DUMMETT AND PUTNAM: REALISM UNDER ATTACK

DUMMETT AND PUTNAM:
REALISM UNDER ATTACK

By

MARK QUENTIN GARDINER, B.A, M.A

A Thesis

Submitted to the School of Graduate Studies

for the Degree of

Doctor of Philosophy

McMaster University

DOCTOR OF PHILOSOPHY (1994) McMASTER UNIVERSITY
Hamilton, Ontario

TITLE: Dummett and Putnam: Realism Under Attack

AUTHOR: Mark Quentin Gardiner B.A. (University of Calgary)
M.A. (University of Calgary)

SUPERVISOR: Professor Nicholas Griffin

NUMBER OF PAGES: vi, 372

ABSTRACT

Realism has traditionally been a philosophical doctrine embodying an ontological element asserting the existence of various types of entities and a meta-theoretic element asserting that the existence of those entities is independent of our knowledge of their existence. Anti-realism, on the other hand, denies that the existence of objects is independent of our knowledge.

Recently, attempts have been made to reinterpret the basic realist/anti-realist dispute in semantic terms. Basically, realism would be the view that the truth (or falsity) of sentences are independent of our knowledge of their truth-values. Anti-realism, on the other hand, would hold that truth is not so independent of our knowledge.

Michael Dummett and Hilary Putnam have presented two of the most famous extended semantic criticisms of metaphysical realism. Dummett argues that realism is committed to an unacceptable theory of meaning. Putnam argues that realism rests upon incoherent assumptions regarding truth and reference.

Unlike many commentators, I accept basic Dummettian constraints. I argue, however, that his conclusions do not follow. Not only can the semantic realist conform to his constraints, a realist construal of truth is in fact ineliminable in such an account. Thus, I turn Dummett's framework against its own conclusions.

Regarding Putnam, I proceed by rejecting his premises. I show that the arguments he constructs do not support the claim of incoherence levelled at metaphysical realism. Often, indeed, his arguments, if carefully understood, actually support realism.

I thus conclude that the two most famous and formidable attempts to reject metaphysical realism on the basis of semantic considerations fail. As such, there is no reason to abandon realism traditionally understood.

ACKNOWLEDGEMENTS

I would foremost like to thank Dr. Nicholas Griffin for his strong encouragement of this project, as well as his invaluable guidance in its direction and comments on its mis-direction. I would especially like to thank his willingness to take on the additional burdens of being a long distance supervisor.

I would also like to thank all my fellow students who guided my thinking about these matters by either agreeing or disagreeing with me. In particular, I want to thank Randy Metcalfe, Anthony Jenkins, Glen Baier, and Felix Ó Murchadha. I wish that they may also soon feel the satisfaction of finishing.

Finally, I would like to thank my cat, Cinder, for reminding me that occasionally arguments need to be shredded.

TABLE OF CONTENTS

INTRODUCTION

1.0 THE DEBATE

> The field and the frogs in it, the sun which shines on them, are there whether I look at them or not.[1]

As modern Anglo-American analytic philosophers, and especially Michael Dummett, owe much to Frege, it is appropriate to open with him. As an introduction to a work on the realism/anti-realism debate, this quote is a bit out of context, but it does serve to express nicely the two main theses falling under the general rubric of 'realism'. First of all, Frege makes an *existential* claim - fields, frogs, and the sun exist: 'realism' unquestionably involves *ontological* or *metaphysical* elements. A moral realist might, for example, assert the existence of values or moral facts. A mental realist might assert the existence of private mental events. A 'scientific' realist might assert the existence of 'tokens of most current unobservable scientific physical types'.[2] An *anti*-realist, in this sense, would be one who denied that such entities exist. Thus, emotivists, behaviourists, and instrumentalists can all be regarded as *anti*-realists. Secondly, Frege implies that the *being* is distinct or independent from the *being perceived*; the things which are exist independently of our perceiving, or knowing, that they exist. In this sense, 'realism' is a overarching or 'meta-theoretic' position concerning, broadly, the relation between metaphysics (or ontology) and epistemology; the realist would maintain that, in the order of conceptual priority, ontology does not depend on epistemology.

---

[1]Frege (1918) p. 12.

[2]Devitt (1991) p. 24.

Realists in general tend to ground epistemology in ontology - we 'know' the things we do in virtue of the access we have to the 'things' which exist independently of our knowledge of them.[1] *Anti*-realists of this second sort, in general, reverse the order of priority - the 'things' which are exist in virtue of our knowledge of them. The quintessential statement of this sort of anti-realism comes from Berkeley: *esse est percipi*; to be is to be perceived. Summing up, we can combine the two elements and say that a realist (towards X) is one who asserts that tokens (of type X) exist independently of our knowledge of them. An anti-realist of the first variety would deny that there exist tokens (of type X) *simpliciter*, while anti-realists of the second variety would deny that tokens (of type X) exist *independently* of our knowledge.[2] It is only the debate between realists and anti-realists of the second sort which will concern us.

In characterizing anti-realism, I have been alluding to the 'epistemological turn' - the 'movement' originating in the 17th century which sought to dislodge metaphysics from its position of 'first philosophy' and replace it with epistemology. This century has seen a similar 'linguistic turn'. We can express the essential realist and anti-realist positions in linguistic terms: given that 'truth' is one of the primary semantic concepts, realism is the view that the truth-values of sentences are independent of our determination of them, whereas anti-realism is the view that the truth-values of sentences are not independent of our determination of them. According to realism, truth is

---

[1]The sceptic, of course, denies that we have a reliable access to the knowledge-independent world, and thus denies that we have genuine knowledge.

[2]See Devitt (1991) Ch. 2 for a similar characterization of realism and anti-realism.

primarily a non-epistemic notion, whereas according to anti-realism, truth is primarily an epistemic one.

However, it is one thing to acknowledge that the core elements of realism and anti-realism can be expressed in linguistic or semantic terms and quite another to accept, as Michael Dummett and Hilary Putnam do, that therefore the debate is essentially a semantic one, or that semantical arguments can settle the issue. I accept the former but reject the latter. Besides the goal of clearly showing why Dummett and Putnam do not succeed in discrediting realism, I hope to lay a strong foundation for the repudiation of any similar semantic approach. The realism/anti-realism debate is, it seems to me, primarily a *metaphysical* one concerning the nature and population of 'reality'. Obviously, many epistemological and semantic issues will have a bearing on such a debate (as will many others), but from that it does not follow that the metaphysical controversy can be solved by merely semantic or epistemological considerations. While I cannot, sadly, claim that *all* semantic arguments against realism are inadequate, I hope that by undermining the two strongest such arguments on the market - Dummett's and Putnam's - there will at least be a strong presumption against such approaches.

2.0 STRATEGY

At its most basic level, realism is the view that reality is discovered, not invented. Human knowledge is a function of attempting to mirror that reality. Anti-realism, on the other hand, is crudely the view that reality is invented, not discovered. Reality - i.e. what we take to be reality (for there is nothing else to be called 'reality') - is a function

of or a projection of human knowledge and human practices. Put in these terms, it seems to me that the view of realism is so dominant and 'natural' in our view of ourselves and the world that there is a certain presumption in its favour. As Rasmussen and Ravnkilde observe, "a demonstration of the superiority of anti-realism over [realism] will necessitate the most radical imaginable revision of our wonted conception of reality."[1] While I agree that such *prima facie* presumption can in no way count as evidence *for* the realistic outlook, it nonetheless has argumentative implications. More precisely, given that there is an initial presumption in favour of realism, the burden of proof initially lies with the anti-realist. Moreover, the anti-realist's burden of proof is two-fold: not only must they provide us with persuasive arguments advocating the adoption of an anti-realist attitude, they must also provide us with persuasive arguments for rejecting realism. That is, it is not enough for an anti-realist to merely argue positively *for* their position - they must also argue negatively *against* realism.

There is, it seems to me, a logical relation between these two burdens. Rejecting realism is a precondition for accepting anti-realism: a persuasive argument for the latter will either involve or presuppose a persuasive argument for the former. Moreover, for an argument to be seen as a successful rejection of some established position as opposed to merely pointing out some difficulties, it must be the case that there is an alternative position proposed. The success of such an argument will in turn depend upon the proposed alternative not falling prey to the very difficulties facing the established position. If the proposed alternative is no better off, then such an argument fails to be

---

[1]Rasmussen and Ravnkilke (1982) p. 382.

a *rejection* of the established position. Therefore, the onus on the anti-realist is to demonstrate that there are some irresolvable difficulties for realism and that anti-realism is immune from them.

This being the case, my support of metaphysical realism can be seen largely in negative as opposed to positive terms. That is, I do not see my project as one of giving persuasive arguments as to why one ought to adopt metaphysical realism, rather I see it as one giving persuasive arguments as to why one ought not to give it up.[1] One ought not to give it up if the arguments advanced against it are not successful or if anti-realism, as the proposed alternative, is no better off. My counter-arguments to those advanced by the anti-realists are all, more or less, aimed at establishing one or the other of these conclusions.

Concerning Dummett's attack on realism, my specific strategy is to largely grant him his premises - that an adequate theory of meaning must harmonize with an adequate theory of understanding (which requires that we be capable of *manifesting* our sentential understanding) - but deny his conclusions.[2] In particular, I will show that a semantic

---

[1]There have been attempts to generate 'a priori' arguments for realism (e.g. McGinn (1979), as well as common scientific realist arguments that only a prior acceptance of reality existing independently of our theorizing could explain the undeniable success of science and the so-called historical 'convergence' of scientific theories (e.g. Putnam (1976a)) as well as some against realism (e.g. Luntley (1988) Ch. 1). I have no interest in either defending or rejecting such attempts - so-called 'a priori' arguments are rarely persuasive.

[2]I am also willing to concede his views about what an adequate theory of meaning would look like. Much of Dummett's work is devoted to rejecting so-called holistic theories of meaning and defending his own *molecular* account. By accepting his constraints I am able to sidestep this entire debate.

realist, which takes a recognition-transcendent (i.e. non-epistemic) notion of truth as its central concept, can meet all of Dummett's challenges.

Concerning Putnam's attack, my strategy is to deny his premises. I deny that we cannot make sense of an epistemically ideal theory failing to be true, that we cannot make sense of our being fundamentally mistaken about the nature of 'reality', and that we cannot make sense of there being a unique and privileged description or theory of 'reality'.

3.0 STRUCTURE OF THE ARGUMENT

3.1 Section I: Dummett

Dummett's overall goal is to demonstrate the inadequacies of a general realist metaphysics - the view that reality is, by and large, unconditioned by human conceptual scheming: that "our sense experiences are [not] constitutive of the world of macroscopic material objects", that "a mathematical proposition describes, truly or falsely, a reality that exists independently of us", that "a person's observable actions and behaviour are [merely] *evidence* of his inner states - his beliefs, desires, purposes, and feelings", that "science progressively uncovers what the world is like in itself", that "an ethical statement is as objectively true or false as one about the height of a mountain", and so on.[1] These various disputes are both diverse and unified; they are diverse in that they range over seemingly distinct subjects while they are unified in that there is a common conception of reality that running through them:

---

[1]Dummett (1991b) pp. 4-6.

> We are swimming in deep waters of metaphysics. How can we attain the shore? These various metaphysical controversies have a wide range of subject matters but a marked resemblance in the forms of argument used by the opposing factions. No doubt, light will be cast upon each of these disputes by studying them comparatively; even so, we need a strategy for resolving them. Our decisions in favour of realism or against it in any one of these instances must certainly make a profound difference to our conception of reality...[1]

Dummett sees the common thread running through all realist positions as this: statements in the disputed class are objectively true or false of a reality independently of our theorizing. A realist position, then, is only as tenable as the view that truth can be so characterized. His goal is to reject the possibility of such a construal of truth.

Truth, he argues, is also the central concept in a theory of meaning. To know the meaning of a sentence is to know under what conditions it would be true. It would seem to follow, then, that admissible characterizations of truth would be conditioned by the kinds of meaning we are capable of attaching to our sentences. Furthermore, the kinds of meaning that we are capable of attaching to our sentences are determined entirely by how we use our sentences. A construal of truth incompatible with facts about how we can actually use sentences would be inadequate.

Thus, Dummett argues that a metaphysically realist position would rest upon what he calls semantic realism - the view that to know the meaning of a sentence is to know under what conditions it would be true, where the sentence's truth-conditions potentially transcend our capacity to recognize when they obtain. A metaphysically anti-realist position would similarly rest upon what he calls semantic anti-realism - to know the meaning of a sentence is to know under what conditions it would be true, where a

---

[1]Dummett (1991b) p. 8.

sentence's truth-conditions cannot transcend our capacity to recognize when they obtain. Under an anti-realist construal of truth, then, truth-conditions would be co-extensive with (hence, for all intents and purposes, reducible to) verification-conditions. Truth, for the semantic anti-realist, just is warranted assertibility. To decide between, for any given area of discourse, a metaphysically realist position and a metaphysically anti-realist position, then, it would suffice to decide between a semantic realist position and a semantic anti-realist position. Dummett's overall strategy is to reject semantic realism (the negative programme - §2.2.1), and demonstrate the necessity of semantic anti-realism (the positive programme - §2.2.2).

He presents two arguments to support the negative programme - the acquisition argument (§2.2.1.1), which argues that we simply could not have acquired a realist conception of truth, and the manifestation argument (§2.2.1.2) which argues that, even if we have acquired such a conception, it cannot be the central one in any adequate theory of meaning. It will be argued that the admissibility of the acquisition argument depends entirely upon the admissibility of the manifestation argument, and thus only the latter need concern us. The admissibility of the manifestation argument, it will be shown, rests entirely upon the existence of undecidable sentences. However, it is argued that, once we arrive at the most reasonable understanding of what it is for a sentence to be undecidable, none of the candidates Dummett presents for such undecidable sentences are genuinely undecidable (§3.1). Formally undecidable sentences, of which there is no such similar dispute, are not of the appropriate type to generate the manifestation argument. Finally, it will be demonstrated that no anti-realist can, on pain

of contradiction, present any sentence as undecidable in the way needed to generate the manifestation argument (§3.2). No anti-realist, then, can issue the manifestation argument against semantic realism, and as such no evidence of the inadequacy of semantic realism will have been presented. The negative programme will simply fail.

The positive programme, on the other hand, assumes that only success in a testing procedure is sufficient for attributions of sentential understanding. It is a precondition of such testing procedures that conditions under which a sentence is correctly assertible be recognizable ones. Thus, given that assumption, no one could manifest an understanding of a sentence if such understanding consisted in realist truth-conditions. Only anti-realist verification-conditions are guaranteed of being recognizable, and thus any adequate theory of meaning must take anti-realist verification-conditions as opposed to realist truth-conditions as its central concept.

However, the fulcrum of the negative programme is the supposed existence of certain sorts of undecidable sentences: namely sentences which we can neither verify nor falsify. But if such sentences exist, then because they lack verification-conditions, they lack *recognizable* verification-conditions. It would follow, then, that whether one is a semantic realist or a semantic anti-realist, *if* the only condition for attributing sentential understanding to someone were their success in a testing procedure, such sentences would have to be incomprehensible (and semantic anti-realism, as well as semantic realism, would be inadequate) (§4.1).

The moral to draw, for either the realist or the anti-realist, is that it is simply a mistake to suppose that *only* success in a testing procedure is sufficient for attributions

of sentential understanding. In particular, given the compositional nature of language, the capacity to issue a (meaningful) sentence consisting of a (potentially novel) configuration of constituents already demonstrated to be understood will also suffice for attributions of sentential understanding. Thus, nothing precludes one from understanding a sentence with unrecognizable truth-conditions on the basis of understanding its constituent components as well as how those components are internally related. Verification-conditions may the *a* central concept in a theory of meaning, but not necessarily the *only* one (§4.2).

In addition, recognizing the role of compositionality in language, it will be argued (§4.3) that no theory of meaning which takes only verification-conditions as its central concept can completely characterize all of the compositional facts of the language. In particular, the meaning of a conditional sentence cannot be recovered merely from the meanings of its constituents (as well as its internal structure) if those constituent meanings are exhausted by their verification-conditions. Meaning must, then, transcend verification-conditions. Quite simply, the positive programme will fail.

Finally, a strong case will be made for construing these 'extra' conditions in terms of realist truth-conditions: i.e. that realist truth-conditions are as ineliminable in an adequate theory of meaning as anti-realist verification-conditions. Thus, at one stroke both the positive and the negative programmes will fail. However, it will also be demonstrated that no theory of meaning which takes only realist truth-conditions as its central concept can completely account for all compositional facts. I will argue that just as verification-conditions must be supplemented with realist truth-conditions in an

adequate anti-realist theory of meaning, so too must realist truth-conditions be supplemented with verification-conditions. In other words, at the end of the day, there will be no difference between (adequate) realist and (adequate) anti-realist theories of meaning. The type of theory envisaged is one in which a sentence contributes both its verification-conditions and, where they differ, its truth-conditions to any compound sentence of which it is a component. When a sentence lacks verification-conditions, then it contributes only its truth-conditions.

I end by conceding the importance of verification-conditions to the determination of sentential meaning, and thus make substantial concessions to Dummett. However, those concessions still allow me to retain (and indeed force me to retain) a robust realist conception of truth. That is all that I need to avoid rejecting any of the metaphysically realist positions we started with. Dummett's excursion into semantics has, I submit, no metaphysical implications.

3.2 Section II: Putnam

Putnam's overall goal is to demonstrate that what he calls metaphysical-realism - defined as the conjunction of various ontological and semantic theses (§1.1) - is inadequate in virtue of its *incoherence*. Nonetheless, he thinks some of its elements are quite correct. Once its undesirable elements are eliminated, and it is supplemented in various ways, we will be left with a 'realism with a human face' - with what Putnam calls 'internal realism' (§1.2).

However, Putnam's support for internal realism is only as good as his attack on

metaphysical realism - if that attack fails then we will have neither good reason to discard metaphysical realism nor good reason to adopt internal realism. His attack consists of three species of argument: a model-theoretic argument, aimed at demonstrating that truth cannot be (radically) non-epistemic (§2.1.1), the so-called 'brain-in-the-vat' argument, aimed at showing that there can be no 'gap' between language and the world (§2.2.1), and two arguments dealing with the consequences of the possibility of alternative empirically equivalent theories, aimed at showing that ontology and truth must be theory-relative (§2.3.1 and §2.3.2).

The model-theoretic argument purports to show that no sense can be made of the claim that a theory which is merely epistemically ideal might, in reality, be false. It draws upon certain results in model theory - in particular how truth can be defined in terms of an 'intended' mapping function between linguistic terms and items in the domain. The argument rests upon two assumptions: that there is no theory-neutral way of understanding the relation of our language to 'reality', and that we can make sense of such an epistemically ideal theory. I argue that neither of these assumptions are warranted (§2.1.2).

The brain-in-a-vat argument purports to show that we "cannot really, actually, possibly *be* brains in a vat"[1] and consequently, upon generalizing, no sense can be made of truth failing to be co-extensive with what we are warranted to assert. I will demonstrate that the argument question-beggingly assumes that truth must be co-extensive with correct assertion (§2.2.2), and that also there is a serious tension between

---

[1]Putnam (1981b) p. 15.

the brain-in-a-vat argument and the model-theoretic argument (§2.2.3). Eliminating that tension will be tantamount to either rejecting the model-theoretic argument or the brain-in-a-vat argument.

Finally, I consider two arguments from the possibility of there being empirically equivalent but distinct theories or descriptions of the world. The first, involving the claim that such theories are mutually incompatible, purports to show that there cannot be a unique and complete true description of the world (§2.3.1.1). However, a strong case is made for denying that there could be such alternative theories (§2.3.1.2), and that once the argument is clearly understood it actually *supports* metaphysical realism (§2.3.1.3). The second, involving the claim that such theories are not, in fact, incompatible, purports to show one may retain a theory-neutral conception of the world only at the cost of relegating the world to an unintelligible Kantian 'thing-in-itself' (§2.3.2.1). I will demonstrate that the argument is circular - it either assumes an prior rejection of the metaphysical realists' ontological theses in order to reject their claims about the nature of truth, or else it assumes a prior rejection of the metaphysical realists' theses about truth in order to reject their ontological claims (§2.3.2.2).

All in all, I contend that Putnam's excursion into semantics has, like Dummett's, no metaphysical implications. We simply have not been given sufficient reason to abandon our ordinary and common-sense conception of reality as existing independently 'out there'.

SECTION I: DUMMETT

1.0 THEORIES OF MEANING

There are and have been many types of 'realisms': mental, moral, modal, mathematical, about the past, about the future, about universals, about theoretical entities, about macroscopic physical objects, and so on. Reflection on the vast variety of realisms at once allows us to appreciate Dummett's main contributions to the topography of the debates that surround them.

Traditionally, realisms have been bound together as ontological doctrines asserting the existence of entities peculiar to them: private mental states for the mental realist, values for the moral realist, possible worlds for the modal realist, and so on. The enemies of these doctrines have tended to deny the existence of such entities: behaviourism against mental realism, constructivism against mathematical realism, emotivism against moral realism, and so on.

Any ontological doctrine essentially contains at its core some set of existential statements; e.g. 'there are private mental events' for the mental realist or 'there are moral facts' for the moral realist. Thus, we can cursorily characterize a realism about X as the position which asserts 'there are X's'.[1] Put this way it is trivial, hence

---

[1]Thus stated, the term 'realism' is seen as containing a disguised relation: 'realism *towards X*'. As such, realism/anti-realism debates are generally conceived of as locally surrounding some particular issue or discourse - realisms are individuated by what they are respectively realisms *about*. Even though Dummett's main argument is perfectly general, he tends to conceive of such debates in local terms, generally eschewing the possibility of a global anti-realism (e.g. Dummett (1982)). Young (1987) argues that, if one is a coherentist, a case for global anti-realism as a legitimate contender can be made. However, it seems to me that, appearances to the contrary, Dummett *does* make

unobjectionable, to say that a realism about X is true just in case its associated existential claims are true. Though trivial, this way of putting things does serve to focus on the fact that the debates surrounding a realism are ultimately debates about the truth or falsity of *statements*. Once this is seen, then it must be admitted that a preliminary debate about the nature of truth is unavoidable in any debate surrounding a realism.[1]

As with Davidson I take it as a datum that "the truth of an utterance depends on just two things: what the words as spoken mean, and how the world is arranged."[2] This being the case, if the realism/anti-realism debate, in whatever form it takes, rests upon a debate over the nature of truth, and the truth of an utterance depends, at least in part, upon what it means, then it is a short step to conceding that debates over realisms involve at their core debates about meaning. This is the first of Dummett's contributions: the traditional ontological disputes about the existence of classes of entities ought to be replaced by semantic disputes about the type of meaning possessed by various classes of sentences:

> This now provides us with a line of attack upon these problems. Instead of tackling them from the top down, we must do so from the bottom up. An attack from the top down tries to resolve the metaphysical problem first, then to derive from the solution to it the correct model of meaning, and the appropriate notion

---

a case for global anti-realism; local disputes would come down merely to whether a particular discourse contained only decidable sentences.

[1]It may be objected that this is true of any debate, whether or not it concerns a realism, and hence is superfluous. That objection is correct in so far as it goes, but as we shall see, in a very important sense, Dummett makes the debate about the nature of truth *constitutive* of any debate concerning a realism - more aptly, it is what constitutes *the* realism/anti-realism debate. It is for this that it is not unreasonable to regard the realism/anti-realism debate as one of the most pressing in philosophy today.

[2]Davidson (1981) p. 309.

> of truth, for the sentences in dispute...
>
> To approach these problems from the bottom up is to *start* with the disagreement between the realist and the various brands of anti-realist over the correct model of meaning for statements of the disputed class, ignoring the metaphysical problems at the outset.[1]

Now it is important to keep in mind, although it tends to be overlooked, that Dummett is offering a way *through*, not a way *around*, the various metaphysical problems we started with; he is, after all, offering a 'logical' basis of *metaphysics*. He starts 'at the bottom' as being the most fruitful way of arriving 'at the top'. Thus, his proposal depends upon there being some intimate connection between the semantic issues of meaning and truth and the metaphysical issues of the nature of reality; namely that reality is however true sentences express it to be, and a sentence is true just in case it accurately states how reality is constituted.[2] As Dummett says, "having first settled on the appropriate notion of truth for various types of statement, we conclude from that to the constitution of reality".[3]

---

[1]Dummett (1991b) p. 12.

