 2000, p. 487 fn. 26). (V) builds in reflexivity and transitivity of consequence, since clearly inclusion (of Σ-ways in A-ways) is reflexive and transitive. So, they say, any system, such as Aristotle's system of syllogisms, which rejects reflexivity (Aristotle n.d., 24b18-20), or Tennant's (Tennant 1987, ch. 17) or Smiley's (Smiley 1959), which reject transitivity of consequence, is simply not a system of logic.

Beall and Restall's actual response to Priest's challenge is to say that we are entitled to infer β from α according to K_1, but not according to K_2 (Beall & Restall 2001, § 6). But this is no answer. That simply repeats the description of the case.

Suppose K_1 is classical logic and K_2 is relevance logic (as Beall and Restall do). We are given that the inference from α to β is classically valid and not relevantly valid. We are also told that α is true. Does this information tell us whether β is true? Apparently so, for classical validity is validity: "classical logic *is* logic .... If the premises of a classically valid argument are true, so is the conclusion" (Beall & Restall 2000, p. 490). So β is true, and not relatively true, but true *simpliciter*. The fact that β does not follow relevantly from α is irrelevant. Classical logic dominates, and β is true.

Two puzzles arise from this. First, relevance logic actually says more than that β does not follow relevantly from α. It says that one is not entitled to infer β from α. Relevance logic is an account of consequence. Beall and Restall describe this as saying that one is not relevantly entitled to infer β from α, whereas one is classically entitled to do so. But that classical entitlement, we saw, allowed us to infer β from α. So, given that α is true (and that β follows classically from α), we can infer that β is true—and not just classically infer it. If α is true then β is true. By their account, classical validity (or whatever is the stronger validity, K_1) dominates. This makes a mockery of relevance considerations. Relevance logic was not put forward as a mere alternative to classical logic. Ackermann, for example, believed that strict implication, which expresses classical validity, was wrong: "Thus one would reject the validity of the formula $A \to (B \to A)$" (Ackermann 1956, p. 113). So too for intuitionistic reasoning. Brouwer wrote:

> "An a priori character was so consistently ascribed to the laws of theoretical logic that until recently these laws, including the principle of excluded middle, were applied without reservation even in the mathematics of infinite systems and we did not allow ourselves to be disturbed by the consideration that the results obtained in this way are in general not open, either practically or theoretically, to any empirical corroboration. On this ba-

> sis extensive incorrect theories were constructed."
> (Brouwer 1923, p. 336)

Can a relevance logician, or an intuitionist maintain, in the face of Beall and Restall's pluralism, that one should not infer that β is true? We will return to this question in § 3.

Secondly, Beall and Restall offer a hostage to fortune here. Although none of their supposedly equally good logics disagrees in inferring contradictory statements from the same (consistent) premises, this would appear to be a possibility. Classical logic, K_1, dominates K_2, so does not disagree with it. Suppose it disagrees with K_3, in that while β follows K_1-ly from α, $\ulcorner{\sim}\beta\urcorner$ follows K_3-ly from α, while α is consistent—that is, there is some world, indeed this one, in which α is true. Should we infer that β is true, or that $\ulcorner{\sim}\beta\urcorner$ is true? We have seen that, according to Beall and Restall's pluralism, classically valid (that is, K_1-valid) arguments are valid. So β is true. But if K_3-valid arguments are also valid, β is false. Unless Beall and Restall accept the truth of a contradiction, they must find some reason for rejecting K_3 as not logic, like the syllogism and non-transitive systems. Such reasons had better not be *ad hoc*. One good reason (or at least, not *ad hoc*) would be if K_3 did not admit a semantics of cases, and so did not fit their guiding principle (V). But that needs argument.

An example is given by Abelian logic (Meyer & Slaney (circa 1984)), whose characteristic axiom is $((A \to B) \to B) \to A$. This is not a classical tautology, but Abelian logic is consistent (and Post-complete), lacking certain classical validities in compensation. Hence in classical logic,

$$\sim A, B \vdash \ \sim(((A \to B) \to B) \to A),$$

that is, A false and B true is a counterexample. But in Abelian logic,

$$\sim A, B \vdash ((A \to B) \to B) \to A,$$

since the conclusion is (Abelianly) logically true. Suppose now that we discover that A is false and B is true. Should we infer that $((A \to B) \to B) \to A$ is true, or false? Classical and Abelian logic give conflicting answers. Here pluralism meets its limit.

Beall and Restall might try to dismiss Abelian logic on the grounds that it does not admit a semantics of cases, and so does not fall under (V). But one should note Routley's proof (Routley n.d.) that every logic admits a two-valued worlds semantics. If he is right, every logic falls under (V). Thus Abelian logic really is a counterexample to Beall and Restall's pluralism.

Beall and Restall's logical pluralism tries to be eclectic and all-embracing (up to a point which excludes Aristotle, Smiley and Tennant), but it falls down on two counts: first, it does not respect the core motivation of the non-classical logics, which first prompted them as rivals to the classical orthodoxy; and it threatens to plunge into inconsistency, if explicitly incompatible logics both turn out to accord with the governing principle, (V).

Let us turn to examine (V) more closely.

3 Truth-Preservation

Beall and Restall describe (V) as a principle of truth-preservation. It states that validity requires truth to be preserved in all cases. Different specification of the cases then yields different logics consonant with (V). Any system not consonant with (V) is not logic, and any system consonant with (V) is equally good as a logic.

This is puzzling, for as Beall and Restall point out, there are, for example, "too many modal logics to hold each of them as the logic of broad metaphysical necessity" (Beall & Restall 2000, p. 489). What is required, they say, is to specify what a logic is meant to do, and then there is scope for disagreement. If we want to capture metaphysical necessity, one modal logic is the right logic. Perhaps if we want to capture moral obligation, a different modal logic will better capture the interpretation we want for the operators, and so too for formalizing the logic of knowledge and the logic of provability.

But this difference of logic is orthogonal to Beall and Restall's thesis of logical pluralism, as Restall observes (Restall 2002, p. 431). The former is Carnapian tolerance, which tolerates logical disagreement as due to difference of language. There is no real disagreement, and nothing the logical monist might object to. Clearly, if $\ulcorner \Box p \urcorner$ expresses '*p* is obligatory', we reject $\Box p \vdash p$, for unfortunately not everyone fulfils their obligations. Again, if $\ulcorner \Box p \urcorner$ expresses '*p* is provable', we reject $\Box p \vdash p$, since not all systems of proof are consistent. But if $\ulcorner \Box p \urcorner$ expresses logical necessity, we insist on $\Box p \vdash p$, for what is necessary must happen. (As a Gifford Lecturer at St Andrews once put it, referring to personal experience: "if you can't breathe, you don't.") These alternative logics are supplementary logics, in Haack's sense (Haack 1974, p. 2), and do not illustrate any real sense in which one and the same inference can be both valid (according to one logic) and invalid (according to another).

The same point applies to another example which Beall and Restall mention, the distinction between formal and material consequence. Take their example, '*a* is red, so *a* is coloured.' There is nothing here for a logical monist to jib at. Every instance of a formally valid argument is valid. But not every instance of a formally invalid argument is invalid. Formally invalid arguments can have valid instances, some of which will be formally valid in virtue of instantiating a different valid form, but others valid not in virtue of form at all. (V) allows validity to arise from many causes, and does not distinguish formal validity from other sorts of validity.

Again, the distinction between first-order and higher-order validity need not disturb a logical monist. Many valid arguments are first-order valid; some are not. Some of the latter are second-order valid, but others are not. To repeat, every instance of a valid form is valid; but invalid forms can have valid instances. Allowing higher-order expressive power, and increasing the range of logical constants (e.g., to include modal and bimodal, e.g., temporal, connectives) both increase the range of formal validity. But these are all part of the one canon of validity for the monist. As (V) puts it: if any case in which the premises are true is one in which the conclusion is true, the argument is valid, and vice versa.

Beall and Restall believe that (V) covers relevant consequence, as well as other logics. This is, however, tendentious. Relevant consequence is paraconsistent, in rejecting the inference from a contradiction to any proposition whatever: formally, $\ulcorner A \;\&\; {\sim}A \urcorner$ does not imply arbitrary B. (Let us call this Ex Falso Quodlibet, EFQ for short.) Beall and Restall distinguish three types of paraconsistent logician (Beall & Restall 2001, § 2): first, there is the regular dialetheist, who believes there are true contradictions—the actual world is inconsistent, as shown, for example, by the paradoxes. One man's *modus tollens* is another's *modus ponens*, so the fact that, say, Naive Set Theory leads to contradiction does not refute Naive Set Theory, but gives the regular dialetheist reason to believe that the ensuing contradictions are true. The light dialetheist is more cautious: the actual world might be inconsistent, but the jury is still out on whether that is the right conclusion to draw from the paradoxes. Both types of dialetheist are paraconsistentists, since even if some contradictions might be true, not every proposition could be true.

In contrast, non-dialetheic paraconsistentists, so-called, do not think contradictions could be true. Beall and Restall describe them as concerned with ways the world could not be—with impossible worlds. (See (Beall & Restall 2001, §§ 1–2).) For EFQ to be invalid, according to (V), there must be cases where $\ulcorner A \;\&\; {\sim}A \urcorner$ is true and B false. But $\ulcorner A \;\&\; {\sim}A \urcorner$ cannot be true—any case where $\ulcorner A \;\&\; {\sim}A \urcorner$ is true is impossible. So the non-dialetheic paraconsistentist seems committed to saying that there are ways the world could not be, and that such cases must be considered in deciding on the validity of an argument. This is an incoherent position, for if such cases cannot arise, it is hard to see why they need to be considered.

The dialetheic paraconsistentist is not in such a bind. For the dialetheist, the actual case is, or at least could be, inconsistent. So there is a real possibility that the premise of EFQ is true, and no guarantee that if it is, the conclusion is true too. So (V) shows that EFQ is invalid.

But if one does not think that $\ulcorner A \mathbin{\&} \sim A \urcorner$ could be true, how can EFQ fail to conform to (V)? Beall and Restall dub this the "Peircean objection" (Beall & Restall 2001, § 2). There cannot be a case in which $\ulcorner A \mathbin{\&} \sim A \urcorner$ is true and B false, for cases in which $\ulcorner A \mathbin{\&} \sim A \urcorner$ is true are impossible. One cannot be led astray by EFQ, whereby a case where $\ulcorner A \mathbin{\&} \sim A \urcorner$ was true would transform itself into one where everything was true, for there can no more be a case where $\ulcorner A \mathbin{\&} \sim A \urcorner$ is true than there can be a case where arbitrary B is true—such cases are impossible.

Beall and Restall's response to the Peircean objection is to claim that there is more than one way of being led astray—that is, that "even if the whole purpose of Logic is to avoid being led astray, there seems to be more than one logic that may arise given this purpose" (Beall & Restall 2001, § 2). That is, there is more than one way of going astray. Relevant validity, endorsed by the non-dialetheic paraconsistentist, fleshes out this purpose in one way; constructive and classical validity in yet others. Proponents of the latter pair are safe: they endorse EFQ, and do not think there is any case where $\ulcorner A \mathbin{\&} \sim A \urcorner$ is true and B false. The first of the three is in a tighter corner: by Beall and Restall's lights (but see § 4 below), the non-dialetheic paraconsistentist thinks there are such cases, but they are impossible. Consequently, Beall and Restall ascribe to the non-dialetheic paraconsistent the thought: "One would be led astray if one's conclusion didn't conform to the canons of relevance. Better put: One would be led astray if one's conclusion failed to *follow relevantly* from one's premises" (Beall & Restall 2001, § 2).

This is to abandon truth-preservation as the criterion of validity. What is now required of validity is not just preservation of truth, but preservation of relevance, too. (V) may look like a statement of truth-preservation as the criterion of validity, but interpreted as Beall and Restall take it, it is not.

For an impossible case is not a case in which anything can be true. If it is a case at all, it is a case in which it is impossible for anything to be true. Thus if the non-dialetheic paraconsistentist is committed to interpreting (V) as ranging over impossible cases, (V) describes at best case-preservation, not truth-preservation.