[2]At this point, the expressed relation is not meant to be more than platitudinous.

[3]Dummett (1976b) p. 89. See also McDowell (1976) §2. Not all philosophers, even of an anti-realist bent, accept Dummett's claim of the relationship between metaphysics and semantics. For example, see Devitt's (1991) Maxim 2: Distinguish the metaphysical (ontological) issue of realism from any semantic issue, and Maxim 3: Settle the realism issue before any epistemic or semantic issues. Essentially, Devitt's maxims amount to a rejection of Dummett's entire proposal of how to attack the issue. Tennant (1987) denies that semantic anti-realism has any metaphysical consequences (pp. 10-11). McGinn (1976), on the other hand, sees Dummett's argument as having just this force, but resists the conclusion by arguing that the requirements of a theory of meaning can be satisfied by a realist semantics. Young (1992), by opting for a coherentist understanding of assertibility-conditions, argues that the semantic positions of realism and anti-realism have *no* metaphysical consequences, and that consequently "anti-realism does not stand or fall with any metaphysical hypothesis. Realists must stop thinking that they refute anti-realism by arguing against idealism and other metaphysical positions."

For Dummett, there is also an intimate relationship between truth and meaning - the nature of truth can be determined from a correct theory of meaning, and a correct theory of meaning is formulated by reference to truth.[1] A meaning-theory (for a

---

(p. 76). However, one can agree with Young that refuting idealism (or 'other metaphysical positions') does not entail a refutation of anti-realism, but from this it does not follow that an acceptance of anti-realism would not entail an acceptance of idealism. On the other hand, Loar (1987) argues that "semantic realism is not *necessary* for realism" (p. 93). My own view is that (semantic) anti-realism does entail some species of idealism but, as we do not have sufficient reason to accept (semantic) anti-realism, we do not have sufficient reason to accept idealism (and indeed, every reason not to).

[1]The following account of what a theory of meaning would look like is intended to be informal and sketchy but it should serve as sufficient background to understand Dummett's main arguments. For a formal and more fully worked out account, see Davies (1981) Ch. 1 and Lycan (1984); for a more informal treatment, see C. Wright (1986). For Dummett's precise understanding of a theory of meaning, see Dummett (1975), (1976b), (1982), and (1991b). Essentially, he sees a theory of meaning as a tripartite structure consisting of a core and two concentric outer shells. At the core lies the theory of truth, which yields theorems specifying "the way in which the semantic value of a sentence is determined by the semantic values of its components, and [gives] the general condition for a sentence to be true, in terms of its semantic value." (Dummett (1991b) p. 61). (In general, the semantic value of a name is an object, of a predicate is a function, and of a sentence is a truth-value.) The theory of sense occupies the first shell, which consists of a specification of the practical abilities speakers may manifest regarding the theorems in the core (i.e. it is a theory of understanding). The second outer shell is occupied by the theory of force, which specifies the various types of conventional linguistic significance, such as assertion, requesting, making commands, etc. Dummett (1991b) Ch. 5 also throws in the notion of *tone*. Tone seems to deal with subtle differences of use, such as that between 'and' and 'but', 'dead' and 'deceased', etc. It "serves to define the proposed *style* of discourse, which, in turn, determines the kind of thing that may appropriately be said." (p. 112). It is a constraint on any theory of meaning that the three (or four) parts harmoniously interact - i.e. concepts at one level must be explicated by reference to concepts at lower levels. As we shall see, Dummett's main complaint against realist theories of meaning is that they are unable to harmoniously blend a theory of truth with a theory of sense. Given the downwards relation, and that the theory of truth lies at the core, I shall tend to use "theory of meaning" as referring merely to the core theory of truth, and "theory of understanding" to refer to the theory of sense.

language L)[1] is a formal deductive theory which yields theorems of the form:

*) 'S' means (in L) that P

which specify the meaning of every sentence expressible in L. There are two main constraints on such a theory. In the first place, given that, in natural languages at least, sentences are formed out of words in such a way that a potentially infinite number of sentences can be generated from a finite stock of words, a meaning-theory must reflect the compositional nature of language. It does this by postulating a finite number of axioms specifying the senses of individual words, and then specifying recursive rules for recovering the sense of sentences from the senses of its constituent words. Languages are compositional in another sense: not only are simple sentences composed of words, so too are compound sentences composed of simpler sentences. Thus, an adequate meaning-theory must also contain recursive rules for recovering the meaning of compound sentences from the meanings of their constituent sentences.[2]

The second constraint hardly needs to be stated: the disquoted sense of 'S' in the meaning-specifying theorems cannot differ from the sense of P. However if this were the

---

[1]Dummett, in a later work, uses the expression 'theory of meaning' in the same way we use 'theory of knowledge' - i.e. as a label for a general branch of philosophy of language. He uses the expression 'meaning-theory' as an axiomatic system specifying the meaning of all words and expressions in a particular language (Dummett 1991b). I tend to use the two interchangeably, as does Dummett in his earlier works.

[2]"A theory of meaning will contain axioms governing individuals words, and other axioms governing the formation of sentences: together these will yield theorems relating to particular sentences." (Dummett (1976b) p. 72). Also: "What a semantic theory is required to do, therefore, is to exhibit the way in which the semantic value of a sentence is determined by the semantic values of its components, and to give the general conditions for a sentence to be true, in terms of its semantic value." (Dummett (1991b) p. 61).

only constraint on an adequate meaning-theory then there could be no complaint against a homophonic theory which generated theorems of the form "'S' means (in L) that S", such as:

1) 'Snow is white' means (in English) that snow is white.

Such a theory satisfies the synonymy constraint but is uninformative. It is uninformative in that it takes the notion of meaning for granted. In the first place, by utilizing the notion of meaning in its meaning-specifying theorems, it offers no philosophical illumination of that concept - i.e. no conceptual analysis of the notion of meaning in terms of other concepts. Secondly, by taking that notion for granted, it offers no explanation of in what the meaning of S consists. If one did not understand what the sentence 'snow is white' meant, it would be unhelpful to be told that it meant that snow was white. To avoid such triviality, a meaning-theory should offer meaning-specifying theorems which do not presuppose the concept of meaning.[1] In other words, the meaning-specifying theorems should take some form like:

M) 'S' is T if and only if P.[2]

All that is left, more or less, in constructing an adequate meaning-theory is to find a

---

[1]Dummett (1975) distinguishes between what he calls a 'modest' meaning-theory and what he calls a 'full-blooded' one. A modest meaning-theory is content to yield meaning-specifying theorems without concern for understanding (such as (1)). A full-blooded theory, on the other hand, would also offer explanations of meaning (and other core concepts) to one who did not already accept the theory. Dummett's position is that no merely modest meaning-theory can be an adequate theory of meaning. See McDowell (1987) for an opposing view, as well as Dummett's (1987) reply.

[2]See Davidson (1967) p. 23 and McDowell (1976) §1 for a fuller discussion of this issue.

suitable candidate for the predicate 'T'. As early as Frege, 'truth' has been presented as such a candidate.[1] Davidson, for example, suggests that a truth-theory in the style of Tarski[2] would serve as an adequate meaning-theory:

> There is no need to suppress, of course, the obvious connection between a definition of truth of the kind Tarski has shown how to construct, and the concept of meaning. It is this: the definition works by giving necessary and sufficient conditions for the truth of every sentence, and to give truth conditions is a way of giving the meaning of a sentence. To know the semantic concept of truth for a language is to know what it is for that sentence - any sentence - to be true, and this amounts, in one good sense we can give to the phrase, to understanding the language.[3]

Dummett follows up on this point: "*Every* semantic theory has as its goal an account of the way in which a sentence is determined as true, when it is true, in accordance with its composition."[4] This may strike many as odd, given the common attitude that Dummett is attacking the realist slogan that the meaning of a sentence is given by its truth-conditions. What Dummett is attacking is the claim that the meaning of a sentence is given by its *realist* truth-conditions. According to his view, while we are permitted to assume that truth is the central notion in a theory of meaning, we are not

---

[1]"We are therefore driven into accepting the *truth-value* of a sentence as constituting what it means." (Frege (1892) p. 63).

[2]Tarski (1931) and (1944).

[3]Davidson (1967) p. 24. McGinn (1976) and Scruton (1976) (criticized by C. Wright (1987) Ch. 7) accept the basic Davidsonian point. See also Putnam (1976c) for an interesting discussion concerning Davidson's relation to Tarski.

[4]Dummett (1991b) p. 31. See also (1959a) p. 8: "The sense of a statement is determined by knowing in what circumstances it is true and in what false."

permitted to assume any particular conception of truth.[1]

There is, according to Dummett, one last vital element in any adequate meaning-theory: it must also harmonize with a theory of understanding.[2] It is difficult to dispute that languages are (human) artifacts, conventional in nature, and designed to facilitate communication.[3] Successful communication presupposes that communicators understand each other. Dummett insists that a theory of meaning must harmonize with a theory of understanding in the sense that it must be "a representation of what it is that is known when an individual knows the meaning of a sentence."[4] Now, understanding is a type of knowledge - one who understands something is in possession of a requisite type of knowledge. Thus, a theory of meaning, as consistent with a theory of understanding, must be a representation of what one needs to know in order to know the meanings of the sentences in the language - i.e. in order to understand the language.

---

[1]See Dummett (1991b) pp. 32-33 and pp. 163-164. Compare also to C. Wright's (1992) argument's for a *minimalist* conception of truth. Kirkham (1989) p. 208 cites Dummettian passages to point out a systematic ambiguity in Dummett's usage: (i) 'truth' is to be properly construed in the manner the realist proposes (to be discussed in detail later) and consequently sentential meaning cannot be characterized in terms of truth-conditions; and (ii) sentential meaning is to be characterized in terms of truth-conditions, consequently 'truth' cannot be properly construed in the manner proposed by the realist. In what follows, I opt for understanding Dummett in sense (ii).

[2]Dummett often says that a theory of meaning must also *be* a theory of understanding. Such an identification potentially confuses the semantic concerns of a theory of meaning with the epistemic concerns of a theory of understanding. To avoid such confusion, I keep the two separate. See Dummett (1973a) p. 92, (1975) p. 99, (1976b) §II, and (1991b) Ch. 4.

[3]See, for example, Tennant (1987) p. 13 and Prawitz (1980) p. 3.

[4]Dummett (1973c) p. 217.

Put in this way, it is tempting to conclude that what a competent language user knows is the correct meaning-theory for their language. Moreover, if a correct meaning-theory for a language takes the form of a truth-theory for that language, such that the truth-conditions for any sentence specify its meaning, then knowing the meaning of a sentence will consist in knowing its truth-conditions. The truth-conditions of each sentence are expressed by the truth-theory's T-sentences:

T) 'S' is true iff S

Thus, it would seem that what a competent language user knows, in knowing the meaning of a sentence S, is S's associated T-sentence. But, what is it to know a sentence's associated T-sentence?

We might suppose that it consists in knowing *what* its associated T-sentence *is*. This will not do; all I need in order to know what a sentence's associated T-sentence is is to know how to substitute the sentence for the variable 'S' in (T). Take some English sentence which is initially unfamiliar to me:

2) Whan that April with his showres soote, thanne longen folk to goon on pilgrimages.

Knowing how to substitute sentences for sentential variables will allow me to know:

T_2) 'Whan that April with his showres soote, thanne longen folk to goon on pilgrimages' is true iff whan that April with his showres soote, thanne longen folk to goon on pilgrimages.

but obviously I need to know more than just how to substitute sentences for sentential variables to understand Chaucer. So, we need to know more than merely what a

sentence's associated T-sentence is in order to know its meaning.[1] We must also know what the associated T-sentence means. The obvious problem is that S's associated T-sentence contains S as a constituent (in its unquoted occurrence on the right-hand side of the biconditional). Given the thesis of compositionality, one will understand S's associated T-sentence only if one understands S, and hence the T-sentence would be impotent in giving S's meaning.

This is a *bit* quick - grasp of T_s[2] can perhaps be salvaged by continuing the truth-conditional analysis. If the meaning of a sentence is given by its truth-conditions, then the meaning of T_s should be given by *its* truth-conditions. We may attempt to carry on the proposed analysis to yield the meaning of any arbitrary T-sentence; i.e. to grasp the meaning of a sentence of form T_s is to grasp *its* truth-conditions:

T_s') "'S' is true iff S" is true iff 'S' is true iff S.

There are strong reasons why we should abandon such an approach. In the first (and least persuasive) place, Dummett argues that sentences like T_s' are unintelligible - we do not attach any clear sense to conditionals whose antecedents are themselves conditionals.[3] A more persuasive reason is that, just as T_s contains S as constituent, T_s' contains T_s as constituent. Grasp of the meaning of T_s must be conceptually prior to

---

[1]See Dummett (1991b) pp. 69-70.

[2]I.e. S's T-sentence.

[3]Dummett (1973a) p. 449 (see also (1990)). His argument is far from convincing - there are, it seems to me, perfectly intelligible conditional sentences whose antecedents are themselves conditionals. Consider this situation: I, not owning a car, desire to travel to a conference with Wilma in her 2-seater. I overhear her say "I am only going if Fred goes with me." I can then formulate the intelligible sentence "If Wilma will only go if Fred goes with her, then I had better arrange for alternative transportation."

grasp of the meaning of T_s', hence the meaning of the former cannot be explicated by appeal to the latter. Finally and decisively, the sole motivation for considering the higher-level T-sentence is ultimately to give the meaning of the object-level sentence S. If grasp of S requires grasp of T_s, which requires grasp of T_s', which would, if the analysis were to be carried out, require grasp of T_s'', etc., then one would have to grasp the meaning of an infinite number of sentences in order to grasp the meaning of any single sentence - human cognitive powers are amazing, but not that amazing.

Perhaps the problem stems from the implicit assumption that the T-theory in question will be homophonic. The main problem we have encountered is that the T-sentence used to express the truth-conditions (and hence the meaning) of a given sentence contains that sentence as constituent. But, only T-sentences of the form "'S' is true iff S" contain S as constituent; a heterophonic T-theory, delivering theorems of the form "'S' is true iff P", while *mentioning* S, would not contain it as constituent. Such T-sentences could then express the meaning of S without presupposing it. Consider a heterophonic T-theory which assigns to the sentence:

3) Men outnumber women in this room.

the following truth-conditions:

> T_3) 'Men outnumber women in this room' is true iff one-to-one pairing of men in this room with women in this room would leave at least one man in this room without a counter-part.

Such a heterophonic T-theory would at least have the advantage of not obviously presupposing the meaning of (3) in the expression of its truth-conditions. However, if grasp of (3) is to consist in knowing its associated T-sentence in the only plausible sense

- i.e. in grasping its meaning - then grasp of (3) presupposes a grasp of T_3, which in turn presupposes a grasp of *its* constituents. In particular, it presupposes a grasp of:

> 4) One-to-one pairing of men in this room with women in this room would leave at least one man in this room without a counter-part.

Now, what does a grasp of (4) consist in? Under the proposal, it consists in a grasp of *its* associated T-sentence. If the T-theory which yielded (T_3) as (3)'s meaning-specifying theorem yields:

> T_4) 'One-to-one pairing of men in this room with woman in this room would leave at least one man in this room without a counter-part' is true iff one-to-one pairing of men in this room with woman in this room would leave at least one man in this room without a counter-part.

as (4)'s associated T-sentence, then we are back where we started from. If it yields:

> T_4') 'One-to-one pairing of men in this room with woman in this room would leave at least one man in this room without a counter-part' is true iff men outnumber women in this room.

then grasp of (4) presupposes grasp of (3) *and* grasp of (3) presupposes grasp of (4). The only other possible alternative is for the T-theory to yield some theorem of the form:

> T_4'') 'One-to-one pairing of men in this room with woman in this room would leave at least one man in this room without a counter-part' is true iff R.

where 'R' is some sentence *other* than either (3) or (4). But, such a manoeuvre merely calls for an account of the meaning of R, which cannot be given either homophonically or in terms of (4), and thus must proceed by reference to another sentence Q, etc. Thus, even a heterophonic T-theory faces the intolerable result that grasp of any single sentence requires the grasp of an infinite number of sentences.

The moral to be drawn, I think, is not one of despair. Let us carefully attempt

to reconstruct what has gone on. At the root of the problem is a thesis of truth-conditionality:

a) The meaning of a sentence S is given by its truth-conditions.

Dummett's claim is that a theory of meaning must be closely related to a theory of understanding:

b) Understanding a sentence S consists in knowing its meaning.

From (a) and (b) it would seem a short step to concluding:

c) Knowledge of S's meaning consists in knowledge of S's truth-conditions.

As we saw, knowledge of S's truth-conditions cannot solely consist in knowing *how* to construct a sentence expressing them. It seems not unreasonable, therefore, to conclude such knowledge consists in understanding the meaning of a sentence which expresses them. According to the proposed theory, the sentence which expresses S's truth-conditions is its associated T-sentence, T_s. Thus:

d) Knowledge of S's truth-conditions consists in knowledge of the meaning of T_s.

Keep in mind also the compositional nature of language:

e) Knowledge of the meaning of T_s requires knowledge of the meaning of its constituent sentences.

Now either our meaning-theory is to be homophonic or heterophonic. If it is homophonic, then:

f) T_s contains S as a constituent.

which, together with (e) yields:

g) Knowledge of the meaning of T_s requires knowledge of the meaning of S.

But, combining (c) and (d) yields:

h) Knowledge of the meaning of S requires knowledge of the meaning of T_s.

which conjoined with (g) engenders a vicious circle. Thus, we have reason to believe that our meaning-theory should be heterophonic. But if it is heterophonic, then:

i) T_s contains R as constituent.

where R is a sentence syntactically distinct from S. (i), with (e), yields:

j) Knowledge of the meaning of T_s requires knowledge of the meaning of R.

which brings us back to (a) substituting R for S; there is no stopping and a regress ensues: knowledge of the meaning of any single sentence would require knowledge of the meaning of an infinite number.

In any event, we now have a clear view of the reasoning leading to this undesirable conclusion. Reflection on it will, I believe, expose its suspect underlying assumption: it assumes that knowledge of the meaning of any sentence S depends upon an *explicit* formulation of its meaning in terms of *another* sentence. In other words, it assumes that knowledge of the meaning of a sentence must be given in terms of an *explicit* statement - in sentential form - of the knowledge required in order to understand the given sentence; i.e. that the knowledge required for understanding be an explicit theoretical knowledge. It is that assumption, and that assumption alone, which warrants the move from (c) to (d).

The moral to draw is that ultimately an account of understanding - grasp of meaning - must be given in non-linguistic terms. This is precisely Dummett's proposal: knowledge of the meaning of a sentence must ultimately be an *implicit* practical

knowledge.[1]

A piece of knowledge is implicit if there is warrant to attribute it to someone independently of whether they are able to give an explicit statement of what they know. Contrast this with such a thing as knowing who wrote *The Canterbury Tales* - if one cannot explicitly state 'Chaucer wrote *The Canterbury Tales*' then one cannot claim to know it. By contrast, failure to be able to explicitly state how to ride a bike cannot be taken as evidence that one does not know how to.

Implicit knowledge tends to be knowledge of a practical sort - a *knowing-how* as opposed to a *knowing-that*. Knowing how to do something consists in being able to do it; one is able to do x if and only if one knows how to do x. Dummett, by supposing the knowledge required for understanding is implicit, is committed to the view that understanding a language consists in nothing other than being able to *use* the language in appropriate ways: hence the slogan borrowed from the later Wittgenstein that *meaning consists in use*:

> The meaning of a [statement] determines and is exhaustively determined by its *use*. The meaning of such a statement cannot be, or contain as an ingredient, anything which is not made manifest in the use made of it, lying solely in the mind of the individual who apprehends that meaning: if two individuals agree completely about the use to be made of the statement, then they agree about its meaning.[2]

---

[1]Dummett is occasionally a little vague on why sentential understanding must involve implicit knowledge - he tends to rely a little too heavily on the Wittgensteinian 'Meaning as Use' slogan. The above, if correct, clearly brings out the technical reason. See Dummett (1976b) §I and (1991b) Ch. 4.

[2]Dummett (1973c) p. 216.

Thus Dummett concludes that the only meanings that sentences of the language bear are those which attach to them by the use they are put to by competent language users.[1] Furthermore, by supposing knowledge of meaning to be implicit, Dummett is able to avoid the problems posed for the theory of meaning: the inability to explicitly state, without circularity, what a language-user knows when they understand a sentence cannot be taken as a failure of the theory.

But what then is the relation between the proposed theory of meaning and truth-conditions? The previous argument concluded that the knowledge involved in understanding a sentence S cannot *consist* in knowing what S's truth-conditions are (i.e. in knowing T_s). This is as it should be if understanding involves implicit knowledge - such knowledge *consists* exclusively in a practical capacity, and hence cannot *consist* in knowing *that* S's truth-conditions are such-and-such. Nonetheless, such knowledge can *express* or *represent* the implicit knowledge involved in understanding - i.e. it can be an

---

[1]The meaning-as-use thesis is certainly far from being universally accepted. Tennant (1987) suggests that the Quinean indeterminacy thesis (see Quine (1960) §12, §16 and (1968)) points out that meaning will always be *underdetermined* by use - no amount of overt behaviour by the natives will determine whether 'gavagai' means rabbit or undetached rabbit parts (Loar (1987) raises a similar point, though he ultimately rejects it). We shall see a similar point raised in Putnam's Model-Theoretic argument. Currie and Eggenberger (1983) p. 272 claim that "there is nothing unintelligible or even unsound about supposing that speakers have knowledge of meanings which cannot fully be manifested in behaviour." They argue that Dummett starts "with a behaviouristic picture of human capacities for language acquisition" and then concludes that such a picture "makes it inexplicable how we could employ meanings in the way we [i.e. the realist] think we do." They advocate rather that we "start by taking our linguistic practices at face value and ask 'what capacities must be possessed by human beings which enable them to achieve the type of linguistic competence and understanding they do have?'." (p. 276). See also Craig (1982) p. 554: "Why should we insist that nothing is communicable unless it can be known (let alone *observed*) to have been communicated?" However, I am willing to let Dummett have this assumption.

explicit theoretical representation of what one knows implicitly when one understands a language.[1]

But, if such knowledge is implicit, consisting in a practical capacity, what is the practical capacity in question? If the meaning of a sentence is given by its truth-conditions, as proposed, then the capacity in question must somehow involve truth-conditions. The most natural answer is the capacity to distinguish between those conditions (states-of-affairs) which would render a sentence true and those which would not.

However, Dummett's insistence still holds: it is a constraint on any admissible theory of meaning that it be consistent with an acceptable account of understanding. Only by being consistent with an account of understanding can a theory of meaning conform to the essential communicativeness of language; a theory of meaning which was such that the knowledge it stated explicitly was unable to represent the knowledge one possesses implicitly when understanding an expression would be inadequate.

So, understanding consists in a possessing a certain kind of capacity. What does possessing a certain kind of capacity consist in? It consists in being able to *manifest* that capacity: one knows how to ride a bike only if one can demonstrate it by peddling down the street. Thus, associated with each capacity is a *testing procedure*: *if*, under proper conditions (e.g. a bike is present, one is not restrained, etc.), one can peddle down the

---

[1]As Dummett says, "knowledge of a language is not merely a species of practical competence but is also genuine knowledge, and [the] meaning-theory is intended as an organized and fully explicit representation of the content of that knowledge." (Dummett (1991b) pp. 103-104). For a fruitful discussion of this issue, see Crosthwaite (1983).

street *then* one has the capacity to ride a bike; and *if* one has the capacity to ride a bike, *then*, under proper conditions, one can peddle down the street. In other words, the potential for succeeding in a testing procedure is both a necessary and sufficient condition for possession of a capacity.[1]

In terms of linguistic understanding, one understands the meaning of a sentence only if one is able to manifest that understanding by being able to distinguish those states-of-affairs which render the sentence true and those which do not. This in turn presupposes that there be a testing procedure such that success in it would warrant an attribution of the capacity and hence would warrant an attribution of understanding.