Moreover, to attribute to the (non-dialetheic) relevantist the view that validity requires more than just preservation of truth, namely, preservation of relevance, falls prey to a variant of Priest's objection. (I posed this question in (Read 1988*a*) and (Read 2004).) For again, suppose the argument from α to β preserves truth, and that α is true. Should we conclude that β is true? According to (V), any case in which α is true is one in which β is true, and by hypothesis, α is true. So it seems clear that β is true. According to the non-dialetheic relevantist, however, we can be led astray here. How? By failing to keep to the "canons of relevance", we are told. Yet what is the sanction of violating these canons? Not that β is not true. For β is true in every case in which α is true, and α is true. What, however, of the impossible cases? In these, α could be true and β not. But these cases are impossible. So the only cases in which β is not true are impossible. So β is true. Apparently, however, according to Beall and Restall's non-dialetheic relevantist, we should not infer that β is true. That is absurd.

4 Classical Semantics

How have Beall and Restall argued themselves into this absurd position? The answer is that they have misunderstood Meyer's Sermon to the Gentiles (Meyer 1985). Semantics is invariably carried out in a classical metalanguage. Modal logic for years—decades—had no semantics, and felt inferior for that reason. Kripke eventually supplied a semantics, developed in a non-modal, extensional metalanguage. Intuitionistic logic had no semantics, at least, no formal semantics, and some dismissed it for that reason. Without a semantics, one could not understand what justified intuitionistic methods. Beth and Kripke provided a semantics, framed in a classical metatheory. Relevance logic lacked a semantics through its first decade, and suffered the same criticism. Meyer, Routley and others provided the semantics in the classical metatheory of their critics. As Meyer put it, they set out "to preach to the Gentiles in their own tongue" (Meyer 1985, p. 1).

There is a common assumption in all these cases, namely, that classical logic is right, or at least, right for doing semantics. It allows classical logicians to understand what modal, constructive and relevance logicians are doing. Except that it doesn't. It provides a classical model, or classical interpretation, of modal, constructive and relevant reasoning. Modal logicians are interpreted as talking about truth-values (extensional properties) of propositions at other possible worlds (sets), rather than about modal properties of those propositions. Intuitionist logicians are interpreted as talking about possible constructions in states of information, and the provability of propositions, rather than about those propositions' (epistemically constrained) truth and falsity. Relevance logicians are interpreted as concerned with truth-preservation in arcane situations, situations which in the interesting cases—that is, the cases where their account of validity differs from that in classical logic—turn out to be impossible. To say that such cases are impossible should mean that there are no such cases, yet Beall and Restall saddle the non-dialetheic paraconsistentist with holding that there are such cases, only they are impossible. But if they are impossible, then it is impossible that there are such cases. If they are impossible, then there is no situation, however arcane, in which they hold.

This may seem a cheap point. After all, Beall and Restall write:

> "These situations are … 'impossible.' Not in the sense that they do not exist (one may well be a realist about these impossible situations) but in the sense that they can never be *actualized*. They are

> never part of any possible world." (Beall & Restall 2001, § 4)

Priest (Priest 1997) in fact distinguishes three notions of impossible world. First, they may be (what Beall and Restall would call) "cases" where A and $\ulcorner\sim A\urcorner$ are both true, for some A; or they may be cases where classical logic does not hold; or they may be cases where one's preferred logic does not hold. For the non-dialetheic paraconsistentist, of course, there can be no cases of the first kind; and no one should think there can be cases of the third. Any non-classical logician should believe there can be cases of the second kind—indeed, that the actual case is one.

But what of Beall and Restall's distinction between whether these cases exist, and whether they can be actualized? Restall (Restall 1997) offers us a modelling of such cases as sets of possible worlds. (Cresswell, (Cresswell 1973, p. 42), called them "heavens.") It is again part of the pluralist project: "we can enjoy the fruits of both paraconsistent and classical logic" (p. 594). However, this is not realism about impossible situations; these impossibilities exist, as sets, but they are not real (as situations).

Varzi (Varzi 1997) offers us a moderate realism: just as there are ways things could be, namely, maximal consistent states of affairs, so there are ways things could not be, namely, maximal inconsistent states of affairs (p. 598). For 'There is no way that a can be F' is equivalent to 'There is a way a cannot be, namely, F'—a "couldn't be *that* way!" (*loc.cit.*) But of course there is a way a couldn't be F—*every* way is a way the impossible cannot be. This book, for example, couldn't be black and white and red all over, indeed, everything is like that. So too, every way is one in which a couldn't be F, if a can't be F. But that does not magically yield impossible worlds. Far from showing that impossible worlds are real, Varzi's argument reinforces the conviction that they are unreal and that there are no such things.

Yagisawa (Yagisawa 1988) argues for the admission of impossible worlds within a Lewisian modal realism by a kind of

Cantor-paradox argument. Consider the collection of all possible worlds—the Lewis universe. Suppose it had been different in some way—more worlds, or different accessibility relations, or whatever. Such a supposition is an impossibility. Hence, he says, the Lewisian universe is an island within a much larger realm of impossibilities. Such an argument shows the danger of conceiving of possible worlds in such a literal way as Lewis'. Talk of possible worlds is a façon de parler, and like all façons de parler, its extensionalist merits must be balanced against the risk of being misled by it. There are not really any possible worlds, and there certainly are no impossible worlds—they're impossible. What there is, is what there is, the actual world. This world has certain actual properties (how it is) and certain modal properties (how it might be, how it must be, and how it could not be).

Hence talk of inconsistent situations (Priest's first kind of impossible world) is a metaphorical way of talking of inconsistency, of how the actual situation cannot be. It may assist the classical logician to model counterexamples to relevantly invalid reasoning. But it should not be allowed to mislead him into supposing that the non-dialetheic paraconsistentist believes there can somehow be (unactualisable but real) impossible situations or cases.

What the classical perspective is insensitive to, is the real motivation for questioning whether, e.g., EFQ is valid, just as it is insensitive to the real nature of modality or of constructivism. The background assumption is that classical logic is one correct way of doing logic. To accommodate the constructivist and relevantist concerns, it is necessary to admit other ways of doing logic as correct. Thus is logical pluralism born. It is born out of combining a non-classical theory with a classical metatheory. If classical logic is right, how can we understand what the non-classical logician is doing? Having understood those non-classical criticisms, there must be at least two ways of doing logic.

Copeland (Copeland 1979) responded to Meyer and Routley's classical semantics by dismissing it as purely technical, a mathematical method of obtaining metatheoretical results, but of no semantic import. In particular, Copeland objected to the Routleys' clause for negation:

(T*) $\ulcorner{\sim}A\urcorner$ is true at w iff A is not true at w^*.

Either this has nothing to do with semantics, but enables one to manipulate the uninterpreted symbol '$\sim$' in pure semantics for relevance logic; or it does explain the meaning of '$\sim$', in which case, classical and relevance logic are discussing different connectives. If (T*) gives the meaning of '$\sim$', then its meaning is different from the negation in classical logic and, as Prior put it, classical and relevance logicians are "simply talking past one another" (Prior 1967, p. 75). Indeed, as Quine famously quipped, "when he tries to query the doctrine, [the deviant logician] only changes the subject" (Quine 1970, p. 81). If Routley semantics is applied semantics, then (T*) shows that '$\sim$' is not 'not'; if the relevance logician really denies classical laws about negation, then it cannot be logic which the semantic techniques are explaining, but some other strange game—pure semantics.

Restall challenges this argument by showing how the so-called classical negation clause:

(T $\sim$) $\ulcorner{\sim}A\urcorner$ is true at w iff A is not true at w

is a special case of (T*) when w is a world (i.e., consistent and complete) (Restall 1999, p. 61). For w^* is the maximal point consistent with w (relative to the ordering that $w \subseteq w'$ iff whatever is true at w is true at w'), which is just w if w is consistent: "for if xCx [i.e., x is self-compatible, i.e., consistent] then if $x \vDash \sim A$ we cannot have $x \vDash A$" (*loc. cit.*). But this assumes that the metalanguage is consistent, in this case, classical. If the metalanguage matches the object-language (where we may have both $x \vDash A$ and $x \vDash \sim A$) then we may have both $x \vDash A$ and $x \nvDash A$. If a dialetheist (about the object-language) accepts a classical (i.e., consistent) metalanguage then of course he is a pluralist—indeed, schizophrenic.

What is a non-dialetheic relevantist to make of all this? Certainly, the suggestion that both $x \vDash A$ and $x \nvDash A$ is absurd. The non-dialetheic relevantist shares an aversion to dialetheism with the classicist. But then, as we have seen, there are no worlds in which both A and $\ulcorner{\sim}A\urcorner$ are true (if '$\sim$' means 'not'): such worlds would be impossible, and so there are no such worlds. Talk of "worlds" (and "truth" etc.) is just a façon de parler, and the semantics (so-called) is just pure semantics, as Copeland observed.

It is a mistake to describe (T $\sim$) as the classical clause for negation. It is only classical if the interpretation of 'not' is classical; and it is only correct if the interpretation of '$\sim$' and 'not' is the same. If one allows object– and metalanguage to drift apart, then a split personality and logical pluralism are just around the corner. The right response is to insist on doing one's semantics in the logic in which one believes. If Beall and Restall insist on doing semantics classically, then they are classical logicians for whom non-classical "logics" are, if not just an intellectual amusement, then an exercise in applying logic to some more particular activity—e.g., database management (see (Restall 1999, p. 69)) or warrant transfer (see (Restall 2004)). In contrast, if one believes that, e.g., double negation elimination, or EFQ are invalid (as constructivist and relevantist do, respectively), then one should reject the canons of classical logic even, or especially, when applied to the semantic study of one's chosen account of validity.

This robust response is an ingredient of what I once dubbed "logic on the Scottish plan" (Read 1988*b*, § 7.8), in contrast to versions of the semantics of relevance logic which were familiarly known as "logic on the American plan" and "logic on the Australian plan", which, e.g., adopts (T*) as the clause for negation in order to work in the Gentiles' own tongue, classical logic. Under the Scottish plan, the truth-conditions of the connectives are homophonic, as in (T $\sim$) read properly. Adopting such clauses in a classical metatheory, relevance logic will appear incomplete: (classically) valid inferences concerning 'not' will not be validated by the (relevant object-)theory. But what a strange approach to take, if one believes relevance logic is the correct logic. Why use an alien logic for one's metatheory—and if one does, why trust the result?

Articulating a relevant metatheory requires thought and reflection. In particular, one needs to consider what the relevant account of truth-preservation (validity) is. Suppose we formalize (V):

$$(\mathrm{V} \Rightarrow) \qquad \Sigma \vdash A \text{ iff } (\forall w)((\forall B \in \Sigma) w \vDash B \Rightarrow w \vDash A)$$

Beall and Restall think to obtain different accounts of '$\vdash$' by varying the range of 'w'—cases may be worlds, constructions or situations, for example. But the range of 'w' should be universal, and unless one is a dialetheist, impossible worlds do not fall under the range of 'w', for there are no such worlds. Rather, different theories of consequence result from varying the interpretation of '$\Rightarrow$'. In classical logic, there is really only one possibility for '$\Rightarrow$', namely, material implication. In relevance logic, there are two. For relevance logic distinguishes material from relevant implication—or better, classical logic conflates them, illicitly warranting their equivalence. Which is the right account of validity?

The right one is the one I dubbed "the Relevant Account of Validity" (Read 1988*b*, § 6.5). For the essential feature of validity is that it should warrant one in proceeding from the truth of the premises to that of the conclusion. But material detachment is invalid:

$$(\mathrm{V} \supset) \qquad \Sigma \vdash_{\supset} A \text{ iff } (\forall w)((\forall B \in \Sigma) w \vDash B \supset w \vDash A)$$

Learning that $\alpha \vdash_{\supset} \beta$ and that α is true does not warrant belief that β is true. That would be a use of Disjunctive Syllogism for '$\vee$' ($A \vee B$, $\sim A \vdash B$), which is well-known to lead to the validity of EFQ in four easy moves (the so-called Lewis argument: see, e.g., (Read 1988*b*, § 2.6)). What does warrant one in moving from the truth of α to that of β is learning that α relevantly implies β:

$$(\mathrm{V} \rightarrow) \qquad \Sigma \vdash_{\rightarrow} A \text{ iff } (\forall w)((\forall B \in \Sigma) w \vDash B \rightarrow w \vDash A)$$

Accordingly, (V→) is the correct account of truth-preservation, and the correct account of validity. There is one true logic, relevance logic, and it consists in rejecting classical logic, including classical semantics.