---

[1]See C. Wright (1987) p. 53: "...knowledge of declarative sentence meaning must involve a recognitional capacity: the ability to recognize, if appropriately placed, circumstances which do, or do not, fulfil the truth-conditions of a sentence and to be prepared accordingly to assert to, or withhold assent to, its assertion." There is, however, considerable debate over the manifestation constraint. Appiah (1986), for example, remarks that no realist should accept the thesis as it is just "verificationism dressed up" (p. 22). Page (1991) argues that Dummett has given no good reason to accept it: "The question is: If X has [a grasp] of a concept, why must X's understanding of the expressions associated with that concept be publicly manifestable? Rather than address that question, Dummett simply assumes that X's understanding of the expression must be publicly manifestable. It may be the case that in order for someone other than X to know whether X understands the word 'square', X must manifest that understanding in some appropriate way. But to assume that X does not understand 'square' unless X can publicly manifest that understanding is, again, to assume that X's understanding of 'square' must be publicly manifestable. As I see it, therefore, [Dummett] begs the question." (p. 336). In other words, Page resists Dummett's move from 'Unless X can manifest her understanding, we cannot attribute understanding to X' to 'manifestability is necessary for understanding'. However, for the most part, I am willing to grant Dummett the assumption (it is generally thought to be supported by Wittgenstein's celebrated 'private language' argument, though Page (1991) and Craig (1992) criticize Dummett for not making explicit how he understands Wittgenstein's argument, or how, exactly, it is supposed to support the manifestation constraint). It will be argued (§4.2) that the realist can meet the constraint. See also C. Wright (1986) and Loar (1987) for an interesting discussion on this issue.

Recall that according to the proposed theory of meaning a sentence's truth-conditions are expressed by its associated T-sentence. It would seem, then, that understanding a sentence requires that one is able to associate, with each sentence, its correct T-sentence; but we must be careful. Knowing *what* a sentence's correct T-sentence *is* is not sufficient for grasp of its meaning (recall sentences (2) and (T_2)) - one must also *understand* the associated T-sentence. What we therefore require is a testing procedure. Dummett offers one: we can attribute a grasp of the meaning of a (true) sentence's associated T-sentence to anyone who, when situated favourably to investigate the state-of-affairs referred to in the right-hand side of the biconditional assents to the sentence mentioned on the left-hand side (we can likewise attribute a grasp of the meaning of a false sentence's associated T-sentence to one who, when similarly placed, dissents from the sentence):

> Our model for such knowledge [i.e. "the explanation of what it is for a speaker to know the truth conditions of S"] ... is the capacity to use the sentence to give a report of observation. Thus if someone is able to tell, by looking, that one tree is taller than another, then he knows what it is for a tree to be taller than another tree, and hence knows the conditions that must be satisfied for the sentence, 'this tree is taller than that one', to be true.[1]

Notice that if either a person fails to assent or dissent correctly from the sentence when placed in the relevant state-of-affairs or if it is not possible to place the person in that state-of-affairs then it would seem that there can be no justifiable reason to attribute a

---

[1]Dummett (1976b) p. 95. It is interesting to note that Dummett's proposed testing procedure for understanding is virtually the same as Davidson's empirical test for the adequacy of a T-theory. The difference is this: Davidson's *assumes* that the testee understands the sentence in order to determine the adequacy of the theory, Dummett's assumes the adequacy of the theory in order to determine that the testee understands the sentence.

grasp of the sentence to that person: i.e. if success in a testing situation is an exclusive indication of grasp of meaning, then either failing the test or failing to take the test precludes an attribution of understanding.

2.0 CRITIQUE OF SEMANTIC REALISM

2.1 What Semantic Realism Is

We have so far outlined the essentials of an adequate theory of meaning. At the core of the theory is the thesis that the meaning of a sentence is given by its truth-conditions, hence the meaning-specifying theorems of such a theory will take the form "'S' is true iff P" where 'P' states S's truth-conditions. Furthermore, the theory is intended to be an explicit representation of what one knows implicitly when understanding a language. Hence, if the meaning of a sentence is given by its truth-conditions, then knowing the meaning of a sentence will consist in knowing its truth-conditions in the practical sense that one has a capacity to distinguish between those conditions which render the sentence true and those which do not. To know the meaning of a sentence is to know under what conditions it would be true.

So far very little has been said about the notion of truth - it has merely been taken for granted. As mentioned in the previous section, the realist and the anti-realist differ over their conception of reality. Given the intimate connection between reality and truth discussed in §1, it follows that they must also disagree over the extension of 'truth'. I suggest that we can view the disagreement over the extension of 'truth' as tantamount to a disagreement over the nature of truth.

Realists and anti-realists do not *completely* disagree over the nature of truth - there is substantial overlap in their conceptions. To illustrate this, we need to characterize truth in as neutral a way as possible. Now, both agree that truth is primarily

a property of declarative sentences - i.e. sentences which can be used to make assertions. We can therefore recast the connection between truth and reality in terms of assertion: reality is however true sentences *assert* it to be and a sentence is true just in case what it *asserts* is correct. The second conjunct can be restated in this way: a sentence is true just in case the conditions for its being correctly assertible obtain. Thus, 'truth' and 'correctly assertible' are, for both the realist and the anti-realist, equivalent: a sentence is true just in case it is correctly assertible. To give the notion of truth in as neutral a way as possible, I propose that we replace the expression 'truth' with the expression 'correctly assertible'.[1] T-sentences (from §1) can thus be restated as:

T′) 'S' is correctly assertible iff S.

Now, as with truth, realists and anti-realists disagree over the extension of

---

[1]Dummett maintains that the fairly robust notion of truth or falsity "take their origin" from the "primitive conceptions of the correctness or incorrectness of an assertion". (Dummett (1991b) p. 83). In (1959a) he criticizes the redundancy theory of truth for failing to take account of the 'point' or 'aim' of truth; it is part of the very concept of truth, he says, that we aim at making true assertions. Hence, the notion of assertion (partially) informs the notion of truth. See also (1991b) pp. 165-166: "The root notion of truth is then that a sentence is true just in case, if uttered assertorically, it would have served to make a correct assertion." and (1990) p. 4: "The concept of truth is born from a more basic concept, for which we have no single clear term, but for which we may here use the term 'justifiability'." The 'more refined' notion of truth, as opposed to the 'coarse' notion of justifiability, is needed, he maintains, in order to understand the use of certain logical connectives in ordinary linguistic practice (principally 'if', but also possibly 'or'; this issue will be taken up again in §4.3). C. Wright (1992), on the other hand, distinguishes a minimalist notion of truth (one which does little more than satisfy various platitudes like "a sentence is true just in case it accurately describes what reality is like") and more substantial conceptions. What both C. Wright and Dummett are driving at is that every conception of truth worthy of the name must satisfy some minimal constraints. I maintain that "correctness of assertion" lies at the core of such criteria, and thus can serve as characterizing a neutral conception of truth. See Sintonen (1982) for a criticism of the proposed priority relation between truth and assertion.

'correctly assertible'. The realist allows for the possibility of the correctness of an assertion even in those cases in which we cannot recognize the conditions for its correctness as obtaining, whereas the anti-realist does not allow for such a possibility. In other words, for the realist, a sentence may *be* correctly assertible even if we do not *know* it is correctly assertible. The realist conception of correct assertibility - i.e. the realist conception of *truth* - is therefore essentially a non-epistemic notion; it attaches to sentences quite independently of our knowledge of it so attaching. However, for the anti-realist, a sentence is only correctly assertible if we possess adequate evidence for it; i.e. a sentence cannot *be* correctly assertible if we do not *know* that it is correctly assertible. The anti-realist conception of correct assertibility - i.e. the anti-realist conception of *truth* - is therefore essentially an epistemic notion; it attaches to sentences depending on our knowledge of it so attaching:

> For the realist, our understanding of [a] statement consists in our grasp of its truth-conditions, which determinately either obtain or fail to obtain, but which cannot be recognised by us in all cases as obtaining whenever they do; for the anti-realist, our understanding consists in knowing what recognisable circumstances determine it as true or as false.[1]

Thus, by associating different conditions with the truth (=correctness of assertion) of a sentence realists and anti-realists thereby assign different meanings to their sentences. For the realist the meaning of a sentence is given by a set of conditions which may unrecognizably obtain, whereas for the anti-realist the meaning can only be given by conditions which we can recognize as obtaining when they do. In terms of

---

[1]Dummett (1959a) p. 23. See also Dummett (1963b) pp. 146-147, (1969) pp. 358-359, and (1991b) Intro.

understanding, a realist account consists in a practical capacity to distinguish between those possibly unrecognized conditions which render the sentence true and those which do not.

There are a number of logical and semantic principles which ride coattails on the realist account of meaning. The realist conceives of truth in terms of a pairing of sentences with conditions (states-of-affairs) such that, associated with each sentence S is a states-of-affairs s such that if s obtain then S is true. Thus, every sentence S cuts all possible states-of-affairs in two: it divides them into those whose obtaining is sufficient for its truth and those whose obtaining is sufficient for its falsehood.[1] It is furthermore unobjectionable that, as complementary states-of-affairs are defined as being mutually incompatible, if s obtains then $\bar{s}$ must fail to obtain and if s fails to obtain then $\bar{s}$ must obtain. In other words, for every set of state-of-affairs, either it obtains or it fails to obtain.

Let S stand for a sentence which asserts that some state-of-affairs s obtains. If s does obtain, then what S asserts is correct and hence is true; if s fails to obtain - i.e. $\bar{s}$ obtains - then what S asserts is incorrect and hence is false. It follows classically, then, that what $\neg$S would assert in such a circumstance would be correct and hence would be true. In other words, analogously to s being complementary to $\bar{s}$, S is complementary to $\neg$S. To say that S is true or false depending upon whether s obtains or fails to obtain

---

[1]The divided collections of states-of-affairs are thus complementary in the standard sense; the complement of some state-of-affairs s is the state-of-affairs $\bar{s} =_{df} \{x: \neg(x \in s)\}$. This reexpresses the realist account of understanding: understanding a sentence S consists in knowing where S divides possible states-of-affairs.

is to say nothing other than that S is true or ¬S is true depending upon whether s or s̄ obtains. We can express the relationship as:

*) $(\forall S)(\exists s)((s \Rightarrow S) \wedge (\bar{s} \Rightarrow \neg S))$[1]

From (*), coupled with the recognition that for each possible state-of-affairs, either it or its complement obtains, it follows that:

LEM) $(\forall S)(S \vee \neg S)$

Furthermore, given that any adequate truth-definition must obey disquotation ("S" is true iff S), we can rewrite (*) as:

*′) $(\forall S)(\exists s)((s \Rightarrow$ 'S' is true) $\wedge$ ($\bar{s} \Rightarrow$ '¬S' is true))

from which, coupled with the same assumption about truth-conditions, it follows that:

*″) $(\forall S)$('S' is true or '¬S' is true)

which in turn, together with the classical identification of the truth of a negation with the falsity of its non-negated component, entails:

BV) $(\forall S)$('S' is true or 'S' is false)

In other words, as long as some basic assumptions about truth-conditions are observed[2], both the logical Law of Excluded Middle and the semantic Principle of Bivalence are validated. Those assumptions are the following: (i) a truth-condition either obtains or fails to obtain; (ii) a truth-condition s obtains if and only if its complement s̄ fails to

---

[1]The sentence form "p ⇒ P" should be read as "if the state-of-affairs p obtains, then P".

[2]In general, the truth-conditions of a sentence are identified with its associated state-of-affairs as originally given by "s ⇒ S" in (*). The anti-realist will conceive of truth-conditions as already being epistemically constrained. They will understand 's' in terms of the state of affairs of our verifying 'S'.

obtain, and conversely: and (iii) 'S' is true iff s (where s is S's truth-condition) and 'S' is false iff $\bar{s}$.

It is important to realize that LEM and BV are grounded in (i)-(iii), which make no (explicit) reference to the specifically realist thesis that truth may transcend provability: (iv) it is possible for a condition sufficient for the truth of some statement to obtain unrecognized. If, for example, all truth-conditions were surveyable (as they are, according to an anti-realist construal of truth-conditions), then the anti-realist would have no complaint against either LEM or BV. It is only if one understands 'true' in clause (iii) realistically that the anti-realist would have any complaint - but then, the anti-realist complaint is not against LEM or BV *per se*, but against importing a realist construal of truth *into* such principles (i.e. including (iv) in one's understanding of LEM and BV).[1]

Put in this way, the realist commitment to LEM and BV *per se* seem harmless; LEM and BV are only as good as their underlying truth-conditional counterparts. Contrary to first appearances, the anti-realist need not reject the first three assumptions - it is only the potential recognition-transcendence of both truth-value and truth-

[1]There is a tendancy in the literature to automatically understand LEM and BV realistically. Luntley (1988), for example, asserts that the realist thesis of the objectivity-of-truth (discussed in more detail in a footnote in §2.2.1.2) - i.e. that "contents have a determinate truth value independently of our being able to verify them" - just is "the principle of bivalence", which he understands as saying that "a content is determinately true or false independently of our ability to verify its truth value." (p. 30). See the next footnote for more detail on this point.

conditions that she must object to.[1]

---

[1]This point is important. In many passages Dummett notoriously makes it seem that acceptance of bivalence is the essence of what distinguishes the realist from the anti-realist. See Dummett (1959a) p. 14, (1959b) pp. 175-176, (1963b) p. 146, p. 155, (1969) pp. 358-359, (1973c) p. 228, (1976a) p. 275, (1976b) p. 93, (1978) p. xxii, p. xxix, (1982) p. 52, p. 60, p. 61, p. 60, and (1991b) p. 9-10, pp. 325-326. Dummett is being somewhat careless in these passages (indeed, he seems to reject the link between bivalence and realism in (1982) p. 69 and (1991b) pp. 304-305). As we shall see, his arguments against realism turn on the issue of the recognizability of truth-conditions, not the acceptability of bivalence. Griffin (1993) argues that the thesis of the recognition transcendence of truth and bivalence are not equivalent (in an earlier manuscript he presented persuasive evidence that many so-called realists do not accept bivalence and that many so-called anti-realists do accept it). Loar (1987) p. 87 points out: "There can be grounds for denying bivalence - e.g. vagueness - that have nothing to do with an anti-realism that asserts that truth requires verifiability." Tennant (1987) denies that anti-realism is essentially tied to the rejection of bivalence. C. Wright (1987) maintains that "Bivalence is merely the natural form for an acceptance of the possibility of recognition-transcendent truth ... [and it is] the status of such a conception of truth which Dummett's proposal, generalized, would make the crucial issue." (p. 4, see also C. Wright (1987) §2 and Currie (1993) §II). McDowell (1976) eloquently argues that an anti-realist can endorse a two-valued logic (and thus refrain from rejecting bivalence) without also being forced to endorse bivalence. Vision (1988) p. 181 nicely expresses the denial that realism itself is tantamount to bivalence: "the central issue dividing global realists and anti-realists is the proper account of sentences having whatever truth-values they do have, *not* whether every sentence has a (classical) one. It is the nature of truth, not its extension, that matters." McGinn (1982b) presents the clearest presentation of the tension. He distinguishes two 'senses' of 'realism' as used by Dummett: "Inspection of [various passages in Dummett's corpus] in search of the distinctive notions of truth adopted by realist and anti-realist turns up, on the face of it, two distinct properties: there is the property of being *epistemic*, and there is the property of being *determinate*. For truth to be epistemic is (roughly) for it to be applicable to a statement S only if S is in practice or in principle verifiable for us; and similarly for falsity. For truth to be determinate is (again roughly) for the statements to which it applies to be susceptible of a classical two-valued semantics, i.e. bivalence holds... But if these properties are indeed inequivalent, then it seems Dummett is tacitly operating with two notions of realism and anti-realism..." (p. 123).

It is my contention that a Dummettian semantic realist is one who takes a epistemically unconstrained notion of truth as central concept. Similarly, the semantic anti-realist denies that such a notion can legitimately serve as central concept (the negative programme) and asserts that only an epistemically constrained notion can do the job (the positive programme). The semantic anti-realist, then, can accept bivalence (in the sense that every sentence is guaranteed to be either true or false) as long as

If this is correct - i.e. that assumptions (i)-(iii) are independent of assumption (iv) - then it should be possible for one to accept the first three while rejecting the fourth. Let us briefly consider a counter-argument to this possibility. Consider the sentence:

> 5) It is 13,099,341°K at such and such a place and such and such a time in the interior of the sun.

Let s be the state-of-affairs such that, if it obtained, it would be 13,099,341°K at such and such a place and such and such a time in the interior of the sun. Given our human limitations of heat tolerance, we cannot, let us suppose, determine whether s obtains (nor, given the second assumption, can we determine whether $\bar{s}$ obtains). A combination of the first three assumptions tells us that exactly one of s or $\bar{s}$ obtains, and those assumptions commit us to the view that a truth-condition may obtain unrecognized (i.e. commits us to assumption (iv)).

It is premature at this point to consider anti-realist responses - they will emerge in subsequent sections. The point of this interlude is only to show that there are at least strong *prima facie* reasons for jointly accepting the underlying realist assumptions; i.e. that there is at least a strong presumption for accepting (iv) if one accepts (i)-(iii).

In other words, *if* LEM and BV can be validated by reference to assumptions (i)-(iii), and those assumptions predispose one towards (iv), and (iv) clearly commits one to a realist conception of truth, then the validations of LEM and BV are only relative - relative to a realist semantics which takes the notion of truth to be that as attaching to

---

'truth' and 'falsity' are not construed realistically. When Dummett declares bivalence to be unwarranted, we should take him as saying that there is no good reason to suppose that every sentence is either true (independently of verification) or false (independently of verification).

sentences independently of our capacities to recognize that their truth-conditions obtain when they do. It is for this reason, I believe, that Dummett claims that classical logic (in the form of commitment to LEM) and realist semantics (in the form of commitment to BV) stand and fall together:

> The validity of the law of excluded middle does not depend absolutely on the principle of bivalence: [but] once we have lost any reason to assume every statement to be either true or false, we have no reason, either, to maintain the law of excluded middle.[1]

If the above is correct, then there are two specific and separable components to a realist conception of truth. On the one hand, there is the Bivalence of Semantic Value. On the other hand, there is the Recognition-Transcendence of Truth (whether in the form of truth-value or truth-condition). We can characterize the semantic realist as the one who accepts both components in their notion of truth and the semantic anti-realist as one who at least repudiates the second. 'Truth', then, is ambiguous between these two senses. This would merely be an interesting bit of linguistic trivia *except that* the anti-realist offers arguments that the realist conception of truth is incoherent and that their own notion is the only acceptable one on the market. It is to these arguments which we must now turn.

2.2 The Critique of Semantic Realism

From the preceding four important theses have emerged:

> (A) a theory of meaning is a theory which attempts to explicate the meaning of any sentence in the language by reference to the central notion of correct

[1]Dummett (1991b) p. 9.

assertion: the meaning of a sentence is given by its conditions for correct assertion

(B) a theory of meaning is adequate only if it can be harmonized with an adequate theory of understanding

(C) an adequate theory of understanding characterizes sentential understanding in terms of a practical capacity to distinguish the conditions under which a sentence is correctly assertible from those which it is not

(D) the semantic realist conceives of correct assertion in such a way as to permit recognition-transcendent conditions for correct-assertion

The anti-realist critique of semantic realism is two pronged: it consists of a negative and a positive programme.[1] The negative programme is aimed at showing that (C) and (D) are mutually inconsistent. It further argues that as (D) is more suspect than (C), it must be rejected. Rejecting it is tantamount to rejecting any theory of meaning which takes recognition-transcendent truth as its central notion. The positive programme is aimed at showing that the only notion of correct assertion - i.e. *truth* - consistent with the constraints on an adequate theory of understanding is that offered by the anti-realist: truth is epistemically constrained.

2.2.1 The Negative Programme

The general strategy of Dummett's attack on semantic realism is pretty clear: a theory of meaning which takes a realist construal of truth as its central concept is incapable of harmonizing with any adequate theory of understanding. Such a theory, then, would be seriously in tension with the obvious fact that we understand our own language and are able to use it successfully in all kinds of ways. Dummett, however,

---

[1]See C. Wright (1987) and Appiah (1986) for a similar division.

presents at least two distinct species of the generic argument. The first, which C. Wright dubs The Acquisition Argument[1], aims at showing that we simply could not have come to acquire a conception of recognition-transcendent truth, and consequently, not having such a concept, it could not be the central one in any genuine theory of meaning. The second, which C. Wright dubs The Manifestation Argument, allows that we may possess such a concept, but is aimed at showing that it could not play the role accorded to it by the realist in a theory of meaning.[2]

---

[1]C. Wright (1987).

[2]Most of the commentators tend to concentrate exclusively on one of these two arguments. For example, McDowell (1976) focuses on the acquisition argument, while C. Wright (1987) tends to concentrate on the manifestation argument. I will consider the two separately, but I will agree with C. Wright that the former collapses into the latter.

C. Wright advances a third anti-realist argument to the effect that the central concept in an adequate theory of meaning need not go beyond an essentially epistemic one. He argues that "is true" and its cousin "is warranted" are both essentially normative predicates, but that necessarily the criteria for correctly predicating either of a sentence are the same ((1992) pp. 12-19). Thus he argues that the only requirement of the central concept in a theory of meaning be that it satisfy some basic normative criteria (which he calls a 'minimalist' conception of truth). A realist construal of truth goes beyond such minimum constraints, and hence is superfluous. He argues that 'superassertibility' (a sentence S is superassertible just in case "it is, or can be, warranted and some warrant for it would survive arbitrarily close scrutiny of its pedigree and arbitrarily extensive increments to or other forms of improvement of our information" (1992) p. 48)) satisfies such a minimalist constraint and is hence all we need in order to construct a theory of meaning. (See also (1987) pp. 295-302).

We can, however, mostly ignore C. Wright's argument. In the first place, Dummett's argument is that a realist construal of truth is an *impossible* one for humans while C. Wright's argument is merely that we do not need it for various purposes (namely, to construct an adequate theory of meaning). Secondly, arguments are raised in §4.3 that we do, contrary to C. Wright's claims, need the richer realist construal in order to construct an adequate theory of meaning.

2.2.1.1 The Acquisition Argument

Languages are learnable. Language learnability, like any other kind, is an epistemic notion: to learn a language L is to bring oneself into a position of knowing L. The knowledge involved is primarily of the implicit practical variety: to know L is to have the capacity to use L in meaningful and significant ways (the core use being, as suggested in §1, successful communication).

However, while knowledge of L may *ultimately* need to be given in terms of an implicit capacity, it need not *exhaustively* be so given. Consider the case of sentential understanding - there we saw that ultimately knowledge of the meaning of a sentence must be given in terms of a capacity to distinguish the conditions under which the sentence is correctly assertible from those under which it is not. We agreed, however, that such implicit knowledge can also be represented explicitly by a meaning-specifying theorem entailed by the correct meaning-theory for the language of which it is a sentence. Those meaning-specifying theorems were generated by appeal to a set of axioms specifying the semantic-values of the sub-sentential components as well as to a set of key concepts such as truth and satisfaction. Explicitly knowing the meaning of S, then, presupposes a knowledge of the axioms and the semantic concepts from which S's meaning-specifying theorem is derived. We can generalize this by saying that explicit knowledge of x consists (partially) in knowledge of the key concepts upon which x rests.[1]

[1]Must the conceptual knowledge be implicit or explicit? I'm inclined to think both - ultimately knowledge of the concepts upon which x rests must be cashed out in terms of a capacity, but there is no reason to think that a further *explicit* characterization of them in terms of other concepts cannot be given.