5 Conclusion

Beall and Restall's logical pluralism is incoherent. It claims that an inference can be both valid according to one account of logic and invalid according to another, and yet that this is not disagreement about validity but about logical purpose. But there is only one purpose of logic: to distinguish the valid inferences from the invalid ones. Among Beall and Restall's "equally good" logics, one dominates: classical logic. This is because they view all their logics from the perspective of classical semantics. Hence their other logics disagree with classical logic only in failing to recognise certain classical inferences as valid.

Other logics might claim in contrast that classically invalid inferences are valid. Then Beall and Restall's eclecticism would collapse into inconsistency. Even without that possibility, Beall and Restall's pluralism ignores the non-classical rejection of classical inference, interpreting it only as an incompleteness, not recognising these validities rather than excluding then as really invalid.

There is one true logic, and it does take (V) as its criterion of validity. But it results from understanding the true nature of truth-preservation, that the conclusion be true whenever the premises are true. (V) needs to be interpreted, and developed, in a relevant metalanguage in which the relevance of the premises to the conclusion is an integral part of truth-preservation: if the conclusion really does follow from the premises then those premises must be, logically, relevant to the conclusion.

BIBLIOGRAPHY

Ackermann, W. (1956), 'Begründung einer strengen Implikation', *Journal of Symbolic Logic* **21**, 113–128.

Alspector-Kelly, M. (to appear), 'Should the empiricist be a constructive empiricist?', *Philosophy of Science*.

Anderson, C. A. (1989), Russell on order in time, *in* C. W. Savage & C. A. Anderson, eds, 'Rereading Russell: Critical Essays on Bertrand Russell's Metaphysics and Epistemology', University of Minnesota Press, Minneapolis, pp. 249–263.

Aristotle (n.d.), *Analytica Priora*.

Asser, G. & Schröter, K. (1958), 'Axiomatisierung der k-zahlig allgemeingültigen ausdrücke des Stufenkalküls', *Mathematische Nachrichten* **19**, 73–86. Reviewed by Theodore Hailperin, *Journal of Symbolic Logic* **25** (1960), p. 176.

Batens, D. (2000), A survey of inconsistency-adaptive logics, *in* D. Batens, C. Mortensen, G. Priest & J.-P. V. Bendegem, eds, 'Frontiers of Paraconsistency', Research Studies Press, King's College Publications, Baldock, pp. 49–73.

Beall, J. C. & Restall, G. (2000), 'Logical pluralism', *Australasian Journal of Philosophy* **78**, 475–493.

Beall, J. C. & Restall, G. (2001), Defending logical pluralism, *in* B. Brown & J. Woods, eds, 'Logical Consequences', Hermes Science Publishers.

Beall, J. C. & van Fraassen, B. C. (2003), *Possibilities and Paradox: An Introduction to Modal and Many-Valued Logic*, Oxford University Press, Oxford.

Beall, J. C. (unpublished), Simply semantic inconsistency, to appear.

Bell, J., DeVidi, D. & Solomon, G. (2001), *Logical Options*, Broadview Press.

Bell, J. L. (1986), 'From absolute to local mathematics', *Synthese* **69**, 409–426.

Bell, J. L. (1993*a*), 'Hilbert's ε-operator and classical logic', *Journal of Philosophical Logic* **22**, 1–18.

Bell, J. L. (1993*b*), 'Hilbert's ε-operator in intuitionistic type theories', *Mathematical Logic Quarterly* **39**, 323–337.

Bell, J. L. (to appear), The development of categorical logic, *in* D. Gabbay & F. Guenther, eds, 'Handbook of Philosophical Logic', Kluwer.

Benacerraf, P. (1985), 'Skolem and the skeptic', *Proceedings of the Aristotelian Society* **Supplementary Vol. 59**, 85–115.

Boolos, G. & Jeffrey, R. (1989), *Computability and Logic*, 3 edn, Cambridge University Press, Cambridge.

Boolos, G. (1994), The advantages of honest toil over theft, *in* A. George, ed., 'Mathematics and Mind', Oxford University Press, Oxford, pp. 27–44.

Bridges, D. & Reeves, S. (1999), 'Constructive mathematics in theory and programming practice', *Philosophia Mathematica (Series 3)* **7**, 65–104.

Bromand, J. (2002), 'Why paraconsistent logic can only tell half the truth', *Mind* **111**, 741–749.

Brouwer, L. E. J. (1923), On the significance of the principle of excluded middle in mathematics, especially in function theory, *in* J. van Heijenoort, ed., 'From Frege to Gödel', Harvard University Press, Cambridge, Mass., pp. 334–345. 1967 reprint, with additions, of 1923 original.

Carnap, R. (1937), *The Logical Syntax of Language*.

Carnap, R. (1946), 'Modalities and quantification', *Journal of Symbolic Logic* **11**, 33–64.

Carnap, R. (1963), Replies, *in* P. Schilpp, ed., 'The Philosophy of Rudolf Carnap', Open Court, LaSalle, IL, pp. 859–1016.

Cartwright, R. (1987), On the origin of Russell's theory of descriptions, *in* 'Philosophical Papers', MIT Press, Cambridge, Mass., pp. 95–134.

Chambers, T. (2000), 'A quick reply to Putnam's paradox', *Mind* **109**, 195–197.

Copeland, B. (1979), 'On when a semantics is not a semantics: some reasons for disliking the Routley–Meyer semantics for relevance logic', *Journal of Philosophical Logic* **8**, 399–413.

Copeland, B. (1996), *Logic and Reality: Essays on the Legacy of Arthur Prior*, Clarendon Press, Oxford.

Copeland, B. (2002), 'The genesis of possible worlds semantics', *Journal of Philosophical Logic* **31**, 99–137.

Corcoran, J. (1980), 'Categoricity', *History and Philosophy of Logic* **1**, 187–207.

Corcoran, J. (1981), 'From categoricity to completeness', *History and Philosophy of Logic* **3**, 113–119.

Cresswell, M. (1973), *Logics and Languages*, Methuen.

Demopoulos, W. & Friedman, M. (1985), 'Russell's analysis of matter: Its historical context and contemporary interest', *Philosophy of Science* **52**, 621–639.

Demopoulos, W. (1999), 'On the theory of meaning of 'On Denoting'', *Noûs* **33**, 439–358.

Demopoulos, W. (2003), Russell's structuralism and the absolute description of the world, *in* N. Griffin, ed., 'The Cambridge Companion to Russell', Cambridge University Press, Cambridge.

DeVidi, D. & Solomon, G. (2001), 'Knowability and intuitionistic logic', *Philosophia: Philosophical Quarterly of Israel* **28**, 319–334.

DeVidi, D. (2004), 'Choice principles and constructive logics', *Philosophia Mathematica* **12**, 222–243.

DeVidi, D. (to appear), Vagueness and intuitionistic modal logic: The Wright way of handling vagueness, *in* K. Peacock & A. Irvine, eds, 'Mistakes of Reason: Essays in Honour of John Woods'.

DiSalle, R. (2002), 'Convention and modern physics: A re-assessment', *Noûs* **36**, 169–200.

Dummett, M. (1977), *Elements of Intuitionism*, Clarendon/Oxford.

Dummett, M. (1978), *Truth and Other Enigmas*, Duckworth, London.

Dummett, M. (1991), *The Logical Basis of Metaphysics*, Harvard University Press.

Dummett, M. (1993), Realism and anti-realism, *in* 'The Seas of Language', Clarendon/Oxford, pp. 462–478.

Eddington, A. S. (1924), *Mathematical Theory of Relativity*, Cambridge University Press, Cambridge.

English, J. (1973), 'Underdetermination: Craig and Ramsey', *Journal of Philosophy* **70**, 453–461.

Field, H. (2003), Is the liar both true and false?, *in* J. C. Beall & B. Armour-Garb, eds, 'Deflationism and Paradox', Oxford University Press, Oxford. Forthcoming.

Friedman, M. (1999), *Reconsidering Logical Positivism*, Cambridge University Press.

Frisch, M. (1999), 'Van Fraassen's dissolution of Putnam's model-theoretic argument', *Philosophy of Science* **60**, 158–164.

Gaifman, H. (1982), Local and non-local properties, *in* J. Stern, ed., 'Proceedings of the Herbrand Symposium: Logic Colloquium '81', North-Holland, Amsterdam.

Gentzen, G. (1936), 'Die Wiederspruchsfreiheit der Stufenlogic', *Mathtematische Zeitschrift* **41**, 214–222.

Gödel, K. (1930), On the completeness of the axioms of the functional calculus of logic, *in* S. Feferman, J. W. Dawson, S. C. Kleene, G. H. Moore, R. M. Solovay & J. van Heijenoort, eds, 'Kurt Gödel: Collected Works', Vol. 1, Oxford University Press, New York. 1986 reprint of German original with facing English translation.

Gödel, K. (1931), 'On formally undecidable propositions of Principia Mathematica and related systems, pt. I', *Monatshefte für Mathematik und Physik* **38**, 173–198. The German original with facing English translation is in S. Feferman, et al., eds, *Kurt Gödel: Collected Works, vol. I*, New York: Oxford University Press, 1986.

Gödel, K. (1933), An interpretation of the intuitionistic sentential logic, *in* J. Hintikka, ed., 'The philosophy of mathematics', Oxford University Press, London, pp. 128–129. 1969 reprint, originally published 1933.

Grayson, R. J. (1979), Heyting-valued models for intuitionistic set theory, *in* M. P. Fourman, C. J. Mulvey & D. Scott, eds, 'Applications of Sheaves', Vol. 753 of *Springer Lecture Notes in Mathematics*, Springer-Verlag, pp. 402–414.

Haack, S. (1974), *Deviant Logic*, Cambridge University Press, Cambridge.

Hailperin, T. (1957), 'A theory of restricted quantification', *Journal of Symbolic Logic* **22**, 19–35, 113–129. Also **25** (1960) 54–56.

Harrington, L. et al., eds (1985), *Harvey Friedman's Research in the Foundations of Mathematics*, North-Holland, Amsterdam.

Hart, W. P. (1970), 'Skolem's promises and paradoxes', *Journal of Philosophy* **67**, 98–109.

Hilbert, D. & Ackermann, W. (1928), *Grundzüge der Theoretischen Logik [Principles of Mathematical Logic]*, Julius Springer, Berlin.

Hintikka, J. (1964), *Knowledge and Belief*, Cornell University Press, Ithaca, N.Y.

Howard, W. A. (1980), The formulas-as-types notion of construction, *in* J. P. Seldin & J. R. Hindley, eds, 'To H.B. Curry: Essays on Combinatory Logic, Lambda Calculus, and Formalism', Academic Press, New York, pp. 480–90.

Jónnson, B. & Tarski, A. (1951), 'Boolean algebras with operators, part I', *American Journal of Mathematics* **73**, 891–939.

Jónnson, B. & Tarski, A. (1952), 'Boolean algebras with operators, part II', *American Journal of Mathematics* **74**, 127–162.

Kanger, S. (1957*a*), 'The morning star paradox', *Theoria* **23**, 1–11.

Kanger, S. (1957*b*), 'A note on quantification and modalities', *Theoria* **23**, 133–34.

Kaplan, D. (1966), 'Review of Kripke "Semantical Analysis of Modal Logic I: Normal Modal Propositional Calculi"', *Journal of Symbolic Logic* **31**, 120–122.

Keenan, E. (2002), Logical objects, *in* C. A. Anderson & M. Zeleny, eds, 'Logic, Meaning and Computation: Essays in Memory of Alonzo Church', Kluwer, Dordrecht and Boston, pp. 149–180.

Kripke, S. (1959*a*), 'A completeness theorem in modal logic', *Journal of Symbolic Logic* **24**, 1–14.