Let me summarize my general remarks about learning. To learn x is to bring oneself into a position of knowing x. Knowing x must *ultimately consist* in some sort of practical capacity, but can, at any given stage, be *represented* explicitly. An explicit representation would take the form of meaning-specifying theorems employing certain key concepts. Likewise, knowledge of the underlying key concepts may be represented in terms of explicit theoretical knowledge, but must ultimately be cashed out in terms of implicit practical knowledge. Take, for example, learning how to ride a bicycle. Learning how to ride a bike involves bringing oneself to a position of knowing how to ride it. That knowledge can be characterized implicitly - i.e. in terms of a capacity to actually ride the thing without falling over - or explicitly - partially in terms of possession of such concepts as peddle-force needed to achieve initial acceleration, the proper adjustment to steering for making turns, proper weight adjustment at arbitrary velocities to maintain balance, etc. However, mere possession of such concepts will not suffice for knowing how to ride a bicycle; possession of such concepts will only so suffice if knowledge of them can ultimately be demonstrated by a practical ability. For example, knowledge of the peddle-force needed to achieve initial acceleration must, if being invoked as an explicit statement of (part of) the knowledge needed to ride a bicycle, consist ultimately in being able to push the peddle hard enough to start moving: if one cannot actually push the peddle hard enough to start the bike moving, then - even if one can accurately state the physical formula governing the action, give the initial values, and calculate the force - one does not know how to ride a bicycle.

Coming back to the original case, there is no harm in characterizing knowledge

of a language L in terms of an explicit knowledge of the axioms of its correct meaning-theory as well as that theory's key concepts *as long* as that explicit knowledge is ultimately grounded in possession of a practical capacity. Contrapositively, if that explicit knowledge *is not* or *cannot* be so grounded (or derived), then it cannot even serve as a theoretical representation of that knowledge. Again, the analogy can help explicate this point.

Consider someone who, as an infant, developed polio to such an extent that they lost all use of their legs. Such a person would, at best, only be able to manifest an explicit knowledge of the key concepts involved in knowing how to ride a bicycle; that explicit knowledge, due to their physical limitations, could not ultimately be manifested or grounded in a practical capacity. The knowledge which they do possess, then, would not be sufficient to warrant an attribution of the knowledge of how to ride a bicycle. Moreover, not only *do* they *not* have the requisite knowledge, such a person *could not* have the requisite knowledge: their physical condition is such as to preclude such knowledge. Why? Precisely because their physical condition is such as to preclude the *possibility* of their acquiring a knowledge of a range of concepts which ultimately must be *grounded* in possession of a practical capacity.

Dummett presents his clearest case for the acquisition argument in "The Reality of the Past"[1]: If a realist construal of truth can be the central concept in an adequate theory of meaning, then it must be possible for us to possess that concept. But, all linguistic concepts must ultimately be grounded in use - i.e. knowledge of them must

---

[1]Dummett (1969).

ultimately be explained in terms of the practical capacities we actually have (or have come to acquire). Acquisition of the concept of recognition-transcendent truth cannot, Dummett claims, be derived from the practical capacities we actually employ in acquiring linguistic concepts. In a nutshell, Dummett claims that we could not have acquired the concept of recognition-transcendent truth and hence it can play no role in a theory of meaning:

> [The anti-realist] maintains that the process by which we came to grasp the sense of statements of the disputed class, and the use which is subsequently made of these statements, are such that we could not derive from it any notion of what it would be for such a statement to be true independently of the sort of thing we have learned to recognize as establishing the truth of such statements...
>
> In the very nature of the case, we could not possibly have come to understand what it would be for the statement to be true independently of that which we have learned to treat as establishing its truth: there simply was no means by which we could be shown this.[1]

The argument rests upon the claim that humans cannot acquire a conception of a recognition-transcendent truth-condition. Dummett's general support for this premise relies upon a broadly empiricist account of learnability; one learns a linguistic concept by experiencing correct uses of it. This in turn requires that the correct use of a linguistic concept be limited to those cases which one can experientially recognise. For example:

---

[1]Dummett (1969) p. 362. He also hints at the argument at (1963a) p. 188: "Teaching a child language is not like teaching a code. One can put a code-symbol and that for which it is a symbol side by side, but one cannot isolate the concept in order to teach the child which word to associate with that concept. All that we can do is to *use* sentences containing the word, and to train the child to imitate that use." See also (1976b) p. 318: the realist faces the problem of "how to account for our acquisition of that grasp of conditions for a transcendent truth-value which he ascribes to us, and to make plausible that description."

> We learn the use of the past tense by learning to recognise certain situations as justifying the assertion of certain statements expressed by means of that tense. These situations of course include those in which we remember the occurrence of some event which we witnessed, and our initial training in the use of the past tense consists in learning to use past-tense statements as the expression of such memories... The only notion of truth for past-tense statements which we could have acquired from our training in their use is that which coincides with the justifiability of assertions of such statements, i.e., with the existence of situations which we are capable of recognising as obtaining and which justify such assertions.[1]

Thus, the only notion of truth as applied to past-tense statements which one could have learned is one which can be gleamed from observing others correctly use it, and that, according to Dummett, requires that its correct use be limited only to experientially accessible (i.e. non-recognition-transcendent) instances.

Dummett is correct here; *if* we accept the broadly empiricist account of learnability, it is difficult to see how a realist construal of truth could have been acquired. As C. Wright says:

> Obviously such a conception cannot be bestowed ostensively. And the challenge is simply declined if the answer is offered 'by description'. For it is of our ability to form an understanding of precisely such a description that an account is being demanded; there could be no better description of the relevant kind of state of affairs than the very statement in question.[2]

What C. Wright is alluding to is that a concept of which only knowledge by description is possible is a concept of which knowledge cannot ultimately be manifested implicitly - i.e. knowledge which we agree could not suffice for attributing linguistic understanding to someone. As long as knowledge by description and knowledge by acquaintance

---

[1]Dummett (1969) p. 363.

[2]C. Wright (1986) p. 13.

exhaust conceptual knowledge, knowledge of any recognition-transcendent concept would be insufficient for attributions of linguistic understanding.[1]

Dummett's basic empiricist account of learnability is highly suspect on a number of counts. McGinn (1982a) labels Dummett's view the *dispositional account of content-ascription* since it requires that the content of a concept be determined by a speaker's disposition to assent to or dissent from certain sentences under appropriate assertibility-conditions - and argues that it is simply mistaken. Consider a Putnamian Twin-Earth scenario:[2] Earthers and Twin-Earthers may be under the same assertibility-conditions when confronted with a sentence containing the term 'water', yet because, by assumption, the substance referred to as 'water' on Earth is distinct from that substance referred to as 'water' on Twin Earth, Earthers and Twin-Earthers do not share the same concept:

> Speakers on earth and twin earth thus acquire the same recognitional capacities and manifest them in the same conditions of evidence, but their sentences do not mean the same... I think that this case already shows that there is something wrong with Dummett's conception of content-ascription: for here we have a case in which agreement in use does not entail agreement in content.[3]

---

[1]Dummett (1969) considers, though ultimately rejects, the realist rejoinder that an recognition-transcendent concept of truth as applied to past-tense statements could be generated by knowledge of truth-value links; e.g. by understanding the link between "'S is true' at t_n" and "'S was true' at $t_{m>n}$".

[2]To be discussed in much more detail in the next chapter.

[3]McGinn (1982a) p. 116. He presents two other structurally similar purported counter-examples. Dummett actually presents a similar argument in (1963a). Gödel's Theorem asserts that there exists some sentence U expressible in some intuitively correct formal system for elementary arithmetic which is true but not provable. This in turn is generally taken to show that no consistent formal system in which mathematics is expressible can be complete. Thus, each individual must , in making sense of mathematical statements, make an implicit appeal to some structure or model which is not formalizable. Not being formalizable, there is no guarantee that we all attach the same meaning to our mathematical statements, *even though* there may be no difference

McGinn's argument may be a bit quick. He assumes that any sentence containing the expression 'water' as used on Earth *must* semantically differ from its counterpart as used on Twin-Earth. He also assumes that Earthers and Twin-Earthers may be in the same assertibility-conditions when confronted with such sentences. The implication of those two assumptions is that assertibility-conditions are insufficient to fix the meaning of a sentence - but that is just to deny semantic anti-realism. Dummett's proper reply, then, should be to resist McGinn's (and Putnam's) assumption that Earth and Twin-Earth counterpart sentences are indeed semantically distinct. The typical reason given for their distinctness is because the substances respectively referred to by 'water' differ, but Dummett's view is that the meaning of an expression is determined solely by its *use*. Earthers and Twin-Earthers do not use the sentences differently, and thus there is no good Dummettian reason for regarding them as semantically distinct.

Be that as it may, Dummett's broadly empiricist account of concept acquisition is dubious on other grounds. There seem no good reasons, except certain outmoded empiricist dogmas, to suppose that conceptual knowledge is exhausted by either acquaintance or description. Consider the concept of the temperature of such and such a location at such and such a time in the interior of the sun. Clearly, given our human limitations, that concept cannot be learned ostensively. However, it is possible to learn the concepts of temperature, of particular locations on an arbitrary spatial grid system, of particular locations on an arbitrary temporal grid system, of spatial relations like interior and exterior, and of the sun - perhaps even ostensively. There seems to be no

---

in our use of such sentences.

reason to suppose that a competent concept user cannot combine these into the complex concept mentioned.[1] In this vein, all that would be required to learn the concept of a recognition-transcendent truth-condition would be the component concepts of a recognizable truth-condition and negation.[2]

In the same vein, McGinn (1976) argues that it is reasonable to assume that such a notion is in fact acquirable. Imagine a community of speakers just like us save that they lack the capacity for locomotion - they are rooted tree-like on the north side of a mountain. They are able to observe all (or at least most) of the goings-on on the north side, but are unable to observe any events on the south side. Suppose, also, that sheep routinely appear on the north side *as if* they had come from the south side, and seem to disappear into the south side from the north side. Given their recognitional capacities, the speakers would have no problem acquiring a realist conception of truth as applied to sentences expressing what sheep are doing on the north side. Dummett's acquisition argument concludes that they could not have acquired a realist conception of truth as applied to sentences expressing what sheep are doing on the south side. However, McGinn argues that members of the community will have formed a 'picture of reality', or a theory, which includes the existence of and events on the south side of the

---

[1]Even the arch-empiricist Locke allows formulation of novel complex ideas from a combination of simple ideas. The archer-empiricist Hume, however, may resist this - but only at the cost of imposing a highly dubious strong skeptical claim regarding induction.

[2]To this the anti-realist may argue that the concept of negation appealed to is the classical truth-functional one. That connective, they will argue, presupposes a realist concept of truth which cannot, without begging the question, be used to support its intelligibility; the intelligibility of classical negation is exactly as secure as the intelligibility of realist truth.

mountain. Such a theory is necessitated, he argues, in order for them to explain observed phenomena, like sheep disappearing and reappearing, etc. In other words, McGinn thinks that there is no reason, other than "empiricist dogma", for denying "the possibility of acquiring conceptions of reality that transcend our recognitional capacities."[1] His point is that our 'total picture of reality' transcends that portion of reality that we observe, and that we furthermore need (i.e. must have acquired) that 'total picture' just in order to explain observed phenomena.[2]

What seems clearly at issue is *not* transcendental arguments for or against the possibility of acquiring certain concepts. Rather, it is whether or not knowledge of certain concepts can be of the appropriate sort - i.e. whether such knowledge ultimately rests upon a *manifestable* capacity. It is no objection to point out that knowledge of the concept of a recognition-transcendent truth-condition can be given by a (question-begging) description. The only relevant objection must centre on the claim that such knowledge must be *exhausted* by description. The argument from acquisition, therefore, is parasitic upon the argument from manifestation (to be taken up in the next section). If knowledge of recognition-transcendent truth-conditions *can* be manifested, then it *can* be acquired - and any question of how it can be acquired would be essentially moot. As

---

[1]McGinn (1976) p. 29.

[2]See also McGinn's (1981) reply to Tennant's (1981) criticism and Tennant's (1984) response. On an aside, Luntley's (1988) anti-realism allows for what he calls an 'objectivity of content' (we do possess a conception of truth as potentially recognitionally transcendent) but rejects what he calls an 'objectivity of truth' (that "the contents we grasp are contents that have a determinate truth value independently of our knowledge of that value". (p. 4)).

C. Wright observes:

> For *if* it could be clear that we did indeed possess a realist understanding of certain statements, the question, how that understanding had been acquired, while no doubt of some independent interest, would cease to be of any importance in the - then defunct - issue between realism and its opponents.[1]

The only argument against the possibility of a realist construal of truth being acquirable, then, must be that it is not manifestable: the acquisition argument collapses into the manifestation argument.

Before examining the argument from manifestability, I want to present a preliminary case for supposing that we *have*, indeed *must have*, acquired a concept of recognition-transcendent truth. As mentioned, the argument from acquisition seems to presuppose that conceptual knowledge is limited to either acquaintance or description. Such a limitation would entail that there can be no intelligible concepts which go beyond possible experience. There is strong reason to deny this claim. Consider the anti-realist account of *sentential* understanding: understanding a sentence S must ultimately consist in a practical capacity to distinguish between the conditions which warrant its assertion and those which do not. Presumably they must offer an analogous account of *conceptual* understanding - understanding a concept must ultimately consist in a practical capacity to distinguish between those objects which are satisfied by it and those which are not.

Consider the general concept of a recognizable truth-condition. Understanding that concept must ultimately consist in a capacity to distinguish between those objects which are satisfied by it and those which are not. The only candidates for objects which

---

[1]C. Wright (1987) p. 86.

are satisfied by the concept are recognizable states-of-affairs, while everything else (in particular, unrecognizable states-of-affairs) will be assigned to the category of objects not satisfied by the concept. Thus, understanding the general concept of a recognizable truth-condition requires a capacity to at least conceptually differentiate between recognizable states-of-affairs and unrecognizable states of affairs. To even have the concept of a recognizable state-of-affairs (which Dummett is committed to, as he holds that concept to be central in a theory of meaning), one must also have the concept of a non-recognizable state-of-affairs. That concept should be sufficient to generate the concept of recognition-transcendent truth. Thus, the realist's inability (or anyone else's, for that matter) to recognize unrecognizable states-of-affairs cannot in any way be taken as evidence of the unintelligibility of that concept. Dummett is simply too quick to assign the concept of realist truth to those "errors of thought to which the human mind seems naturally prone".[1]

The upshot is that the anti-realist critique depends upon whether or not - and in particular *how* - such conceptual knowledge can be manifested. It is to the issue of manifestability that we must now turn.

2.2.1.2 The Manifestation Argument

Dummett concisely expresses the argument from manifestation:

> In fact, whenever the condition for the truth of a sentence is one that we have no way of bringing ourselves to recognize as obtaining whenever it obtains, it seems plain that there is no content to an ascription of an *implicit* knowledge of

[1]Dummett (1969) p. 374.

> what that condition is, since there is no practical ability by means of which such knowledge may be manifested. An ascription of the knowledge of such a condition can only be construed as *explicit* knowledge, consisting in a capacity to *state* the condition in some non-circular manner; and that, as we have seen, is of no use to us here.[1]

More explicitly, Dummett has argued that any adequate theory of meaning must harmonize with an adequate theory of understanding understood as a theory correlating a speaker's linguistic knowledge with capacities to overtly manifest correct use of their language. Secondly, he has argued that grasp of a sentence's meaning consists in a grasp of the conditions under which an assertion of that sentence would be correct and the conditions under which it would not be correct. It follows from his first point, then, that grasp of a sentence's meaning requires one to overtly manifest a capacity to distinguish between those conditions under which its assertion would be correct and those under which it would not be correct.[2] According to the realist, truth-conditions (or, conditions under which the assertion of a sentence would be correct) may transcend recognition; i.e. there are genuine sentences - call them of type U - whose truth or falsity is beyond our determination. It would seem to follow, then, that one could not manifest a capacity to distinguish between conditions under which an assertion of such sentences would be correct and conditions under which it would not, and thus no one could be said to

---

[1]Dummett (1976b) p. 82. Versions of the argument appear in almost every one of Dummett's writings.

[2]One may concede that knowledge of the meaning of a sentence may be manifested in many different ways, just as I can manifest my belief that a particular substance is poisonous by avoiding it, but also by taking "steps to ensure my family avoid it, or take steps to ensure that they don't!" (C. Wright (1986) p. 33). All varieties of manifestation must, however, according to Dummett, be such as to provide evidence of a capacity to distinguish conditions which justify assertion of a sentence from those which do not.

understand such sentences. As no one could understand sentences of type U there is no reason to regard them as genuine sentences - truth-conditions, as a matter of fact, are limited to the recognizable. As Dummett says, "[a] truth-conditional meaning-theory [i.e. semantic realism] violates the requirement that meaning be correlated with speaker's knowledge."[1]

The crucial premise is, of course, the existential one regarding sentences of class U, and it requires a fuller explication of the class U. In a nutshell, U consists of the class of sentences whose truth-conditions are unrecognizable. Consider this situation: a single die is placed in an opaque box and then sealed. The box is shaken and the pronouncement:

6) Right now the upper face of the die in the box shows six pips.

is made, and then the box is opened to reveal the die with six pips showing on its upper face. The realist will say that the utterance's truth-conditions consisted in the state of the die at the time the utterance was made. In this case (6)'s truth-conditions are recognizable as is revealed by the fact that they are recognized as obtaining at the time the box is opened. However, suppose that instead of the box being opened, it was immediately shaken thereby disturbing the condition of the die. The realist will say that it makes no difference *viz-a-viz* (6)'s truth-conditions whether the box was opened or shaken - all that its truth-condition depends on is the state of the die *at the time the*

---

[1]Dummett (1991b) p. 306. Tennant (1984) and (1987) expresses the manifestation argument in this way: MANIFESTATION (grasp of meaning must be fully manifestable) + REALISM (classical, bivalent truth is the central concept in a theory of meaning) + FACT (we understand undecidable sentences) forms an inconsistent triad.

utterance was made. What happens to the die *after* the utterance is made is irrelevant to its truth-conditions.

Notice, however, what happens to the truth-conditions: they change from being recognized to being unrecognized. Being recognized, however, is not the same as being recognizable. The anti-realist will say that (6)'s truth-conditions under either scenario are recognizable ones; in the first case they are recognized *and* recognizable, while in the second they are unrecognized *but* recognizable. In what sense are they unrecognized but recognizable? In the sense that we have a capacity to determine (6)'s truth-*value* (by determining whether or not its truth-*conditions* obtain) at the time the utterance was made. In the second case that capacity was not exercised - but as long as we recognize that capacities are dispositional in nature this need not bother us.

So, (6) fails to qualify for membership in U. Membership in U requires that a sentence's truth-conditions be unrecogniz*able* - i.e. be such that a determination of their truth-value would require capacities going beyond the (humanly) possible. Now, what is the relationship between the unrecognizable and the recognition-transcendent? Recall that the realist wants to say that all sentences admit of (potentially) recognitionally-transcendent truth-conditions, but clearly no realist would hold that all sentences admit of unrecognizable truth-conditions.[1] Realists would happily accept (6) as having recognizable truth-conditions under at least the first scenario. They do want to say, however, that nonetheless they are (potentially) recognitionally-transcendent.

---

[1]Except, of course, the global sceptic. However, the realist, while perhaps sympathetic, need not be a sceptic.

At the heart of the realist position is the notion that the truth or falsity of a sentence has nothing whatsoever to do with epistemic facts about humans.[1] Consider sentence (6) - its truth-conditions were said to be recognizable even though in the second scenario they were unrecognized. Suppose that upon opening the box six pips were clearly visible on the top face of the die. In that case it would have been determined that (6) had the value *true*. It is tempting to think that the anti-realist is committed to the view that it has the value *true* in virtue of its truth-condition *being recognized* as obtaining. However, in the second scenario its truth-condition is not recognized as either obtaining or failing to obtain, yet nonetheless the anti-realist need not deny that, *ex hypothesi*, it has the value *true*. Why? Precisely because we do have a capacity to determine whether (6)'s truth-conditions obtain or not. This is only to repeat, however, that according to the anti-realist, a sentence's having a truth-value depends upon whether or not we have a capacity to determine its truth-value where that capacity need not actually be exercised. So, on the anti-realist account, a sentence's truth-value depends upon epistemic facts about humans - i.e. facts about what sorts of evidential capacities we do or do not possess.

Consider a third scenario. Shortly before the box is shaken all humans lose their capacity for visual and tactile sensations. Under such a plight (6)'s truth-conditions would cease to be recognizable. Now if, as the anti-realist maintains, a sentence has only recognizable truth-conditions, then under this scenario (6) would cease to have truth-

---

[1]Unless, of course, the sentence in question concerns epistemic facts about humans.

conditions at all.[1] If it lacks truth-conditions, it lacks a truth-value.

We could, I suppose, preserve bivalence by a stipulation to the effect that a sentence is true just in case its truth-conditions recognizably obtain and false otherwise. In other words, we might want to contemplate two sorts of conditions for falsehood: a sentence is false just in case either its truth-conditions recognizably fail to obtain *or* its truth-conditions fail to recognizably obtain.[2]

This stipulation, while preserving bivalence, would preclude (6) being assigned the same truth-value under either of the three scenarios. Whereas under the first proposal (correlating the falsity of a sentence with its falsity-conditions recognizably obtaining) (6) would not be determined as false under any of the scenarios, under the second (correlating the falsity of a sentence with its truth-conditions not recognizably obtaining) (6) would be determined as true under the first scenario but false under either the second or third. From an anti-realist perspective, this is as it should be: the truth-value of a sentence is parasitic upon epistemic facts about humans. Alter the facts, as we did in the third scenario, and you *ipso facto* alter its meaning.

The realist, on the other hand, will maintain that (6) retains the same truth-value

---

[1]Either that or they would shift to something else - i.e. (6)'s conditions for truth would be something other than it being the case that the die visibly or tactually shows six pips on its top face. However, in such an event we can retain recognizable truth-conditions for (6) only at the cost of adopting a non-standard set of truth-conditions and hence we will have shifted (6)'s meaning. In other words, we could not regard this as a third scenario *regarding the same sentence.*

[2]Dummett hints at this sort of manoeuvre in (1976b) p 12: "Thus, on *one* way of using the words 'true' and 'false' ... instead of distinguishing between the singular statement's being false and its being neither true nor false, we should have distinguished between two different ways in which it could be false."

across the three scenarios, even though its truth-conditions become unrecognizable under the third. Facts about the recognizability of particular truth-conditions are, from the realist perspective, extrinsic to the relationship between those truth-conditions and their attendant truth-values; altering the epistemic facts will not suffice to alter the alethic facts. In other words, the truth-conditions of a sentence transcend, in the sense of being completely independent of, epistemic facts concerning their recognizability. Truth-conditions, and hence *truth* in general, are, according to the realist perspective, recognition-transcendent in precisely this sense. Thus, there is no prima facie paradox in supposing there to be particular truth-conditions which are both recognizable and recognition-transcendent.[1]

One must be careful to keep the notions of 'recognition-transcendence' and 'unrecognizability' distinct. A sentence has unrecognizable truth-conditions just in case we are unable to determine whether they obtain. A sentence has recognition-transcendent truth-conditions just in case their obtaining (or failing to obtain) is independent of our capacity (or potential lack thereof) to determine whether they obtain. It is the claim that sentences have recognition-transcendent truth-conditions which

---

[1]McGinn (1976) concurs - undecidability is not a pre-condition for recognition-transcendence (pp. 22-23). He notes, then, that if the anti-realist argument revolves around recognition-transcendence *per se*, it should be just as applicable to decidable sentences (p. 23). It seems not unreasonable to infer, from McGinn's observations, that as the anti-realist issues the manifestation argument only in terms of undecidable sentences, it is aimed not at a recognition-transcendent notion of truth in our sense, but rather at whether that notion can reasonably be extended to undecidable sentences. But *if* one can retain a notion of truth as recognitionally transcendent in our sense, then a realist ontology should remain untouched *even if* semantic realism is ultimately untenable.

constitutes realism; conversely, it is the denial that sentences have recognition-transcendent truth-conditions which constitutes anti-realism.