Kripke, S. (1959*b*), 'Semantical analysis of modal logic (abstract)', *Journal of Symbolic Logic* **24**, 323–324.

Kripke, S. (1963), 'Semantical analysis of modal logic I: Normal modal propositional calculi', *Zeitschrift für mathematische Logik und Grundlagen der Mathematic* **9**, 67–96.

Ladyman, J. (2000), 'What's really wrong with constructive empiricism? Van Fraassen and the metaphysics of modality', *British Journal for the Philosophy of Science* **51**, 837–856.

Lambek, J. & Scott, P. J. (1986), *Introduction to Higher Order Categorical Logic*, Cambridge Studies in Advanced Mathematics 7, Cambridge University Press, Cambridge.

Lawvere, F. W. & Rosebrugh, R. (2003), *Sets for Mathematics*, Cambridge University Press, Cambridge.

Lawvere, W. (1976), Variable quantities and variable structures in topoi, *in* 'Algebra, Topology and Category Theory: A Collection of Papers in Honor of Samuel Eilenberg', Academic Press.

Lewis, D. (1970), 'How to define theoretical terms', *Journal of Philosophy* **67**, 427–445.

Lewis, D. (1990), *Parts of Classes*, Blackwell, Oxford.

Lewis, D. (1993), 'Mathematics is megethology', *Philosophia Mathematica* **1**, 3–23. Reprinted in his *Papers in Philosophical Logic*, Cambridge University Press, 1998.

Maietti, M. & Valenti, S. (1999), 'Can you add power sets to Martin-Löf's intuitionistic set theory?', *Mathematical Logic Quarterly* **45**, 521–532.

Martin-Löf, P. (1984), *Intuitionistic Type Theory*, Bibliopolis, Naples.

Martin, R. L., ed. (1984), *Recent Essays on Truth and the Liar Paradox*, Oxford University Press, New York.

McGee, V. (1991), *Truth, Vagueness, and Paradox*, Hackett, Indianapolis.

McGee, V. (1997), 'How we learn mathematical language', *Philosophical Review* **106**, 35–68.

McKinsey, J. & Tarski, A. (1948), 'Some theorems about the sentential calculi of Lewis and Heyting', *Journal of Symbolic Logic* **13**, 1–15.

Meredith, C. A. & Prior, A. N. (1956), 'Interpretations of different modal logics in the "property calculus"', Mimeograph, University of Canterbury Philosophy Department. Reprinted in (Copeland 1996).

Meyer, R. K. & Slaney, J. K. ((circa 1984)), Abelian logic (from a to z), Technical Report 7, A. N. U. R. S. S. S. Logic Group Research Paper.

Meyer, R. (1985), Proving semantical completeness "relevantly" for R, Technical Report 23, A. N. U. R. S. S. S. Logic Group Research Paper.

Montague, R. (1960), 'Logical necessity, physical necessity, ethics, and quantifiers', *Inquiry* **3**, 259–69. Reprinted in (Montague 1974).

Montague, R. (1974), *Formal Philosophy: Selected Papers of Richard Montague*, Yale University Press, New Haven.

Myhill, J. (1950), 'A complete theory of natural, rational and real numbers', *Journal of Symbolic Logic* **15**, 185–196.

Newman, M. (1928), 'Mr Russell's causal theory of perception', *Mind* **37**, 137–148.

Parsons, C. (1990), 'The structuralist view of mathematical objects', *Synthese* **84**, 303–346.

Priest, G. & Routley, R. (1989), Systems of paraconsistent logic, *in* G. Priest, R. Routley & J. Norman, eds, 'Paraconsistent Logic: Essays on the Inconsistent', Philosophia Verlag, pp. 151–186.

Priest, G. (1987), *In Contradiction: A Study of the Transconsistent*, Martinus Nijhoff, The Hague.

Priest, G. (1991), 'Minimally inconsistent LP', **50**, 321–331.

Priest, G. (1997), 'Editor's introduction', *Notre Dame Journal of Formal Logic* **38**(4), 481–487. Special Issue on Impossible Worlds.

Priest, G. (2001), Logic: one or many?, *in* B. Brown & J. Woods, eds, 'Logical Consequences', Hermes Science Publishers. 1999 typescript cited in (Beall & Restall 2001).

Prior, A. N. (1967), *Past, Present and Future*, Clarendon Press, Oxford.

Psillos, S. (1999), *Scientific Realism: How Science Tracks Truth*, Routledge, London.

Psillos, S. (2000), 'Rudolf Carnap's 'Theoretical Concepts in Science'', *Studies in History and Philosophy of Science* **31**, 151–172.

Putnam, H. (1977), 'Realism and reason', *Proceedings of the American Philosophical Association* **50**, 483–498.

Putnam, H. (1981), *Reason, Truth and History*, Cambridge University Press, Cambridge.

Putnam, H. (1983), 'Vagueness and intuitionistic logic', *Erkenntnis* **19**, 297–314.

Quine, W. V. (1970), *Philosophy of Logic*, Prentice Hall, Upper Saddle River, N.J.

Quine, W. (1953), Logic and the reification of universals, *in* 'From a Logical Point of View', Harvard University Press, Cambridge, Mass.

Quine, W. (1963), *Set Theory and Its Logic*, Harvard University Press, Cambridge, Mass.

Ramsey, F. P. (1960), Theories, *in* R. B. Braithwaite, ed., 'The Foundations of Mathematics and Other Logical Essays', Littlefield and Adams, Paterson, NJ, pp. 212–236. Originally published 1929.

Read, S. & Wright, C. (1985), 'Harrier than Putman thought', *Analysis* **45**, 56–58.

Read, S. (1988*a*), 'The irrelevance of the concept of relevance to the concept of relevant consequence', Conference on Knowledge, Logic, Information, Darmstadt, February 1998. Published as 'Logical Consequence as Truth-Preservation', *Logique et Analyse*, 2004.

Read, S. (1988*b*), *Relevant Logic*, Blackwell, Oxford.

Read, S. (2004), In defence of the dog: Reply to Restall, *in* S. Rahman & J. Symons, eds, 'Logic, Epistemology and the Unity of Science', Kluwer.