Nevertheless, the characterization of 'unrecognizability' expresses something of importance to the manifestation argument - namely the criteria for membership of class U. The argument from manifestation is that the existence of such sentences is enough to cast serious doubt on semantic realism. Recall the premise of the manifestation argument that knowledge of the truth-conditions of a sentence S is at least part of the knowledge of the meaning of S. As we have seen, that knowledge must ultimately consist in a manifestable capacity. As such, correct ascriptions of that knowledge depend upon (the potential for) a favourable outcome in a testing situation. The testing procedure of Dummett's consists in determining if one is capable of manifesting a knowledge of S's truth-conditions. Now, S's truth-conditions are stated by its associated T-sentence "'S' is true iff P". Thus, we can attribute knowledge of S's truth-conditions to anyone who is capable of displaying a knowledge of its associated T-sentence. We can correctly attribute that knowledge to anyone who, when situated favourably to investigate the state-of-affairs referred to in the right-hand side assents to the sentence (if true) mentioned on the left-hand side (or who dissents from it, if it is false). Thus, there is one obvious condition under which ascription of a grasp of S's meaning would be incorrect; namely if they dissent from (or assent to) S when its truth-conditions obtain (or fail to obtain). For example, if, when situated in a field of snow, someone dissents from 'Snow is white', they clearly fail to grasp its meaning. However, there is a second condition under which such an ascription would be mistaken; namely if it is not possible

to so situate someone. If the only admissible evidence for ascribing knowledge of sentential meaning to someone is success in a testing situation, they either failure to pass the test or failure to take the test counts as sufficient reason for withholding such ascriptions:

> ...a grasp of the condition under which the sentence is true may be said to be manifested by a mastery of the decision procedure, for the individual may, by that means, get himself into a position in which he can recognize that the condition for the truth of the sentence obtains or does not obtain, and we may reasonably suppose that, in this position, he displays by his linguistic behaviour that the sentence is, respectively, true or false.[1]

Recall the bicycle analogy. In that case we can distinguish between two grounds for withholding ascriptions of knowledge of how to ride a bike. On the one hand, if someone attempts to ride and immediately falls off, then we have sufficient reason for withholding that ascription. On the other hand, if someone is not capable of taking the test - e.g. our polio victim - then we similarly have sufficient reason for such a withholding. Quite simply, if one cannot possibly take a test whose success constitutes sole grounds for attributing knowledge of x, then they cannot manifest their knowledge of x. And if they cannot manifest their knowledge of x, then they do not have the knowledge of x.

The problem for the realist is that it is a precondition of such a testing situation that the state-of-affairs constituting a sentence's truth-conditions be ones that the testee is capable of recognizing. Members of U, by definition, have *unrecognizable* truth-conditions. Thus, even if one were suitably placed *viz-a-viz* a state-of-affairs whose

---

[1]Dummett (1973c) pp. 224-225.

obtaining would be sufficient for S's truth, one would not be aware *that* they obtain and hence could not assent to the sentence on the basis of that recognition (or, alternatively, they could not be aware *that* they fail to obtain and hence could not dissent from it on that basis). Members of U have truth-conditions which determine testing situations no one is capable of satisfying.

Thus, if members of U have truth-conditions which are unrecognizable to humans as a class, then humans cannot manifest their knowledge of their truth-conditions. Given that manifesting a knowledge of a sentence's truth-conditions is (at least part of) what it is to manifest a knowledge of its meaning, it would seem that humans are incapable of manifesting an understanding of the members of U. But, if there are compelling reasons for supposing that humans *do* understand members of U, then they would constitute compelling reasons for holding that knowledge of their realist truth-conditions can be no part of what it is to understand them, and subsequently such truth-conditions can be no part of their meanings. This could furthermore be generalized: realist truth-conditions can be no part of the meanings of sentences of any class.

The generalization is a bit suspect - it appears to equate recognition-transcendent truth-conditions with unrecognizable truth-conditions. As we have seen, that identification is unwarranted. If it turned out that all sentences have recognition-transcendent truth-conditions (as the realist maintains) but that none have unrecognizable truth-conditions (i.e. U is empty), then the manifestation argument would attack a straw person. Thus, the argument cannot be aimed at the recognition-transcendence of truth *per se*. It only gains currency if the proponent of the recognition-

transcendence of truth is *committed* to the existence of sentences with unrecognizable truth-conditions.

Thus, the success of the manifestation argument rests upon admissibility of two key assumptions. First of all, it assumes that the only acceptable means of manifesting sentential understanding is the capacity to succeed in a described testing situation. Secondly, it assumes that U is non-empty - or at least that the realist is committed to it so being. If either of these assumptions fails, then so too does the manifestation argument. In §3.1.2.2 and §4.2 we will consider serious objections to both assumptions.

2.2.2 The Positive Programme

The negative programme was aimed at establishing an inconsistency between the thesis that a correct theory of meaning takes recognition-transcendent truth-conditions as its central concept and the thesis that an adequate theory of meaning must harmonize with an adequate account of understanding. If the two theses are inconsistent, and the latter is unassailable, then semantic realism is in serious difficulty. The positive programme is aimed at establishing the precise form that a theory of meaning must take - in particular what its central notions must be - in order that it harmonize with such an account of understanding. It does so by arguing that only a theory of meaning which takes an epistemically constrained notion of truth as central is able to achieve the blend.

Through the acquisition and manifestation arguments, we are aware of a number of features which an adequate theory of meaning must have. In §1 it was argued that a theory of meaning must take the notion of correct assertion as its central concept.

Thus, in order to achieve the desired blend with the theory of understanding, the specific form of that concept must be such that it be acquirable and overtly manifestable.

The problem with semantic realism (claims the anti-realist) is that it conceives of conditions for correct assertion so as to allow for some of them being unrecognizable. If a truth-condition is unrecognizable, then it is unclear how one could either acquire an idea of it or, if one could, how one could manifest that knowledge. Thus, conditions for correct assertion, in order to avoid the difficulty, must be conceived as non-unrecognizable.[1]

Recognizability is an epistemic property in that whether or not a set of conditions is recognizable (for humans) depends upon epistemic facts (about humans). By altering those epistemic facts - e.g. by altering the evidence-gathering capacities of humans - one *ipso facto* alters the range of recognizable conditions. Thus, a theory of meaning must take this constraint into account by closely tying the very notion of correct assertion to the actual epistemic capacities of humans.

The anti-realist thus conceives of conditions for correct assertion in terms of *verification*-conditions. The meaning of a sentence, then, is given by or consists in the

---

[1]The double negation is interesting. Both the acquisition and manifestation arguments involve a *reductio* - both are aimed at deriving the contradiction that understanding of certain sentences is both actual and impossible for humans. Now, assuming intuitionism (the preferred logic of the anti-realist), only a negated conclusion is warranted from an indirect proof. Thus, only the conclusion that conditions for correct assertion must be non-unrecognizable is warranted - in other words, the stronger thesis that such conditions must be recognizable is *not* intuitionistically warranted by either argument. In the following I will ignore this problem and assume the arguments demonstrate that truth-conditions, if they are to play any role in a theory of meaning, must be recognizable.

conditions under which a sentence would be verified as opposed to those under which it would be true. Now verification, like recognition, is an epistemic notion. A sentence is verified by certain conditions only if we have knowledge of those conditions obtaining. Similarly, verifiable stands to verified exactly as recognizable stands to recognized: a sentence is verifiable just in case we have a capacity to gather evidence sufficient for its correct assertion. As the verifiable is determined by the range of our evidence-gathering capacities, it follows that verification-conditions cannot be recognition-transcendent in the sense established in §2.2.1.2. Moreover, they must be recognizable conditions - verifiability requires that it be possible to gather supporting evidence, which in turn requires a capacity to recognize the obtaining (or otherwise) of certain conditions *as* evidence.

In terms of understanding, the anti-realist takes the capacity to distinguish between conditions under which a sentence would be verified and conditions under which it would not as constituting sentential understanding. One who was capable of manifesting such a capacity would be one who understood the sentence. In the context of the testing procedure, one who assented to (6) while placed in the recognizable state-of-affairs of the die showing six pips on its upper face would be one who manifested their understanding, while one who dissented from the same sentence under the same conditions would be one who failed to manifest their understanding and hence failed to grasp its meaning. As we saw, it is a presupposition of the testing procedure that the conditions of the test be recognizable ones. Verification-conditions satisfy this constraint.

We must be careful, however, to distinguish between a sentence's being unverified and its being *falsified.* A sentence would be falsified just in case we have sufficient evidence that the conditions for its correct assertion failed to obtain. Similarly, a sentence is falsifiable just in case we have a capacity to gather evidence sufficient for its correct denial.

In any event, there are four distinct epistemic properties a sentence might have. It might be either (i) verified; (ii) unverified; (iii) falsified; or (iv) unfalsified. However, consider:

7) Right now the upper face of the die in the box shows five pips.

Sentence (6) is verified in the first scenario and unverified in the second. Sentence (7) is falsified in the first scenario and unfalsified in the second. Moreover, sentences (6) and (7) are both unverified and unfalsified in the second scenario even though they are verifiable and falsifiable respectively in that situation.

What about the cognate notions of unverifiability and unfalsifiability? In the first scenario, sentence (6) is unfalsifiable *given* that, because its conditions for correct assertion recognizably obtain, we cannot have (i.e. manifest) a capacity to recognize that they fail to obtain. Similarly, sentence (7) is unverifiable in the same situation *given* that, because its conditions for correct assertion recognizably fail to obtain, we cannot have (i.e. manifest) a capacity to recognize that they do obtain. Consider the second scenario. As we have a unexercised capacity to recognize that the conditions for (6)'s correct assertion obtain, we cannot manifest a capacity to recognize that they fail to, and hence (6) remains unfalsifiable. Similarly with (7)'s being unverifiable.

The point is that verification and falsification are unlike our normal notions of truth and falsity. In our normal (i.e. classical) inferential practices, truth and falsity commute with negation:

i) S is true ≡ S is not false
ii) S is false ≡ S is not true

However, their epistemic cousins do not:

iii) S is verified ≢ S is not falsified
iv) S is falsified ≢ S is not verified
v) S is verifiable ≢ S is not falsifiable
vi) S is falsifiable ≢ S is not verifiable

(iii) and (iv) are both established by sentence (6) in the second scenario. (v) and (vi) are both established as long as there are sentences which are *effectively undecidable*: i.e. sentences for which we do not have (i.e. cannot manifest) a capacity to determine whether either their verification-conditions obtain or their falsification-conditions obtain. For present purposes, sentence (5) (expressing the exact temperature of some point on the sun) will suffice for illustration.

With these distinctions in mind, we can see that there are at least two distinct anti-realist theories of meaning:

MT_1) The meaning of a sentence S consists in the conditions under which it would be verified.

MT_2) The meaning of a sentence S consists in the conditions under which it would be falsified.[1]

---

[1]Price (1983) follows up a suggestion (Dummett (1976b) p. 112 and pp. 117-118) that ın adequate meaning-theory can be formed which takes the conjunction of verification ınd falsification as its central concept. The main divergence over such theories as MT_1 ır MT_2 will be that the hybrid theory will contain the clauses "'not-S' is assertible iff 'S' ; deniable" and "'not-S' is deniable iff 'S' is assertible". (p. 167). For our purposes, it

and hence two distinct anti-realist accounts of understanding:

> UT_1: Understanding a sentence S consists in being able to distinguish between the conditions under which S would be verified and the conditions under which it would not.
>
> UT_2: Understanding a sentence S consists in being able to distinguish between the conditions under which S would be falsified and the conditions under which it would not.[1]

2.2.2.1 Logical Concerns

The classical interpretations of the logical connectives are given truth-functionally. The truth-table for negation identifies the truth of any given sentence with the falsity of its negation. Consider the negation of sentence (5):

> 8) It is not the case that it is 13,099,341°K at such and such a place and such and such a time in the interior of the sun.

According to the classical truth-table for negation, (8) is true just in case (5) is false (and *vice versa*). Now, given the identification of a sentence's truth-value with whether or not its truth-conditions obtain, it follows that (8) is true just in case the truth-conditions for (5) fail to obtain. It was agreed, however, that humans do not have the capacity to recognize whether the conditions for (5)'s truth obtain, and hence (5)'s truth-conditions must, it seems, be deemed unrecognizable. But, if (5)'s truth-conditions are

---

makes little difference what specific form an anti-realist meaning-theory takes, as long as its central concept is a non-recognition-transcendent one. In general, however, by 'anti-realist meaning-theory' I will tend to mean a theory along the lines of MT_1. See also Prawitz (1987) for a fruitful discussion.

[1]Dummett's general view is along the lines of MT_1 and UT_1, but in (1976b) §V he considers a view along the lines of MT_2 and UT_2. At this point we do not need to assume one over the other as expressing the general anti-realist position.

unrecognizable, then so too must (8)'s (if they obtain just in case (5)'s fail). In other words, the truth-table assumes that truth-conditions are (potentially) unrecognizable, and as such invokes a realist conception of truth (or at least fails to invoke an anti-realist conception). The same goes with the truth-functional interpretations of the other connectives.

Some have used this consequence to support a realist semantics.[1] The argument runs like this: Meaning is determined by use, therefore the meaning of the logical connectives will be determined by our use of them incorporated in our actual inferential practices.[2] As our actual inferential practices incorporate classical logic, and as classical logic assumes a realist construal of truth, our actual inferential practices warrant a realist semantics. For example, we accept an inference as valid just in case its conclusion is true whenever its premises are jointly true. Consider modus ponens: $P \rightarrow Q$, $P \vdash Q$; it is the case that wherever the premises are true the conclusion is also true, as is shown by a truth table analysis. Thus, logical validity and the meaning of the connectives are intimately related. What is not clear in the realist argument is whether the meanings of the connectives, taken as primitive, determine which inference patterns are valid, or whether valid inference patterns, taken as primitive, determine the meanings of the connectives.

---

[1]E.g. Scruton (1976). See also Dummett (1973a) p. 468, (1973c), and (1991b) Intro.

[2]The 'consequences of utterance' (i.e. the inferences we actually allow from some statement) are just as much an aspect of linguistic use as are 'conditions of utterance' (i.e. our propensity to assent to or dissent from certain sentences under certain conditions). Dummett (1973c) p. 221.

In one sense it does not really matter. Under the assumption that meaning is exhausted by use, the connectives will have whatever meaning is demanded by our use of them. In particular, we use connectives to draw certain conclusions - i.e. they are an ineliminable element in any adequate description of our inferential practices. If our inferential practices are such that we accept any application of, say, the disjunctive syllogism, then the conditional must have a meaning consistent with that use. The realist argument is this: as all applications of the disjunctive syllogism are in fact accepted by us, then the meaning of the disjunction must be that as expressed in its classical truth table. The notion of truth employed in expressing the meaning of the disjunction is that of recognition-transcendence, thus our inferential practices demand a realist semantics - i.e. it is only such a semantics which is consistent with our actual practices.

The more resonant example is this. Conjoining the truth-table for negation with that of disjunction yields "S $\vee$ $\neg$S" as a tautology.[1] Thus, if our inferential practices accept the truth of all instances of excluded middle, the truth-functional senses of both the negation and the disjunction would be validated. Thus, acceptance of an unrestricted application of excluded middle necessitates a realist understanding of the logical connectives and thus vindicates a realist semantics.

This argument rests upon two assumptions. First of all, it assumes that our

---

[1]This syntactic establishment of LEM as a tautology relies upon the specific truth-tables for negation and disjunction. Those truth-tables presuppose bivalence by assuming that all possible combinations of truth-values are accounted for. So *if* the truth-tables capture our actual inferential practices, then those practices validate both bivalence and excluded middle. Given this, it does no harm to treat bivalence and excluded middle on a par in the context of this discussion.

inferential practices are sacrosanct. In other words, there can be no distinction between what practices we in fact employ and what practices we ought to employ. Secondly, it assumes that it is clear what inferential practices we do employ.

Consider a simple thought experiment. Suppose it is the case that our (current) inferential practices validate an understanding of the negation and disjunction allowing unrestricted assertion of excluded middle - i.e. are such that all sentences of the form "$S \vee \neg S$" are deemed correctly assertible. Imagine, however, that a successful case is made for denying truth-valuedness to sentences containing vacuous referring-expressions; e.g. 'The present King of France is bald'. Under that supposition, 'Either the present King of France is bald or he is not' fails to be correctly assertible, and hence our current inferential practice would be deemed unacceptable and hence in need of revision. If such revision-inducing evidence is possible, then the fact that our (current) inferential practices warrants a realist semantics would be insufficient as an argument in support of that semantics. Appeal to an inferential practice would only suffice to warrant a semantics if that practice were the one we *ought* to adopt. The only admissible notion of an inferential practice which we *ought* to adopt is one in which the inferences it warrants are valid *absolutely* and not merely *relative* to that practice. The notion of absolute validity reverses the order of priority: an inferential practice would be acceptable *because* the inferences it permits are valid instead of an inference being valid *because* it is permitted by an inferential practice. Under the first assumption this is a mistake: practices are just what they are and no critique of them are possible. However, imagine a person whose inferential practices permitted applications of "$P \vee Q, P \vdash \neg Q$".

Surely someone, upon remembering being told of either Hitler or Stalin (but not remembering which) that they invaded Poland, then remembering that it was of Hitler, proceeded to answer "no" on a history exam question "Did Stalin invade Poland?" would have strong evidence to revise her inferential practice once she had her exam paper returned. I am inclined to agree with Dummett that:

> It cannot be a matter of taste whether a form of argument is valid or not: the meanings of the premises and the conclusion must determine whether or not the latter follows from the former.[1]

The second assumption is similar to the first - it assumes that we clearly understand our own practices. However, there may be two distinct practices, potentially delivering different results in certain applications, which we may confuse. Consider another thought-experiment. Before a certain time, t_n, our practice for assigning colour-predicates to swans was by appeal to the conditional statement:

a) If x is a swan, and x appears white, then x is white

Before t_n, all observed swans appeared white and hence all were deemed to be white. The sample of observed swans was so large that by t_n we, either overtly or covertly, accepted the universal statement 'All swans are white' - or:

---

[1]Dummett (1991b) p. 11. Later in the work he considers Prior's plonk connective, *, which uses $\vee$'s introduction rule ($P \vdash P*Q$) and $\wedge$'s elimination rules ($P*Q \vdash Q$). Inferences involving * can be criticized for allowing the derivation of contradictions from true sentences: let P be any true sentence, then by the introduction rule we can derive $P*\neg P$, and by the elimination rule we can derive $\neg P$ which, with $\wedge$-introduction, yields $P \wedge \neg P$. (p. 209). His conclusion is that an inferential practice "can be flawed or defective... With that, we perceive that our [inferential] practice is no more sacrosanct, no more certain to achieve the ends at which it is aimed, then our social, political, or economic practice." (pp. 214-215). (Dummett is a bit quick - the classical sentential calculus can be proved consistent, even if it does use itself as meta-language.) See also C. Wright (1987) Ch. 1.

b) If x is a swan, then x is white

After t_n, it may not be clear what our practice actually is. It may continue to be by an implicit appeal to (a), which makes reference to appearances, or it may be by an implicit appeal to (b) whose reference to appearance has disappeared.

There would be no problem if the two practices never in fact delivered a different result. Suppose, however, that at some time $t_{m>n}$, a swan was discovered which appeared black. According to our first possible practice, we would not assign whiteness to it, whereas according to our second possible practice, we would. Disagreement over which its colour was would indicate an unclarity about what our current practice was. Proponents of (b) as expressing the current practice would argue that the swan *must* be white in virtue of what they consider to be *the* uncriticizable practice of assigning colour-predicates to swans. That argument would utterly fail to move proponents of the first view, as they refuse to accept (b) as expressing *the* practice. Similarly proponents of the first view would argue that the swan *could not* be white in virtue of what they consider to be *the* uncriticizable practice of assigning colour-predicates to swans. And again the proponents of the second view would not be moved. The proponents of the first will claim that the others are merely confused about what the practice is, as will proponents of the second.

How can the dispute be resolved? It is tempting to think that it should be resolved by appeal to which purported practice *ought* to be the current practice, where the practice that ought to be adopted is, of course, the one which assigns the correct colour to particular swans. This does not help *if* we think that the notion of a colour

ascription *being correct* is given by the practice itself; i.e. both sides can equally claim that their practice delivers the correct colour - the colour which is correct *relative* to that practice. In other words, such a method must presuppose that there is an *absolute* notion of correct colour-ascriptions. But then, such a notion would reverse the former order of priority: instead of a colour being correctly ascribed to a particular swan *in virtue* of its conforming to a particular practice, a particular practice would be acceptable *in virtue* of it assigning the correct colour to particular swans. In other words, we would need a notion of correctness of colour-ascriptions which transcended our colour-ascribing practices.[1]

This is, of course, the same problem faced in our consideration of the first assumption. Unless we are prepared to accept that our inferential practices are sacrosanct, we should be prepared to concede that an inferential practice is acceptable just in case it sanctions valid inferences rather than an inference being valid just in case it is sanctioned by a particular inferential practice. In other words, we would need a notion of validity of inference which transcended our particular inferential practices; i.e. a notion of *absolute* validity. Thus, no appeal to a particular inferential practice could tell us whether or not a particular inference ought to be considered valid and hence whether or not a particular interpretation of a logical connective ought to be considered unavoidable.

In the case of the swans, such an absolute conception might be easier to find: the practice of ascribing colours to swans is embedded in the more widespread practice of

---

[1]See Weiss (1992) for similar remarks concerning alternative counting practices.

ascribing colours to objects in general. If it is agreed that the practice of ascribing colours in general proceeds on the basis of observation, then there is good reason for adopting (a) over (b) as an expression of the more acceptable practice. Finding practices which are more general than our inferential ones is considerably more difficult. In particular, it is difficult to see what would count as a non-circular conception of *absolute* validity - or how such a notion could possibly transcend particular inferential practices or systems.[1]

In any event, proponents of the first view may be able to side step such problems by an appeal to history. Prior to t_n, it was undeniable that the practice was characterized by appeal to (a). (b), they may claim, was essentially proposed as a short cut whose acceptability lay in their conviction that there would never be a tension between it and (a). That conviction was shown to be unfounded, and thus it must be given up.

The proponent of (b) may argue, however, that while it is true that the order of *discovery* goes from (a) to (b), the order of *justification* does not; (b) need not be justified by appeal to its relationship to (a). (a), they may hold, is part of an older legacy - one which we have wisely replaced by a more acceptable practice.

Returning to our inferential practices, it may not be clear what they currently are. Take excluded middle, for example. Is our practice to accept unrestricted applications of it, or is it to accept only instances for which one disjunct can be established? The

---

[1]This is essentially Dummett's point in (1973b) and (1991b) Ch. 2: "One of the tasks of a semantic theory is to explain the meanings of the logical constants." p. 54. See also Prawitz (1980).

classicist will claim the former while the intuitionist will claim the latter.[1] Each will accuse the other of being confused about what our practices in fact are.

We have seen the problems with attempting to argue that one's triumphed practice is the one which ought to be adopted. Going that route precludes one from making a fundamental appeal to practices at all in justifying inferences and hence a theory of meaning. Similarly, the classicist can rebuff the intuitionist's claim that historically instances of excluded middle were admitted for cases in which one of the disjuncts was known to be true; it was only after subsequent experience suggested that, for every pair $<S, \neg S>$, exactly one of them could be shown to be true, that an unrestricted acceptability of it was incorrectly assumed. They will merely claim that while there may have been empirical inputs in our discovery of the laws of logic, it does not follow that the laws themselves have any empirical content.

To confront the original realist argument, however, it does not matter if there is any admissible method for determining what our inferential practices actually are; as long as there can be reasonable disagreement over what they are, an appeal to a purported practice will fail to sufficiently establish a particular semantics. In other words, while it may be true that an acceptable semantics must conform to actual practices, one cannot non-circularly appeal to a set of practices as an argument in support of a particular semantics.[2]

---

[1]See Brouwer (1908), Heyting (1956), and Dummett (1977).

[2]Prawitz (1980) remarks that there is a form of equilibrium between our inferential practices and our inferential theory. I suspect that this is probably right, but it helps neither the realist nor the anti-realist on this point.

There is one further thing to note. Suppose that, as a matter of fact, no observable non-white swan ever came into existence. Under this supposition, proponents of (a) would never differ in their colour-ascriptions to swans from proponents of (b). Nevertheless, if meaning is exhausted by use - i.e. by the linguistic practices governing their use - the predicate 'is a white swan' must differ in meaning between proponents of (a) and proponents of (b) *even though* there may be in fact no noticeable difference in their respective use of that predicate.[1] The two practices are simply not the same; the one relies exclusively on observation and the other not at all. The upshot is that there may be a systematic unclarity in our own linguistic practices without it even being the case that we ever be aware of it.[2]

In discussing the manifestation argument a distinction was drawn between recognition-transcendent truth-conditions and unrecognizable truth-conditions to the effect that all unrecognizable truth-conditions are recognition-transcendent but it was unclear whether the converse held. Suppose it did not in the sense that there are recognition-transcendent but not unrecognizable truth-conditions. In this case, a classicist might hold that all instances of excluded middle be admissible independently of knowledge of at least one of their disjuncts holding. The intuitionist, on the other hand, will only admit instances where at least one of their disjuncts was known to hold. By supposition, at least one disjunct of an instance of excluded middle must have

---

[1]Thus Dummett's premise that meaning is exhausted by use is in jeopardy, calling into question the foundations of his central argument.

[2]Of course, methodologically, it is best to resist assumptions of wide-spread ignorance (see Currie (1993)). It is, however, only a methodological principle.

recognizable truth-conditions, and hence could be known to hold if only we exercised an appropriate recognitional capacity. Thus, by supposition, there would never be an instance of excluded middle to which a classicist and anti-classicist would deliver a differing assessment. Nonetheless, if meaning is exhausted by use - i.e. by linguistic practice - then the meanings of negation and disjunction must be different between classicists and anti-classicists *even though* there may be no noticeable difference between them.

If the foregoing is correct, then the realist argument is in even worse shape. It may be the case that the use of a particular inference be governed equally well by two distinct descriptions of our inferential practices. Call one such description D_1 and the other D_2. Suppose that both D_1 and D_2 equally support an inference I involving a single connective. According to D_1, that connective must be understood as meaning M_1 while according to D_2 it must be understood as meaning M_2. In the absence of an adequate method to decide between D_1 and D_2 (indeed it might never occur to us that our actual practice is ambiguous between them), appeal to D_1 in support of the connective meaning M_1 would be inconclusive. Such an appeal is warranted only if (i) it is known that D_1 and D_2 are distinct, and (ii) there is an decisive method (employed) for accepting one over the other. If either of these conjuncts fails, then the realist argument fails. We have seen reason to deny each of them in the case of our inferential practices, and hence there is strong reason to reject the realist argument.

Where does that leave us? It remains the case that validity and the meanings of the connectives are intimately linked. The validity of an inference at least partially

depends upon the meanings of any connectives employed, and the meanings of the logical connectives are at least partially constrained by which inferences are valid. Finally, classical logic, with its truth-functional interpretation of the logical connectives, assumes a realist conception of truth. An anti-realist semantics, it would seem, must assign non-classical meanings to the logical connectives, with the result that the range of valid inferences will be altered. In other words, a logic appropriate for an anti-realist semantics must, it seems, be revisionary. It is to these two issues that we must now turn.

Dummett finds the core of the logic he needs in the intuitionism of Brouwer.[1] At base, intuitionism replaces the classically central notion of truth with the notion of provability. Classically, in harmony with disquotation, an assertion of a simple sentence S should be understood tacitly as asserting "'S' is true". Thus, an assertion of ¬S will be understood classically as an assertion of "'S' is not true". Similarly, assertions of S∨P, S∧P, and S→P will be understood respectively as asserting "either 'S' is true or 'P' is true", "both 'S' is true and 'P' is true", and "if 'S' is true then 'P' is true".

The core intuitionistic understanding of the assertion of a simple sentence S is in terms of "'S' is provable". 'Provable' differs from 'truth' in that the former cannot, by definition, be a recognition-transcendent notion. In other words, the assertibility of a sentence S requires the actual existence of a recognizable procedure for determining S's

---

[1]See Brouwer (1908), (1923), (1952) and (1975) for pioneering work in intuitionism. Tennant (1987), on the other hand, opts for what he calls intuitionistic *relevant* logic. This difference need not concern us, as the main issue is whether the logical connectives should be understood in terms of truth (classically) or provability (intuitionistically).

truth-value, although it is an open question at this point whether that procedure must actually be carried out by us. The compound assertions $\neg S$, $S \vee P$, $S \wedge P$, and $S \rightarrow P$ will then be understood respectively as asserting "'S' is not provable"[1], "either 'S' is provable or 'P' is provable"[2], "both 'S' is provable and 'P' is provable", and "if 'S' is provable then 'P'

---

[1]Of course, we cannot read the 'not' classically - the intuitionist defines the negation in terms of contradiction rather than defining, as the classicist does, contradiction in terms negation. In mathematical discourse, for example, $\neg P$ is assertible just in case it can be shown that any purported proof of P could be transformed into a proof of a contradiction, say $1=0$. There are, however, some serious problems facing intuitionistic negation. Generalizing from the mathematical case, a negated empirical sentence, say "It is not the case that grass is white", is assertible just in case a purported proof of "Grass is white" could be transformed into a proof of some contradiction. But, there does not seem to be an empirical contradiction analogous to the mathematical $1=0$. Faced with this problem, anti-realists have offered alternative accounts of negation. C. Wright (1987) Ch. 10 offers this: a total state of information (TSI) justifies the assertion $\neg P$ iff it justifies the assertion that a TSI justifying assertion of P could not be obtained, no matter how thorough-going the investigation were conducted (he offers such a TSI interpretation for all the connectives). Luntley (1988) Ch. 4, on the other hand, offers: a proof of $\neg P$ is a construction which can be applied to a proof of P to yield a canonical derivation of a contradiction in the class of ϵ-sentences, where an ϵ-sentence is one which possesses 'experienceable' true or falsity (essentially, a sentence possesses experienceable truth or falsity if it conforms to his manifestation constraint by referring to an experienceable state-of-affairs and one's knowledge of P's meaning makes a detectable difference to experience). (Luntley also offers a similar account for all of the connectives.)

Even if anti-realists succeed in offering a more or less reasonable and broadly intuitionistic account of negated empirical sentences, other problems remain. Hossack (1990) argues that introducing contradiction as a primitive fails to account for the basic *incompatibility* involves in a contradiction, and that such an account can only be secured by eventually (i.e. in one's meta-theory) introducing a truth-functional negation. He further argues that if negation is interpreted truth-functionally at *any* level it must be interpreted truth-functionally at every level. B. Harrison (1983) and Price (1990) argue that a truth-functional construal of negation is not incompatible with manifestation requirements. Edgington (1981) argues that an intuitionistic interpretation is simply unreasonable (see her remarks concerning the disjunction in the next footnote). See also Meyer (1980) and Daniels (1990).

[2]Edgington (1981) offers some interesting arguments against the reasonableness of construing normal uses of the disjunction intuitionistically. Suppose I take my umbrella

is provable".[1]

On an aside, there is already an interesting difference between classical and intuitionistic logic. In both cases, reiterations of their 'truth'-predicates are allowed; "¬S" can be read classically as '"'S' is not true" is true' and intuitionistically as '"'S' is not provable" is provable'. However, classical reiterations are redundant - 'S' and '"S" is true' have the same truth-conditions - while intuitionistic reiterations are *not* redundant - 'S' is provable just in case a proof of it exists, while '"S" is provable' is provable just in case a proof of its provability exists, and there is no reason to suppose that the two proofs are

---

to work and upon returning realize that I do not have it. She (reasonably) remarks that 'either my umbrella is in my office or in my car' is assertible even though neither 'my umbrella is in my office' nor 'my umbrella is in my car' is. Secondly, if all we know is that Wilma is not an only child, then 'Wilma has a sibling' is assertible, but 'Wilma has either a brother or a sister' is not (as neither 'Wilma has a brother' nor 'Wilma has a sister' is assertible) even though 'has a sibling' and 'has either a brother or a sister' are synonymous. Dummett's reply, supplied by Edgington, is that in either case it is only a contingent fact that we have no evidence for either disjunct - the disjuncts are assertible in virtue of it being possible to gather the appropriate evidence. However, she notes that such a response would have the assertibility of normal disjunctions depend upon the assertibility (truth?) of associated subjunctive conditionals which Dummett thinks are themselves generally undecidable (see §3.1.2.2). Her argument is interesting but does, ultimately, attempt to establish a realist semantics by appeal to what (we think) our linguistic practices are (which we have just seen is, at best, extremely tentative).

On an aside, Edgington proposes justification-based but non-intuitionistic interpretations for both the negation and the disjunction. In a nutshell, where 'J' is read as 'the justification for asserting', the clauses 'J(A∨B)≥J(A)', 'J(A∨B)≥J(B)', and 'J(A) varies inversely with J(¬A)' capture the senses of the connectives. She furthermore argues that the principle of bivalence drops out as valid under these proposals.

[1]See Heyting (1956) and Dummett (1977). Slater (1988) pp. 63-64 thinks it a mystery why "intuitionism continues to be thought of as an alternative *propositional logic*, rather than a (mis-symbolized) *modal logic*" which takes 'provable' as its modal operator. The reason, I suspect, is that intuitionists want to eliminate *everywhere* the classical construal of the logical connectives, and thus cannot merely be offering a modal operator which operates on sentences containing truth-functionally interpreted internal connectives.

the same.

The most resonant difference between the two logics is revealed by the disjunction. As mentioned, conjoining classical negation and disjunction yields all instances of excluded middle as tautologies. Thus, every instance of excluded middle is classically assertible independently of whether one is capable of determining which of the disjuncts holds. Conjoining intuitionistic negation and disjunction yields an understanding of an assertion of excluded middle as "either 'S' is provable or '¬S' is provable", which in turn is understood as asserting "either there is a recognizable procedure which, if carried out, would prove 'S' or there is a recognizable procedure which, if carried out, would prove '¬S'" which is *not* a tautology unless it is assumed that at least one of such a pair of recognizable procedures must exist. That assumption would appear to be unwarranted for at least the instance of excluded middle formed by a disjunction consisting of (5) and (8) (i.e. concerning the temperature of the sun).

Quantifiers will similarly be understood. Classically, an assertion of $(\forall x)(Fx)$ will be understood as asserting "for all objects a, b, c, ... in domain D, 'a is F' is true and 'b is F' is true and 'c is F' is true and ..." and an assertion of $(\exists x)(Fx)$ will be understood as asserting "for at least one object a, b, c, ... in domain D, either 'a is F' is true or 'b is F' is true or 'c is F' is true or ...". Intuitionistically, they will be understood as respectively asserting "for all objects a, b, c, ... in domain D, 'a is F' is provable and 'b is F' is provable and 'c is F' is provable and ..." and "for at least one object a, b, c, ... in domain D, either 'a is F' is provable or 'b is F' is provable or 'c is F' is provable or ...".[1]

---

[1]See Heyting (1956), and Dummett (1977) and (1982)

Intuitionism thus seems to be the logic required for the anti-realist. It is the only - or at least the best known - logic on the market which takes a non-recognition-transcendent notion of truth in the form of provability as its central concept, and thus is in harmony with an anti-realist semantics. There is one last thing to note, however. We have seen that there are two distinct forms that an anti-realist semantics might take - it might take either verification or falsification as its central notion. A specific intuitionistic understanding of the logical connectives would similarly have to harmonize with whichever of the two competing central notions was adopted. The core of an intuitionistic logic underlying MT_1 would understand the assertion of a simple sentence S as asserting "'S' is verifiable". It is more difficult to see what the core of an intuitionism underlying MT_2 would be. The obvious candidate would be that an assertion of a simple sentence S should be understood as asserting "'S' is not falsifiable", but that would assume a pre-analyzed meaning for the negation. A less obvious candidate would be to assign a primitive meaning to 'unfalsifiable' (i.e. it would not be equivalent to 'not falsifiable') and render an assertion of S as "'S' is unfalsifiable". Such a move would necessitate two non-interderivable central notions: falsifiable and unfalsifiable. Perhaps difficulties like these should be seen as tipping the scale towards accepting MT_1 over MT_2.

At any event, if our linguistic practices demands a semantics which takes a non-recognition-transcendent notion of truth as its central concept, and our inferential practices form a sub-class of our linguistic practices, then our inferential practices had better also take a non-recognition-transcendent notion of truth as its central notion. As

our inferential practices are guided by an underlying logic, that underlying logic in turn had better involve such a notion of truth. Thus, classical logic must give way to (at least something like) intuitionistic logic in an anti-realist theory of meaning.

Before moving on, we must note that this (admittedly brief) argument presupposes that a choice of logic requires a semantic validation - i.e. that classical inferences are valid in virtue of the meanings assigned to the connectives. The rejected realist argument attempted to reverse the order of priority: it assumed our accepted inferences were valid and attempted to justify a realist semantics on that basis. However, it is not clear that a choice of logic requires such a validation at all.[1] Dummett has recently offered a new approach to the problem.[2] He discusses the possibility of offering a purely syntactic (as opposed to semantic), or proof-theoretic, validation for a choice of logic. The idea is this: a logic would be justified if its connectives display *harmony*: being able, by appeal only to a connective's introduction rules (i.e. those rules of inference allowing one to derive a sentence containing some connective from sentences not containing it, taken to be self-justified) to justify that connective's elimination rules (i.e. those rules of inference allowing one to derive a sentence not containing some

---

[1]Dummett (1963b), (1969), (1973b) and (1973c), Prawitz (1980) and (1987), and Luntley (1988) all insist on a semantic validation, and as classical logic rests on a realist semantics (they suppose), which is inadequate (they argue), classical logic is not acceptable. Pearce and Rantala (1982) agree that classical logic entails a realist semantics, while C. Wright (1987) Ch. 11 and Rasmussen and Ravnkilde (1982) §6 do not (although the latter do argue that semantic realism is incompatible with an intuitionistic logic). One of my main conclusions is that a realist semantics is not in serious jeopardy, and thus classical logic would not be in any serious danger from this count (Haack (1982) hints at a similar response).

[2]Dummett (1991b) Chs. 11-13.

connective from a sentence which does).[1] For example, the elimination rules for conjunction ($A \wedge B \vdash A$ and $A \wedge B \vdash B$) can be proof-theoretically justified by appeal to its introduction rule ($A,B \vdash A \wedge B$), for if we have a proof of A (or A is a premise) and we have a proof of B (or B is a premise), which are needed to entail $A \wedge B$, then we have a proof of A (and also of B) - namely the proof that established A as used in the introduction rule.

It would seem, of course, that the choice of introduction rules as basic is arbitrary - surely harmony would also be shown if the introduction rules could be justified by appeal to the elimination rules (if the introduction rules can be derived from the elimination rules, and *vice versa*, the logic is said to exemplify *stability*). However, Dummett notes that classical and intuitionistic logic disagree only over the elimination rule for negation[2]; thus, taking the introduction rules as basic, the logic which can canonically establish the elimination rules would be proof-theoretically justified. Dummett calls this the 'fundamental assumption': "if we have a valid argument for a complex statement, we can construct a valid argument for it which finishes with an application of one of the introduction rules governing its principle operator".[3] The anti-realist hope, then, is that intuitionistic logic will permit, while classical logic will not, a canonical proof for its negation elimination rule.

---

[1]A proof which involves only atomic premises and the introduction rules is termed *canonical*.

[2]Classical: $\neg\neg A \vdash A$; intuitionistic: $A,\neg A \vdash B$.

[3]Dummett (1991b) p. 254. I.e. if the logic is proof-theoretically justified, no application of an elimination rule need be used, as they can all be replaced by the canonical proofs justifying them.

While there is some promise for intuitionism, Dummett's investigations (while interesting, pioneering, and obviously important) are somewhat inconclusive.[1] Dummett himself sums up by admitting that "our examination of the fundamental assumption has left it very shaky. As applied to the disjunction operator, we have had to interpret it quite broadly; the need for this exemplified a general feature of reasoning about empirical matters, namely, the perverse decay of information."[2] Finally, he admits that the proof-theoretic method can only demonstrate the validity of a logic - failure to demonstrate a proof-theoretic justification is not sufficient to establish a logic's invalidity. Rejection of a logic, therefore, requires a non-syntactic (i.e. a semantic) argument.[3] Thus, the excursion into proof-theoretic justification has not discharged the need for a semantic validation (at least in terms of criticizing a choice of logic). All in all, it seems best to take the semantic argument as the strongest one guiding a choice of logic (if such a choice even requires validation). We can also tentatively agree that classical logic is best correlated with a realist semantics and intuitionistic logic is best correlated with an anti-realist semantics.

It would seem, then, that if you change the meaning of an expression you change

---

[1]Luntley (1988) Ch. 4 offers strong criticism against such an approach, concluding that only semantic considerations can justify a logic.

[2]Dummett (1991b) p. 277. Related to this last point, Weir (1986) argues that at best intuitionism offers a reasonable interpretation of the connectives in areas of discourse dealing with indefeasible (e.g. mathematics) as opposed to defeasible (e.g. empirical) sentences.

[3]Dummett (1991b) Ch. 14.

the underlying logic governing inferences involving that expression. But, if you change the underlying logic governing inferences, it would seem that you must also change the inferential practices involving that expression. In other words, it would seem that the price to pay for an anti-realist semantics would be a wholesale revision of our inferential practices:

> [Replacement] of the notions of truth and falsity, as the central notions for the theory of meaning, by those of verification and falsification must result in a different logic, that is, in the rejection of certain forms of argument which are valid on a classical, i.e. two-valued, interpretation of the logical constants. In this respect, the linguistic practice which we actually learn is in conformity with the realists' conception of meaning: repudiation of realism as a philosophical doctrine entails revisionism about certain features of actual use.[1]

---

[1]Dummett (1973a) p. 468. See also (1991b) p. 302: "The view that a revision of [the logical laws] involves a change in the meaning of the logical constants is inshakable." Not all philosophers of an anti-realist bent accept Dummett's claim that an anti-realist semantics must be revisionary. For example, C. Wright (1987) Ch. 10 and Putnam (1976a) argue that intuitionism need not be revisionary. C. Wright's argument is this: anti-realism is bound to be revisionary if it rejects various theorems of classical logic - in particular, if it rejects either LEM or DNE. It will reject those theorems if there are sentences which are not effectively decidable. According to (C. Wright's interpretation of) intuitionistic negation, $\neg P$ is warranted just in case it can be shown that no proof of P can be constructed. Let Q be an undecidable sentence. As such, neither Q nor $\neg Q$ can be proven. But, if it can be shown that Q cannot be proven, then such a demonstration will suffice to establish $\neg Q$, contrary to the assumption that Q is undecidable. Therefore, there cannot be any undecidable sentences, so there is no reason to reject either LEM or DNE, and hence anti-realism need not be revisionary (of course, C. Wright's argument depends upon his interpretation of intuitionistic negation - see Luntley (1988) Ch. 4 and Weiss (1992) for serious criticism of it). (Compare this to Dummett (1991b) p. 319: "...it is plausible that a semantic theory could be constructed for empirical statements that would yield standard intuitionistic logic. Under such a semantic theory, it will be impossible to identify any statement as being neither true nor false, just as, in intuitionistic mathematics, there are no statements identifiable as neither provable nor refutable: for to say of a statement that it was not true would be to declare that it could never have been verified, which is just to declare it false." Compare the claim that it is impossible to identify a sentence as neither verifiable/true nor falsifiable/false with the argument in §3.1.2.2.) Putnam argues that the connectives of either classical or intuitionistic logic can be reinterpreted - without pragmatic difference

There are at least two examples where it seems clear that an anti-realist semantics must involve revisionary consequences. Consider, first of all, the law of Double Negation Elimination - i.e. $\neg\neg S \vdash S$. According to the classical truth-table S and $\neg\neg S$ will have the same truth-value under all the same conditions, and thus DNE is guaranteed to be truth-preserving and hence valid. Intuitionistically, however, an inference is valid just in case it is proof-preserving in the sense that in an inference $S \vdash P$, any proof of S is sufficient to prove P. DNE is intuitionistically valid, therefore, if and only if any proof of S is sufficient to prove $\neg\neg S$ and vice versa. Now according to the intended intuitionistic meanings, we are to understand respectively the assertions of S as "'S' is provable'", $\neg S$ as "it is provable that 'S' is not provable"[1], and $\neg\neg S$ as "it is provable that '"S" is not provable' is not provable".[2]

Let (5) replace S, and assume that neither (5) nor (8)[3] is provable (i.e. we do not have capacities which would allow us to prove either (5) or (8)). Because we can prove that we cannot prove (8), the negation of (8) is warranted. The negation of (8) is, of course, the double-negation of (5). However, whereas we have a proof for the double negation of (5), we do not, by supposition, have a proof for (5) itself. Thus, a proof for

---

- in terms of each other. I will not consider his arguments, but they appear dubious given the seemingly obvious fact that, as will be discussed, intuitionism straightforwardly rejects certain inferences accepted by the classicist. See also McDowell (1976) §7-8 and Horwich (1982).

[1]I.e. that a purported proof of "S" would lead to a proof of a contradiction. See Heyting (1956), Dummett (1977), and Dummett (1991b) Chapter 13.

[2]I.e. that a purported proof that a purported proof of "S" would lead to a proof of a contradiction would lead to a proof of a contradiction.

[3]I.e. the classical negation of (5).

$\neg\neg(5)$ is not sufficient to prove (5) and thus DNE cannot be deemed valid from an intuitionistic perspective.

The second example is even more resonant. Classical logic allows both an unrestricted application of excluded middle and the standard $\vee$-elimination rules. Let T(S) mean "'S' is true" and V(S) mean "'S' is verifiable". If anti-realism is not revisionary in rejecting either unrestricted application of excluded middle or standard $\vee$-elimination rules, then it would seem to be provably self-refuting:

a) $T(5) \vee T(8)$	premise[1]
b) $\neg V(5) \wedge \neg V(8)$	premise
c) $\neg V(5)$	b - $\wedge$ elim.
d) $\neg V(8)$	b - $\wedge$ elim.
e) $T(5)$	assumption
f) $T(5) \wedge \neg V(5)$	c,e - $\wedge$ intro.
g) $(\exists S)(T(S) \wedge \neg V(S))$	f - $\exists$ intro.
h) $T(8)$	assumption
i) $T(8) \wedge \neg V(8)$	d,h - $\wedge$ intro.
j) $(\exists S)(T(S) \wedge \neg V(S))$	i - $\exists$ intro.
k) $(\exists S)(T(S) \wedge \neg V(S))$	a,b,e,g,h,j - $\vee$ elim.

The conclusion establishes that there is at least one sentence which is true but not verifiable, and hence truth cannot be co-extensive with verifiability; i.e. it must *transcend* the verifiable. This is a direct denial of the anti-realist position. Thus, there seems to be good reason to suppose that anti-realists must accept a revisionary attitude towards our inferential practices; such an attitude appears to be necessitated (i) by proposed new meanings of the logical connectives and (ii) on pain of self-refutation. However, the revisionary attitude, if adopted, need not be seen as *ad hoc*. If the arguments from acquisition and manifestation go through, then there cannot be an admissible notion of

---

[1]Line (a) is, of course, just an instance of excluded middle.

truth underlying classical logic; it would simply be hoisted on its own petard.

If that were the case, however, it is difficult to see *how* intuitionism *could* be revisionary. A call for revision could only be acceptable if the practice to be revised were a possible, though ultimately inadequate, one; in other words, only if there *were* an inferential practice invoking classical logic which *ought* to be replaced by one invoking intuitionistic logic. But, if the arguments from acquisition and manifestation go through, an inferential practice invoking classical logic could simply *not* be a possible one - not for humans anyway.[1]

There may be another sense in which it could be considered revisionary - namely that it involves a revision not of our inferential *practices* but of our *understanding* of our inferential practices. Recall the divergent practices regarding colour-ascriptions to swans. If members of that culture never in fact disagree over particular cases, then we can regard members not as having two distinct practices, characterized by (a) and (b), but as having two distinct understandings of their practices. In this case (a) and (b) would represent an understanding that some member would have towards their own practice of assigning colours to swans.