Restall, G. (1997), 'Ways things can't be', *Notre Dame Journal of Formal Logic* **39**, 583–596.

Restall, G. (1999), Negation in relevant logics (How I stopped worrying and learned to love the Routley star), *in* D. Gabbay & H. Wansing, eds, 'What is Negation?', Kluwer, pp. 53–76.

Restall, G. (2000), *An Introduction to Substructural Logics*, Routledge.

Restall, G. (2002), 'Carnap's tolerance, meaning and logical pluralism', *Journal of Philosophy* **99**, 426–443.

Restall, G. (2004), Logical pluralism and the preservation of warrant, *in* S. Rahman & J. Symons, eds, 'Logic, Epistemology and the Unity of Science', Kluwer.

Routley, R. (n.d.), Every logic on a λ-categorial (i.e. Church type-theoretical) language has a two-valued worlds semantics—thereby supply logics and semantics for English, Typescript.

Russell, B. (1908), 'Mathematical logic as based on the theory of types', *American Journal of Mathematics* **30**, 222–262. Reprinted in Russell, *Logic and Knowledge* (ed. Marsh) (London: Unwin, 1956) and (van Heijenoort 1967).

Russell, B. (1912), *The Problems of Philosophy*, Hackett, Indianapolis. A reprint of the 1912 Home University Library edition.

Russell, B. (1919), *Introduction to Mathematical Philosophy*, George Allen and Unwin, London.

Russell, B. (1924), Logical atomism, *in* R. Marsh, ed., 'Logic and Knowledge: Essays 1901-1950', Routledge, London.

Russell, B. (1927), *The Analysis of Matter*, Dover, New York. Reprint of 1954.

Sainsbury, R. M. (1995), *Paradoxes*, second edn, Cambridge University Press, Cambridge.

Scanlan, M. (1991), 'Who were the American Postulate Theorists?', *Journal of Symbolic Logic* **56**, 981–1002.

Scroggs, S. (1951), 'Extensions of the Lewis system S5', *Journal of Symbolic Logic* **16**, 112–120.

Shapiro, S. (1985), Epistemic and intuitionistic arithmetic, *in* S. Shapiro, ed., 'Intensional Mathematics', North Holland, Amsterdam, pp. 11–46.

Shapiro, S. (2004), Simple truth, contradiction, and consistency, *in* 'The Law of Non-Contradiction: New Philosophical Essays', Oxford University Press, pp. 336–354.

Simmons, K. (1993), *Universality and The Liar*, Cambridge University Press.

Smiley, T. (1959), 'Entailment and deducibility', *Proceedings of the Aristotelian Society* **59**, 233–254.

Stein, H. (1967), 'Newtonian space-time', *The Texas Quarterly* **10**, 174–200.

Tait, W. (1994), The law of excluded middle and the axiom of choice, *in* A. George, ed., 'Mathematics and Mind', Oxford University Press, Oxford, pp. 45–70.

Tarski, A., Mostowski, A. & Robinson, R. M. (1953), *Undecidable Theories*, North-Holland, Amsterdam.

Tarski, A. (1959), What is elementary geometry?, *in* L. Henkin et al., eds, 'The Axiomatic Method', North-Holland, Amsterdam. Reprinted in J. Hintikka, ed., *Philosophy of Mathematics*, London: Oxford University Press, 1969.

Tennant, N. (1987), *Antirealism and Logic*, Clarendon Press, Oxford.

Tennant, N. (2000), 'Deductive versus expressive power: a pre-Gödelian predicament', *Journal of Philosophy* **97**, 257–277.

Thompson, S. (1991), *Type Theory and Functional Programming*, Addison-Wesley, Wokingham, England.

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

van Fraassen, B. (1989), *Laws and Symmetry*, Oxford University Press, Oxford.

van Heijenoort, J., ed. (1967), *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*, Harvard University Press, Cambridge, Mass.

Varzi, A. (1997), 'Inconsistency without contradiction', *Notre Dame Journal of Formal Logic* **38**, 621–639.

Vaught, R. (1986), Introduction to 1934c and 1935, *in* S. Feferman, J. W. Dawson, S. C. Kleene, G. H. Moore, R. M. Solovay & J. van Heijenoort, eds, 'Kurt Gödel: Collected Works', Vol. 1, Oxford University Press, New York, pp. 376–379.

Wajsberg, M. (1933), 'Ein erweiterter Klassenkalkül [an extended calculus of classes]', *Monatshefte für Mathematik und Physik* **40**, 113–126.

Weyl, H. (1922), *Space–Time–Matter*, Dover, New York. Reprint of 1922, Henry L. Brose translation.

Whitehead, A. N. & Russell, B. (1912), *Principia Mathemtatica*, Vol. 2, Cambridge University Press, Cambridge.

Williamson, T. (1994), 'Never say never', *Topoi* **13**, 135–45.

Williamson, T. (1996), 'Putnam on the sorites paradox', *Philosophical Papers* **25**, 47–56.

Williamson, T. (1997), Vagueness and ignorance, *in* R. Keefe & P. Smith, eds, 'Vagueness: A Reader', MIT Press, Cambridge, Mass., pp. 265–280.

Williamson, T. (2000), *Knowledge and its Limits*, Oxford University Press, Oxford.

Winnie, J. (1967), 'The implicit definition of theoretical terms', *British Journal for the Philosophy of Science* **18**, 223–229.

Winnie, J. (1970), Theoretical analyticity, *in* R. Cohen & M. Wartofsky, eds, 'Boston Studies in the Philosophy of Science', Vol. 8, Reidel, Dordrecht and Boston, pp. 289–305.

Wright, C. (1987), 'Further reflections on the sorites paradox', *Philosophical Topics* **15**, 227–290.

Wright, C. (1992*a*), 'Is higher order vagueness coherent', *Analysis* **52**, 129–39.

Wright, C. (1992*b*), *Truth and Objectivity*, Harvard University Press, Cambridge, Mass.

Yagisawa, T. (1988), 'Beyond impossible worlds', *Philosophical Studies* **53**, 175–204.

Zermelo, E. (1909), 'Sur les ensemble finis et le principe de l'induction complète', *Acta Mathematica* **32**, 185–193.

INDEX