In the case of inferential practices, we might be able to regard classical logic and intuitionistic logic not as sanctioning divergent inferential practices but as expressing divergent understandings of what our inferential practices are. In other words, it need

---

[1]Cooper (1978) invokes a similar argument in defending LEM from the semantic paradoxes: "[If] one says the class-paradoxical sentences are meaningless [then] the disjunction of a class-paradoxical sentence and its contradictory [is not] a valid substitution-instance of LEM." (p. 166).

not be the case that each accuses the other of either accepting as valid inferences which are invalid absolutely, or rejecting as invalid inferences which are valid absolutely; it may be the case that each merely accuses the other of holding inadequate understandings of the logical connectives employed.

This would seem to be closer to the mark. If the anti-realist is correct and *no* intelligible meaning can be given to classical truth-conditions, then it is simply incoherent to suppose that inferences warranted by those truth-conditions may or may not be valid. On the other hand, one may draw a valid inference even though they only partially (or not at all) understand the connectives sanctioning the inference. The anti-realist revisionary call, therefore, may be seen as a call to abandon illusory classical understandings for genuine intuitionist ones.

The upshot is that if there is any interesting relation between semantics and logic - i.e. if any possible meanings of the connectives must conform to an adequate theory of meaning - and if classical logic involves a realist semantics while intuitionistic logic involves an anti-realist semantics, then there must be a divergence in the meanings each assigns to the connectives. That in turn entails that there must be some manifestable difference in the respective uses of expressions containing them; i.e. there must be some inferences which are classically but not intuitionistically valid, or vice versa. The anti-realist would be advised, then, to advocate abandoning current inferential practices purportedly captured by classical logic in favour of those captured by intuitionism. We saw a problem with this: if the core problem with classical logic is that it assigns unintelligible meanings to its connectives, then its 'inferences' must involve unintelligible

sentences, and thus can be neither valid nor invalid. If this were the case, then we could not understand the call for revision in terms of replacing one system which allows as valid absolutely invalid inferences with one which only allows as valid absolutely valid ones.

Recall the analogy. Proponents of (a) may claim that there are certain applications of (b) which are perfectly acceptable - namely those which are limited to previously observed swans. (b), they may argue, is only unacceptable when it becomes extended to included cases of non-observed swans. In other words, it has an acceptable core use, but an unacceptable extension.

The anti-classicist could say the same thing about the classical connectives - they have an intelligible use when limited to sentences with recognizable truth-conditions, but become unintelligible when extended to sentences with unrecognizable truth-conditions. What is wrong with the classical meanings is that they allow application to such sentences - it is under such extensions that they lose their intelligibility.[1] The call for revision, then, can be seen as a call to stop such unrestricted extension, the result being that inferences extended inadmissibly no longer be considered valid. The restricted classical senses of the connectives would, the argument goes, turn out to coincide with those of the intuitionist. It seems to me that the intuitionist should not be disheartened

[1]As we shall see, Dummett's argument is that semantic realism founders when we attempt to extend our theory of meaning from sentences with recognizable truth-conditions to sentences lacking such truth-conditions.

by her revisionary tendencies, and that we should understand revision in this latter sense.[1]

While the relation of logics to theories of meaning is interesting and important, it is not centrally relevant to my concerns. Realism is the view which accepts, while anti-realism is the view which rejects, the thesis that truth may be recognitionally transcendent. Classical negation and disjunction, for example, allow for the possibility that sentences which resist verification or falsification will continue to possess determinate truth-values, while intuitionism does not. Thus, while I agree that the validity of bivalence is not the central issue in the debate, I will assume that classical logic is the preferred logic of the realist while intuitionistic logic is the preferred logic of the anti-realist.

---

[1]Much of the preceding discussion has tended to link classical logic to semantic realism and intuitionistic logic to semantic anti-realism so closely as to almost make them equivalent. However, as mentioned, Rasmussen and Ravnkilde (1982) §6 argue that semantic realism is compatible with various multi-valued logics and hence, at best, semantic realism is *consistent* with classical logic (it is not, they argue, consistent with intuitionistic logic). In an influential paper, McDowell (1976) attempts to completely sever the two links. Anti-realism, he argues, may endorse a two-valued logic which refuses to countenance any counter-examples to bivalence. Furthermore, semantic realism may embed its theory of sense in an intuitionistic proof theory. In essence, McDowell thinks that there need not be a lot of difference between realists and anti-realists.

3.0 RESPONSES TO THE NEGATIVE PROGRAMME

3.1 Problems with Unrecognizability

3.1.1 Recognition-Transcendence

Both the manifestation and acquisition arguments depend upon the realist being committed to sentences with unrecognizable truth-conditions. They do not, we should note, depend *per se* upon a commitment to sentences with recognition-transcendent truth-conditions. Thus, the arguments tacitly assume that a commitment to recognition-transcendence entails commitment to unrecognizability. Is that assumption acceptable?

As argued, a truth-condition is recognition-transcendent just in case it is wholly independent of any epistemic facts about humans. Now, a sentence S is true just in case its truth-conditions obtain, and is false just in case they fail to obtain. We can say, then, that a set of truth-conditions determines - by its obtaining or otherwise - the truth-value of its associated sentence. Thus, for the realist both the obtaining or otherwise of a set of truth-conditions and the truth-value determined for a sentence are independent of any epistemic facts about humans. In particular, they are independent of facts regarding the capacities of humans to recognize if and when they obtain. Thus, for the realist, it must remain an open possibility that a truth-condition obtain unrecognized. It does not follow from this, however, that there are sentences with unrecognizable truth-conditions.

A truth-condition is recognizable just in case we have a capacity to determine that such a condition obtains when it does (or fails to obtain when it does), although we need not actually exercise that capacity. A truth-condition is unrecognizable, on the other hand, just in case we do not have a capacity to determine that such a condition obtains

when it does (or fails to when it does). It is the case, then, that whether a truth-condition recognizably obtains depends upon certain epistemic facts about humans.

The anti-realist may, at this point, raise a *prima facie* damaging dilemma for the realist. Either all truth-conditions are recognizable, or some are unrecognizable. If some truth-conditions are unrecognizable, then the acquisition and manifestation arguments appear to go through. If all truth-conditions are recognizable[1], and recognizability depends upon epistemic facts about humans, then truth-conditions *simpliciter* depend upon epistemic facts about humans and the recognition-transcendence conception of truth is lost.

The apparent strength of this dilemma rests upon a simple but subtle mistake; namely a question-begging identification of the obtaining of a truth-condition with the (potential) recognition of its obtaining. Consider the truth-conditions associated with sentence (6) asserting that the upper face of the die in the box shows six pips. The realist will accept *both* that they are recognition-transcendent *and* that they are recognizable; i.e. that whether they obtain is independent of epistemic facts about humans but that we nonetheless have a capacity to recognize that they obtain when they do (that capacity is exercised in the first but not the second scenario). The realist, therefore, will insist that there are two 'facts' involved in this case: (i) the *obtaining* of

[1]The anti-realist is, of course, committed to this view; by definition, verification-conditions must be recognizable, and the anti-realist conceives of truth in terms of verification.

the truth-conditions and (ii) the *recognition* of their obtaining.[1]

The realist therefore accepts at least the weaker claim that *some* sentences have recognizable truth-conditions without in any way compromising their recognition-transcendent notion of truth. Is there anything which prevents them from generalizing this attitude to *all* sentences? In other words, is the following a coherent position?: associated with each fact concerning the obtaining or otherwise of any sentence's truth-condition is an *independent* fact concerning the potential for recognizing the first. The anti-realist cannot merely *assume* that it is incoherent - they must supply an *argument* to that effect. The dilemma posed above does not do the job.

Our assessment of the coherence of the position will depend upon a clear understanding of what it involves. Our envisaged realist will accept:

> a) If it is a fact that a set of truth-conditions recognizably obtain, then it must be a fact that they obtain.

and:

> b) If it is a fact that a set of truth-conditions obtain, then it is a fact that they recognizably obtain.

but not:

> c) If it is a fact that a set of truth-conditions obtain, then it *must* be a fact that they recognizably obtain.

If (c) held, then, as the fact mentioned in the consequent depends upon certain epistemic facts about humans, the fact mentioned in the antecedent would subsequently also

---

[1]McGinn (1982a) distinguishes two senses of 'recognition transcendence': (i) no evidence can be found bearing on the truth of the statement, and (ii) evidence can be found but is inconclusive (p. 123). His distinction, while interesting, does not affect my argument.

depend upon those facts, and the recognition-transcendent notion of truth would be lost.

Therefore, our envisaged realist must take it as a non-logical fact that all truth-conditions are recognizable ones. Being a non-logical fact, they must concede the possibility that some truth-conditions are unrecognizable - they need not, however, concede it as an *actualized* possibility. This may seem an incredibly weak distinction, but recall that the manifestation argument depends upon the realist being committed to the *existence* of sentences with unrecognizable truth-conditions, not merely to the possibility of such sentences.

Let me support it another way. All that is required of our hypothetical realist is acceptance of:

d) It is possible for some sentence to have unrecognizable truth-conditions.

The *possibility* of a sentence's truth-conditions being unrecognizable is consistent with their *actually* being recognizable. To say that a sentence's truth-conditions are possibly unrecognizable is only to say that it is possible that humans lack the capacity to recognize when they obtain. Thus, all the realist need commit herself to is that no intelligible sentence (that is, a sentence correctly expressible in a language L) is such that its truth-conditions are, in fact, unrecognizable.

Thus, the recognition-transcendent notion of truth, while consistent with the existence of unrecognizable truth-conditions, need not entail it. A separate argument that such sentences actually exist must therefore be given. It may seem that this point is trivial - it has been suggested that the existence of sentences with unrecognizable

truth-conditions is as well established a fact as any in philosophy.[1] Nonetheless, I beg the reader's indulgence on this point (until, at least, §3.1.2.2).

3.1.2 Decidability

3.1.2.1 Decidability and Recognizability

Instead of talking in terms of recognizable and unrecognizable truth-conditions, Dummett talks about sentences being decidable or undecidable:

> Our language contains many sentences for which we have no effective means, even in principle, of deciding whether statements made by means of them are true or false; let us label them 'undecidable sentences'. If it is assumed that truth is subject to the principle of bivalence - that every sentence is determinately either true or false - the language also contains sentences for which we have no ground for thinking that, if true, we must in principle be capable of being in a position to recognize them as true.[2]

The core notion is that a sentence is decidable just in case there exists "an effective procedure for determining whether or not their truth-conditions are fulfilled."[3] Conversely, a sentence is undecidable just in case there is no such procedure. Our notion of a sentence having either recognizable or unrecognizable truth-conditions is obviously quite similar to the present notion of a sentence being either decidable or undecidable. Recall sentence (6). There is, unproblematically, a procedure for determining whether it is true or false - namely opening the box and observing the state of the die. In the first scenario that procedure was in fact carried out determining (6)

---

[1]E.g. that involved in Gödel's Theorem.

[2]Dummett (1991b) pp. 314-315.

[3]Dummett (1976b) p. 81.

to be true, and hence it is unquestionably decidable. In the second scenario, while that procedure was not in fact carried out, it could have been carried out; in other words, even in the second scenario there *is* a procedure for determining its truth-value, and hence it remains decidable. In general we can say that if a sentence has recognizable truth-conditions - i.e. conditions that we have a capacity for recognizing as obtaining when they do - then the (potential) exercise of that capacity serves as the procedure for determining truth-value, and hence all sentences with recognizable truth-conditions are decidable.

Does the converse hold; i.e. is it the case that all sentences which are decidable have recognizable truth-conditions? That depends upon whether it is a constraint on the admissibility of such a procedure that we have a capacity to recognize the results that it would deliver if carried out. It is not so obvious that it is. Consider sentence (5) (concerning the temperature of a particular spot on the sun). *Prima facie* it seems not unreasonable to suppose that there exists a procedure for determining its truth-value - namely go to that spot and time and take a temperature reading - but, due to our limitations of heat tolerance, we do not have a capacity to recognize the results of that procedure. Thus, there is some reason to suppose that not all decidable sentences have recognizable truth-conditions.

To contest this view, one would have to argue that there is no good reason to suppose that such a procedure exists. As we saw, it is not a constraint on the existence of such a procedure that it actually be carried out (recall sentence (6) in the first scenario), but one could argue that it is a constraint that it *could* be carried out. A

procedure is an extended operation; it requires potential completion for its existence. Consider the following proposed procedure for determining the number of digits in the expansion of π:

a) map the digits in the expansion of π one-to-one with the natural numbers

This at best can only be considered a *partial* procedure; to express a full procedure it would have to be supplemented with something like:

b) carry on until the digits are exhausted, then the highest mapped natural number indicates the number of digits in the expansion of π

Now, as it is not possible that such a procedure be completed - the digits will never be exhausted - the proposed method cannot count as a procedure for determining the number of digits in the expansion of π. Recall the proposed procedure for determining the truth-value of sentence (5):

c) go to that spot and time and take a temperature reading

which at best expresses a partial procedure. It requires, for completion, something like:

d) observe the results of the temperature reading

which, due to our human limitations, is not capable of being completed. Thus, it is only an illusion to suppose that there is a procedure for determining the truth-value of (5). Sentence (5), then, would seem to be undecidable.

This result can, I think, be generalized: it is a constraint on any purported decision procedure that it *could* be completed, whether or not it actually is; i.e. a decision procedure must be *completable*. A purported decision procedure is completable just in case its results can be recognized. Thus, it seems that we have good reason to suppose that all decidable sentences have recognizable truth-conditions. We are thus, it seems,

in no danger if we identify sentences with recognizable truth-conditions with decidable sentences, and sentences with unrecognizable truth-conditions with undecidable sentences. We can therefore (provisionally) accept:

> DEC) A sentence S is decidable if and only if (i) it has recognizable truth-conditions, if and only if (ii) there is a procedure such that, if carried out, would determine S's truth-value.
>
> UND) A sentence S is undecidable if and only if it is not decidable; i.e. just in case either (i) or (ii) in (DEC) fail for S.

There are, however, three residual ambiguities in this account: (a) what are the temporal constraints on the existential quantification in (ii)?, (b) for whom must the truth-conditions be recognizable?, and (c) what are the capacities by means of which a truth-condition is recognizable?

A decision procedure, like anything else, can exist in either the past, present, or future. Let 'exists' be construed tenselessly. We can say, then, that a decision procedure existed if it exists it the past, currently exists if it exists in the present, and will exist if it exists in the future. Is satisfaction of *any* of these existential possibilities sufficient for the decidability of some sentence? There are a number of possible positions regarding this question. First of all, one might maintain that decidability is, as it were, an atemporal notion; i.e. if a sentence is *ever* decidable, then it is *always* decidable. The converse of course holds - if a sentence is *always* decidable, then it is decidable at *some* time. Let "D_tS" abbreviate "'S' is decidable at t", which in turn is to be understood as "there exists, at t, a decision procedure which, if carried out, would determine S's truth-value". The proposed thesis, then, is:

DT_1) $(\exists t)(D_tS) \equiv (\forall t)(D_tS)$[1]

Recalling the (pragmatic) identification of decidability with the recognizability of truth-conditions, under this proposal as long as a sentence is decidable at any time, it has recognizable truth-conditions and hence poses no special problems for a realist theory of meaning.

Alternatively one might hold that decidability is not an atemporal notion - the property of being decidable is one which a sentence may have at certain times and lack at others. A sentence is decidable for only as long as its decision procedure exists - i.e. can be carried out. Letting 'P' be a decision procedure for some sentence, this position can be expressed as:

DT_2) $(\forall t)[(\exists P)(P_tS) \equiv D_tS]$[2]

This position will offer some special problems for a realist semantics. Consider the sentence:

9) Halley's comet is composed mostly of ice.

For simplicity, suppose the only admissible procedure for determining its truth-value is

---

[1]Notice that (DT_1) is similar to a standard realist atemporal construal of truth: if a sentence is ever true then it is always true. Compare DT_1 to C. Wright's (1987) I/S Anti-Realism.

[2]DT_2 captures, I believe, the essence of a position which has come to be called *actualism*. Griffin (1993) argues that anti-realism entails actualism, which is too high a price to pay. The relation between anti-realism and actualism is, I believe, analogous to that between intuitionism and strict finitism (which limits the range of determinately true or false mathematical statements to those whose verification or falsification are humanly feasible (See C. Wright (1987) p. 112 for examples)), though Dummett (1970), C. Wright (1987) Ch. 4 and Mitchell (1992) argue that intuitionism does not collapse into finitism. Compare DT_2 to C. Wright's (1987) I/N Anti-Realism.

by way of directly taking and observing samples from the comet. Suppose, also, that in 1986 we had short range probes capable of collecting and retrieving such a sample (and that we will continue to have such probes). Thus, according to DT_2, (9) *was* decidable in 1986, *will be* decidable in 2061, but is *not* decidable in 1994; in 1994, it has unrecognizable truth-conditions and semantic realism is in jeopardy.

There is, finally, a middle position. One might maintain that if a sentence is decidable at any given time, it will remain decidable at all subsequent times:

DT_3) $(\exists t)(D_tS) \rightarrow (\forall t' \geq t)(D_{t'}S)$[1]

Call DT_1 the atemporal conception of decidability, DT_2 the strictly temporal conception, and DT_3 the partially temporal conception. Which of the three is most acceptable?

Problems arise for DT_2 from the fact that procedures are temporally extended operations. Consider a future tensed sentence:

10) Clinton will serve a second term as American President.

On an initial reading of DT_2, as there will be a decision procedure for determining (10)'s truth-value in 1996, (10) will be decidable in 1996, but is not decidable now. That procedure can be expressed something like:

> P_{10}) After the closing of the polls in the 1996 U.S. Presidential election, count the ballots cast. If Clinton has more votes than any other candidate then (10) is true and if he has less votes than some other candidate then (10) is false.

---

[1]Compare to C. Wright's (1987) I/NP Anti-Realism. This seems to be Dummett's preferred version; his proposed interpretation of intuitionistic logic in terms of Beth Trees presupposes something like it: "It is evident that we ought to admit as an axiom $[(\vdash_n A) \rightarrow A]$: if we know that, at any stage, A has been (or will be) proved, then we are certainly entitled to assert A." ((1973c) p. 233).

However, there is no reason for ruling out an alternative decision procedure for (10) in the form:

> P_{10}') *Wait until* the polls close in the 1996 U.S. Presidential election, then count the ballots cast...

(P_{10}') is procedure which *now* exists (and indeed existed at any arbitrary time in the past) such that, if carried out to its completion, will determine (10)'s truth-value. Thus, a decision procedure for (10) exists in 1994 and thus it is *now* decidable. This result can be generalized - any future-tensed statement for which a decision procedure *will* exist is a statement for which a decision procedure *does* (and *did*) exist.[1] It is indeed even more general than that: *any* statement for which a decision procedure exists at time t_m is a statement for which a decision procedure exists at all times $t_{o \leq m}$. Thus, if a statement is decidable at any time, it must be decidable at any prior time. Thus DT_2 collapses into:

DT_4) $(\exists t)(D_t S) \rightarrow (\forall t' \leq t)(D_{t'} S)$[2]

Notice that DT_4 conjoined with DT_3 is equivalent to DT_1. Let S be a sentence decidable at some future time. By the first conjunct it is decidable at all past times. If it is decidable at all past times, then it is decidable at some past time. By the second conjunct it is decidable at all future times. Therefore, if S is decidable at any time, then

---

[1]C. Wright (1987) Ch. 5 remarks that this result follows once one realizes that no decision procedure is instantaneous.

[2]Compare to C. Wright's (1987) I/NF Anti-Realism. The converse should unobjectionably hold. It states that if a sentence is decidable at all times prior to t_m, then it is decidable at t_m. At most it commits us to holding that a sentence retains its decidability for short periods - only from the instant *before* t_m to t_m itself. There should then be nothing controversial - nor anything interesting - about replacing the conditional in DT_4 with a biconditional.

it is decidable at all times, as DT_1 states. It follows, then, that if a proponent of DT_3 rejects DT_1 then she had better reject DT_4; similarly, if a proponent of DT_4 rejects DT_1 then she had better reject DT_3. We are yet, however, no closer to determining which of the three is preferable.

We can, I think, eliminate DT_3 as exhaustively expressing the temporal relations inherent in the notion of decidability. In order to be distinct from DT_1, it must allow there to be sentences which are decidable at some future time but not at either the present or any past time - i.e. for it to be adequate DT_4 must be inadequate. However, we have just seen an argument to the effect that DT_4 must, taken individually, be considered a component of *whatever* we accept as our final conception of decidability. Thus, DT_3 cannot constitute a complete understanding of the temporal relations inherent in decidability.

Can DT_4 constitute such a complete conception? I know of no conclusive arguments either way, but there are some rhetorical arguments in favour of each side. In order to be distinct from DT_1, it must allow there to be sentences decidable in either the past or the present which are not decidable in the future. In other words, it requires it to be possible that a sentence can lose its property of being decidable.

A case can be made for such a possibility. Recall our culture that made a practice of ascribing colours to swans. Suppose that they settled their differences, and opted for (a) (making reference to appearances) as expressing their social practice. (a) also determines, it should be noted, a procedure for determining the truth-value of any sentence of the form:

11) Swan A is white.

That procedure can be described by something like:

P_{11}) Suitably place yourself in front of swan A under normal lighting conditions and observe it. If it appears white, then (11) is true, otherwise it is false.

An instance of (11) is decidable at any time t_m just in case either P_{11} can be carried out at t_m or any $t_{n \geq m}$. Suppose that it can be carried out at t_n but that at t_{n+1} global blindness inflicts our culture. Thus, (11) is decidable at t_n but not at $t_{m>n}$ - i.e. it loses its property of being decidable.

On the other hand, the possibility of a sentence losing its decidability can be quite disconcerting, as reflection on the relation between truth and decidability will show. If a sentence is decidable at t_1, then it is true (or false) at t_1. According to a standard atemporal construal of truth, if a sentence is true (or false) at any time, say t_1, then it is true (or false) at all times. Now the anti-realist advocates an identification of the condition of a sentence being true with the condition of its being verifiable. Moreover, the anti-realist is committed to a certain relationship between verifiability and decidability. If a sentence is verifiable, then we have a capacity to determine it as true. It does not follow, however, that if a sentence is not verifiable then we have a capacity to determine it as false; the anti-realist is not committed to the view that there is, associated with *every* sentence S, a pair of sentences $\langle S, \neg S \rangle$ such that exactly one member is verifiable.

Suppose, however, that S is decidable (at t). That means, by definition, that there is a procedure (at t) which would, if carried out, determine S's truth-value. S's being decidable (at t) therefore presupposes that it *has* a truth-value (at t). Now ¬S (even

understood intuitionistically), given that S *has* a truth-value, must take the alternate value to whatever S takes. Hence, as long as S is decidable (at t), there must be associated with it a pair of sentences <S,¬S> such that exactly one member must be verifiable - even an anti-realist can accept bivalence for decidable sentences: the anti-realist is committed to the view that if a sentence S is verifiable then it is decidable, and if it is decidable, then either it or its negation is true.

If our normal practice accepts that a sentence, once true, is always true, and we accept the identification of truth with verifiability, then we are faced with the consequence that a sentence, once verifiable, is always verifiable. Thus, if a sentence is decidable in virtue of its being verifiable at any time, then it must be decidable at all times.

This argument assumes, of course, that the anti-realist accepts the normal practice of supposing that a sentence, once true, is always true. An anti-realist proponent of DT_4 can take the above argument as an argument *against* accepting the atemporal construal of truth. However, it is unlikely that even an anti-realist will so reject that practice. Truth must be assumed to be at least a fairly stable property of sentences. For example, if a mathematical statement M is verified at t_1, and hence deemed true at t_1, we do not hesitate to assume it is true at all subsequent times. That is, we do not feel the need to reprove M each time it is employed in a subsequent proof. But if it is possible for M to cease to be true (in the anti-realist sense) even after it has been proved at t_1, then we can have no guarantee that a subsequent proof employing M will be correct.

Proponents of DT_4 will be unimpressed by the mathematical case. They will agree

that a subsequent proof employing M will depend upon M continuing to be true, but that assumption can always be checked by reproving it. It is only if we lost our capacity to reprove M that a subsequent proof employing it would be suspect; but if we lost our capacity to reprove M then surely we would have lost our capacity to construct the subsequent proof anyway. Mathematics tends to be thought of as an atemporal enterprise - passage of time is not generally thought to diminish our capacity to reprove theorems.

The empirical case is more difficult. Suppose that at t_1 our culture succeed in verifying sentence (11). They then, at t_2, formulate the hypothesis:

12) Swan A was white at t_1 and sea water tastes salty.

If their practices are sufficiently similar to ours, they will attempt to verify (12) by tasting sea water - they will not bother to verify the first conjunct but will merely assume it to be true in virtue of (11) being true at t_1. In other words, they would exploit a certain truth-value link:

a) If "S" is true at t_1, then "'S' was true at t_1" is true at t_2.

It seems undeniable that (a) represents a significant aspect of our linguistic practices; Dummett accepts them as "fundamental features of our understanding of tensed-sentences, [ones which play] a predominate role in our training in the use of these statements."[1] However, as the natives do not, owing to global blindness, have a capacity to verify "'S' is true at t_1", *nor* to verify "'¬S' is true at t_1", they cannot assume that S is decidable at t_2, hence cannot assume that (a) represents an admissible practice.

---

[1]Dummett (1969) p. 364.

Therefore they cannot assume that the first conjunct of (12) is decidable. In other words, DT_4 and the acceptance of certain truth-value links like (a) are at odds. If we are more committed to such truth-value links than we are to DT_4 being a complete account, then we would be well advised to reject DT_4 in favour of DT_1.[1]

The proponent of DT_4 may raise this response. While we cannot assume that such truth-value links will hold unconditionally, in the absence of any reason to suppose that they break down, we are warranted to accept them. In the case of the swan, if (11) switches from being decidable at t_1 to not decidable at t_2, then we would have to assume that it switched from being true at t_1 to not being true at t_2. We have no reason to believe that it ceases to be true, and hence have no reason to believe that the truth-value link is not admissible in this case. Moreover, we may have reason to believe that it does *not* cease to be true. Suppose that at t_1 the natives actually performed the procedure to determine (11)'s truth-value, and it was found to be true. They will then have memories of that test being carried out and of what its result was. Those memories will then constitute sufficient evidence that (11) does not cease to be true at t_2.

In other words, the proponent of DT_4 may offer one of the following defenses for (a):

> i) As long as there is sufficient evidence to doubt that S loses its truth-value, then (a) is admissible.
>
> ii) As long as there is no evidence to suppose that S loses its truth-value, then (a)

---

[1]Dummett (1969) rejects the realist's use of such truth-value links. His argument, however, is that they cannot ultimately ground a realist conception of truth for past-tensed sentences, not that they cannot ground an atemporal notion of decidability (which is our present concern). See also C. Wright (1987) Chs. 3 and 5.

is admissible.

However, it is difficult to see what sort of evidence could be raised for rejecting any particular application of (a). Any such evidence could just as easily be construed as demonstrating that we were merely wrong in our initial assignment of a truth-value to S. Of course, if at t_1 our verification was conclusive this would not be possible, but empirical sentences are simply not the sorts of things which admit of conclusive verification. Thus, while (i) and (ii) indicate a possible line of defense, they strike me as insufficient.

Where does that leave us? While there are not conclusive grounds for ruling out DT_4 in favour of DT_1, it seems to me that DT_1 is more in line with our actual linguistic practices. I do not hold that our linguistic practices are sacrosanct, but we need a stronger motivation for revising them than that such a revision would allow for DT_4. Therefore, unless a stronger argument is mounted, we are safe to assume that DT_1 captures the sense of the existential quantifier utilized in (DEC) and (UND).

So there is reason to understand DEC in terms of a tenseless construal of the existential quantifier; there is no significant difference between a sentence being decidable *at* t_1 and being decidable *at* t_2 - all temporal indicators in DEC are intersubstitutable.

The decidability of a sentence, according to DEC, also depends upon it having recognizable truth-conditions. What is unclear, however, is to whom the 'we' refers. There are a number of possible positions.

In the first place, one might hold that there is a significant distinction between a sentence being decidable *for one person* P_1 and being decidable *for another person* P_2; i.e. a sentence S is decidable for a person P just in case P has a capacity to recognize whether or not S's truth-conditions obtain. Let 'CxY' mean that person x has the capacity to recognize whether or not the truth-conditions for sentence Y obtain, and let 'DxY' mean that sentence Y is decidable for person x. This first position can then be represented as:

$RR_1)$ $(\forall p)(DpS \equiv CpS)$

On the other hand we can conceive a position which denies that there is any significant difference between a sentence being decidable *for* P_1 and being decidable *for* P_2 - i.e. that in terms of decidability, personal indicators are intersubstitutable. Yet, we cannot ignore the fact that not everyone has the same capacities - how can decidability be universal if possession of relevant capacities are not? There are two ways to go on this question. In the first place, we could concede that universality of decidability cannot be compatible with the non-universality of relevant capacities and hence restrict genuine decidability to cases where universality of relevant capacities is secured. In other words, we might hold that a sentence is only genuinely decidable if everyone is capable of determining whether its truth-conditions obtain. A sentence will then be decidable for one person just in case it is decidable for everyone:

$RR_2)$ $(\forall p)(DpS) \equiv (\forall q)(CqS)$

On the other hand, we may simply deny that universality of decidability requires

universality of relevant capacities.[1] We may attempt to get around the *prima facie* problem by supposing that possession of the relevant capacity by at least one person would be sufficient for its associated sentence to be decidable for all:

$RR_3)$ $(\forall)(DpS) \equiv (\exists q)(CqS)$

Let me illustrate these three positions. Consider sentence (11). Imagine again that (a) represents the culture's linguistic practice of ascribing whiteness to swans and hence serves as a member's procedure for determining the truth-value of any sentence like (11). Suppose further the culture is populated by exactly two sighted persons, P_1 and P_2, and one blind person, P_3. According to RR_1, (11) is decidable for P_1 and P_2, but not for P_3. According to RR_2, (11) is not decidable for any of the three in virtue of P_3 lacking the capacity to determine its truth-value. Finally, according to RR_3, (11) is decidable for each of the three in virtue of either P_1's or P_2's capacity to determine its truth-value.

However, the three positions become ambiguous when temporal considerations are thrown in. Recall that, according to DT_1, the second clause of DEC is satisfied even if global blindness strikes our hypothetical culture between t_1 and t_2. It is not clear, however, whether the first clause would be satisfied in that event. According to either RR_1 or RR_3, (11) is decidable for no member at t_2 as no member at t_2 any longer has the capacity to determine whether or not its truth-conditions obtain, and hence is not

---

[1]McGinn (1976) notes that Dummett seems to assume that recognitional capacities are constant across a given linguistic community and that, as this assumption is clearly false, Dummett's argument is weakened (p. 23). Charity, however, requires us to refrain from attributing this assumption to anti-realism in general.

decidable at t_2 *simpliciter*. (11) is similarly undecidable at t_2 according to RR_2 as not every member has the relevant capacity at t_2. The results of one interpretation of RR_1-RR_3 seem clearly in tension with DT_1.

What we require is some way of understanding RR_1-RR_3 which does not generate this *prima facie* tension. The most obvious way would be to temporally modify RR_1-RR_3. RR_1 should thus be replaced by something like:

RR_1') $(\forall p)(DpS)_t \equiv (\exists t')((CpS)_{t'})$

which reads that a sentence S is decidable for any person P at time t just in case there is a time t′ at which P has the capacity to determine whether or not S's truth-conditions obtain. Thus, even in the event of global blinding between t_1 and t_2, as P_1 had the relevant capacity at t_1, (11) is decidable for her at t_2 (and indeed at all times). If P_3 was born blind, however, and never acquires sight, then (11) is undecidable for her at all times.

RR_2 and RR_3 need to be similarly modified:

RR_2') $(\forall p)(DpS)_t \equiv (\forall q)(\exists t')((CqS)_{t'})$

RR_3') $(\forall p)(DpS)_t \equiv (\exists q)(\exists t')((CqS)_{t'})$

The realist will resist the claim that such temporal modifications will suffice to capture the range of the recognizable. Consider the sentence:

13) The first swan was white.

Under the reasonable assumption that swans predate humans, no one at any time has the capacity to determine whether (13)'s truth-conditions obtain. Thus, according to any of RR_1'-RR_3', (13) will fail to be decidable. If it fails to be decidable, then it has

unrecognizable truth-conditions. Thus, if the arguments from acquisition and manifestation go through, (13) would cause serious problems for semantic realism. Positions RR_1'-RR_3' thus represent possible *anti-realist* construals on the limitations of the persons for whom a truth-condition must be recognizable in order for its associated sentence to be decidable.

In regards to (13), the realist will typically say that *had* there been someone around when the first swan was hatched and matured, they *would* have been able to determine whether or not its truth-conditions obtained.[1] It is in virtue of the subjunctive conditional, they will argue, that (13) should be considered decidable. What such a realist is aiming at is that the decidability of sentences depends not upon the capacities of *actual* persons but rather on the capacities of *possible* persons. In other words, they would call for a further modification of RR_3' to the effect:

RR_4) $(\forall p)(DpS)_t \equiv \Diamond(\exists q)(\exists t')((CqS)_{t'})$[2]

Under RR_4, sentence (11) will be deemed decidable for each of P_1, P_2, and P_3 at either t_1 or t_2 (or any other time). For now, let us leave the question of whether the realist construal or a generic anti-realist one is more acceptable and concentrate on which of RR_1'-RR_3' is most acceptable. The only arguments are, I think, pragmatic ones. RR_2' clearly places too stringent a criteria on decidability - no English sentence describing

---

[1]See, for example, Appiah (1986): "Why should we accept that we could not construct ways of placing people which would allow (or would have allowed) them to confirm sentences of the kind [the anti-realist] offers?... Surely if we had been placed in the remote past, we would have recognized whether or not, for example, Cleopatra ate dates before clasping the asp to her bosom." (p. 44).

[2]The precise limitations on the modal operator will be discussed shortly.

how things look would be decidable as long as there is at least one blind English speaker. RR_1', by stressing the capacities of individuals, would preclude any general theory of meaning: at best we could formulate a theory of meaning for a language L *relative* to an individual P. We could, I suppose, view a theory of meaning for L as a (possibly infinite) conjunction of such meaning-theories for individuals, but then the possibility of constructing such a theory - which in a sense Dummett's whole case depends upon - would seem utterly implausible. By default, then, it seems that something like RR_3' would be the most likely anti-realist position. It seems not unreasonable, then, to suppose that from an anti-realist perspective a sentence S is decidable just in case there tenselessly exists a person who, at some time, is capable of determining whether or not S's truth-conditions obtain. Accepting something like RR_4 as the most likely realist position extends the range of the decidable to those decidable by non-actual but possible persons.

What admissible substitutions in the first existential quantifier in RR_4 are intended?[1] There are two options. We may wish to restrict the evidence gathering powers of admissible substitutions to those of actual persons. For example, actual persons do not have the capacity to identify medium sized physical objects situated more than a certain distance away, or to transport themselves to certain spatial or temporal locations, etc. The capacities of possible persons would be co-extensive with such a list of capacities of actual persons. Alternatively, we may wish to considerably extend the

[1]The following views expressed owe a strong debt to Appiah (1986) Ch. 4.

evidence-gathering powers of possible persons. Reminiscent of the late (and recently resurrected) Superman, we may wish to conceive of our possible person as having x-ray and telescopic vision, virtually limitless in his spatial and temporal movements, etc.[1]

Consider, for illustration, sentence (5) concerning the temperature of some portion of the sun. No actual person has the capacity to tolerate such intense degrees of heat needed in order to determine whether its truth-conditions obtain. However, Superman would have no problem doing so. Under the proposed restriction of the capacities of possible persons to those of actual persons, it would appear that even under RR_4 (5) would be undecidable, whereas under the proposed extension to the capacities of hypothetical super-beings it would be decidable.

The main problem with the first proposal is that it is very difficult to see how to adequately restrict the range of actual capacities. Utilization of instruments greatly enhance our evidence gathering capacities. For example, prior to the invention of the telescope, the sentence:

14) Mars has two moons.

would have been deemed, incorrectly, either false or undecidable. Unless we can determine the upper limit on the possible evidence-gathering-aiding instruments we can or will construct, we cannot determine the upper limit on the extent of actual human capacities.

---

[1]C. Wright (1987), for example, suggests that a sentence is decidable just in case "an appropriately large but finite extension of our capacities would confer on us the ability to verify it or falsify it." (p. 113). He later adds "at some time or other" (p. 180) thereby inserting a thesis of temporality.

On the other hand, it is question-begging to extend the notion of a possible person to include omniscient gods, although Dummett assumes, inappropriately, that realists wish to do exactly that:

> The fundamental difference between the anti-realist and the realist lies in this: that, [the] anti-realist interprets 'capable of being known' to mean 'capable of being known *by us*', whereas the realist interprets it to mean 'capable of being known by some hypothetical being whose intellectual capacities and powers of observation may exceed our own.[1]

Dummett's assumption is uncharitable - the reasonable realist will admit *some* significant restrictions on admissible capacities.[2] Let me make a distinction between *internal* and *external* evidence-gathering capacities. An internal capacity is one which actual humans have unaided by such instruments - i.e. capacities limited to the unaided exercise of our five senses. An external capacity is one which humans have in virtue of their utilization of instruments. It seems reasonable to restrict the notion of a possible person to those whose internal evidence-gathering capacities are co-extensive with those of actual humans - these capacities can, I imagine, be determined empirically. The difficulty then is in locating the admissible extension of external evidence-gathering capacities.

Unquestionably there would be logical limits placed on such an extension. Consider:

15) An undetectable gremlin lives in my refrigerator.[3]

---

[1]Dummett (1959a) p. 24.

[2]Appiah (1986), for example, argues that decidability requires only "that there be a logically possible test, one that *we human beings with our actual powers* might or might not be able to carry out." (p. 57, emphasis added).

[3]The example is taken from Rosenberg (1983).

As there can be no possible evidence by which one can detect the presence of an undetectable gremlin, no capacity for gathering such evidence can be possible. Therefore, under no extension of human capacities would (15) be rendered decidable.

We may also wish to impose certain theoretical limits. It is generally thought to be theoretically impossible to determine simultaneously the position and velocity of particular electrons. Thus, no extension of a human capacity would allow:

16) At time t electron e is in position p moving at velocity v.

to be decidable. Alternatively, we may wish to impose theoretical restrictions. For example, we may exclude instruments whose results of application are unrecognizable or non-understandable by the internal capacities of humans. Suppose we were able to construct a machine with near infinite heat toleration which beeped only when the surrounding temperature was exactly 13,099,341°K. Such a machine would detect and indicate whether sentence (5)'s truth-conditions obtained, but we would be unable to recognize the evidence that it provides. The existence and operation of such a device would not then constitute an admissible extension on our evidence-gathering capacities. Similarly, some mathematicians suppose that the Four Colour Problem has now been solved by use of a high-speed computer.[1] However, the proof generated by the computer is so extensive that it is beyond the lifespan of any (or several) human(s) to 'go through' the proof, and thus the use of such a computer may be deemed inadmissible.

Clearly, however, even imposing such restrictions leaves the extent of the

---

[1]Appel and Haken (1977).

capacities of our possible person terribly open-ended.[1] One might attempt further restrictions by imposing temporal limitations - a possible human at a given time t is one whose capacities are co-extensive with the capacities of actual humans at t. Besides being *ad hoc*, such a restriction is at tension with the results of previous sections. By DT_1 (and both RR_3' and RR_4), if some actual future person utilizes a currently non-existent (but logically and theoretically admissible) instrument to determine the truth-value of some sentence S, then S is *now* decidable. By the current proposal, however, such an actual future person cannot *now* be deemed a possible person (which is odd in itself) and hence S cannot now be decidable *even though* it will be decidable.

The best it seems we can do is to understand a possible person as one whose internal evidence-gathering capacities are co-extensive with those of actual humans, and whose external evidence-gathering capacities do not exceed logical and theoretical limitations.

At any event, we are finally in a position to understand the most reasonable construals of decidability and undecidability. According to the anti-realist:

> ARD) A sentence S is decidable just in case there exists a person P such that there is some time t at which P can manifest a capacity to determine S's truth-value.

and according to the realist:

> RD) A sentence S is decidable just in case it is possible that there exist a person P such that there is some time t at which P can manifest a capacity to determine S's truth-value.

---

[1]Putnam (1990) p. ix distinguishes his non-realism from positivism by refusing to "*limit in advance* what means of verification may become available to human beings."

On both accounts, a sentence is undecidable just in case it is not decidable.

3.1.2.2 The Extent of the Undecidable

The anti-realist must argue, and not merely assume, that sentences with unrecognizable truth-conditions exist. Such sentences are identified with undecidable sentences. Thus, the anti-realist owes us an argument that genuinely undecidable sentences exist. Dummett offers these:

> Three features of our language may be singled out as especially responsible for the occurrence of undecidable sentences.
> (i) Our capacity to refer to inaccessible regions of space-time, such as the past and the spatially remote.
> (ii) The use of unbounded quantification over infinite totalities, for example, over all future time.
> (iii) Our use of the subjunctive conditional.[1]

Let me go through the case for each of these in turn.

Recall Dummett's general testing procedure for determining whether a person P grasps the meaning of some sentence S: (i) determine S's associated T-sentence; (ii) situate P appropriately to investigate the states-of-affairs relevant to determining whether the condition mentioned on the right-side of the T-sentence holds; (iii) see if P assents to S (or dissents if S is false). If a sentence refers to a space/time region inaccessible to us, then it will be impossible to situate someone appropriately to investigate the states-of-affairs in that region, and hence it will be impossible to demonstrate a grasp

[1]Dummett (1991b) p. 315. See also (1976b) p. 81 and p. 98. Although Dummett leaves open room for other cases, he nowhere, to my knowledge, discusses any.

of the sentence's meaning. In other words, we would be unable to manifest our understanding of the sentence, violating Dummett's basic constraint on any acceptable theory of meaning. Consider the sentence:

17) Caesar crossed the Rubicon.

Grasp of (17), on Dummett's account, requires the capacity to assent to (17) when situated in a position from which the state-of-affairs of Caesar crossing the Rubicon can be observed. But of course, no one can now so situate themselves and hence, if grasp of (17) consists in grasp of its truth-conditions, no one now could grasp its meaning.

On the other hand, we have just seen arguments to the effect that as long as a sentence is decidable it has recognizable truth-conditions, and as long as it has recognizable truth-conditions it poses no problems for a realist semantics. On our best understandings of decidability - either ARD or RD - (17) comes out decidable; there exists a person - Caesar himself for example - who is capable, at the time he crossed the Rubicon, of determining its truth-value.[1]

How do we resolve the apparent tension? We can either appeal to the testing procedure to restrict the range of the decidable, or modify our understanding of the testing procedure in light of our understanding of decidability.

It is important to notice that Dummett's proposed testing procedure is a

---

[1]Slater (1988) offers much the same defense for subjunctive conditionals (to be discussed later): "Indeed it is undecided whether Jones was brave or not, but our inability to *project* how Jones would have behaved in a test situation still leaves the disjunctive decidable, and hence, 'valid', since the sovereign body in question, namely Jones himself, had the power, by pure and simple will, to decide it." (p. 63). See pp. 61-62 for a similar point regarding quantification over infinite domains.

procedure for testing the understanding, and hence intelligibility, of particular sentences for particular individuals; it is *not* a procedure for testing the general intelligibility of sentences *simpliciter*. For example, it is a condition on P's understanding of S that P is capable of succeeding in such a testing procedure. However, P's failure to succeed cannot, in itself, be taken as having any implication regarding the understanding another has of S. Only if it is a general feature about P - a feature which she shares in common with all other persons - which is responsible for her failure in the testing procedure would that failure have an implication on the understandings of others. Recall our paralytic would-be bicycle rider. Her failure to manifest a capacity to ride a bicycle is due to features peculiar to herself and hence cannot be taken as implying anything about the implicit knowledge of bicycle riding that others may have. Only if the feature responsible for her failure were universal - i.e. if everyone were paralytic - would her failure be generalizable into a general lack of knowledge of bicycle riding.

The same argument can be raised in the case of linguistic understanding - the failure of an individual P in a testing procedure for S cannot be taken as implying a general lack of understanding of S *unless* the features responsible for P's failure were universal ones. Being situated in the 20th century is the feature responsible for, say, *my* failure to succeed in (17)'s associated testing-procedure, but that feature is not a universal one.

We have, then, two distinct views about the relation of a testing procedure to the understandability of a sentence. On the one hand, we have the view that a sentence S is understandable for P just in case P is capable of succeeding in S's associated testing

procedure. Let 'UxY' mean that sentence Y is capable of being understood by person x, and let 'TxY' mean that person x is capable of succeeding in Y's associated testing procedure. Thus, according to this first view:

TU_1) $(\forall p)(UpS \equiv TpS)$

TU_1 precludes the possibility of *my* understanding sentence (17) under a realist account of understanding. On the other hand, it may be held that understandability requires only that *someone* is capable of succeeding in a testing procedure:

TU_2) $(\forall p)(UpS) \equiv (\exists q)(TqS)$

Compare TU_2 to RR_3. There is a certain isomorphism between both the notions of understandability and decidability and the notions of a decision-procedure for truth-value and a testing-procedure for understanding. Such isomorphisms are precisely what we want if we wish to modify our notion of a testing-procedure to harmonize with our understanding of decidability (i.e. ARD or RD). We can, in fact, continue to modify our notion of a testing procedure to harmonize with RR_3':

TU_2') $(\forall p)(UpS)_t \equiv (\exists q)(\exists t')((TqS)_{t'})$

Now, not unexpectedly, the realist will wish to modify even this notion of a testing procedure to allow for the non-actualized possibility that there exist a person capable of succeeding in a testing situation. For example, even if there does not in fact exist a person who could have succeeded in the testing procedure for sentence (13) (concerning the colour of the first swan), it is possible that there existed such a person in the relevantly restricted notion of a possible person. They will thus advocate acceptance of something like:

TU_3) $(\forall p)(UpS) \equiv \Diamond(\exists q)(TqS)$

which can be similarly modified to include temporal constraints:

TU_3') $(\forall p)((UpS)_t \equiv \Diamond(\exists q)(\exists t')(TqS)_{t'}))$

Opting for either of TU_2' or TU_3', it would seem that (most) sentences referring to space/time regions inaccessible to *us* would pose no problem for a realist semantics.

There is a common problem in both of these accounts - how can the fact that someone in the 1st Century BC can manifest their understanding of (17) entail that people in the 20th Century are capable of understanding it *even though* no person in the 20th Century can manifest that understanding? If no one in the 20th century can manifest their understanding of it, then no one in the 20th century can understand it *simpliciter.* This consequence appears just as destructive for a realist semantics as a universal failure to succeed in (17)'s associated testing procedure.

There is a possible reply to this objection. One might argue that it has not been established that persons in the 20th century are precluded from succeeding in (17)'s associated testing procedure. The potential for a person P succeeding in such a procedure presupposes the truth of the sentence:

> 18) If P were suitably placed, then P would have the capacity to determine (17)'s truth-value.

Now the anti-realist cannot conceive of (18) as a material conditional with a false antecedent - that would, classically at any rate, render it true, and hence would provide sufficient evidence that P has the required capacities for understanding (17). They must rather conceive it as a subjunctive conditional with an unactualized antecedent. Thus, the undecidability of (17) will presuppose the undecidability of (18). Generalized, the

undecidability of any sentence referring to a region of time/space inaccessible to us would presuppose the undecidability of an associated subjunctive conditional. Thus, Dummett's arguments for the undecidability of sentences refering to spatio/temporally remote regions will depend upon his arguments for the undecidability of subjunctive conditionals.

It may be objected that not only does the antecedent refer to a non-actualized possibility, it refers to an impossibility. Due to constraints on the temporal movements of humans, it is not possible for a 20th century person to be relevantly situated. As such, no one in the 20th century can understand (18), and hence no one could understand (17).

The reply to this objection mirrors some considerations made above. It may be argued that P's understanding of a sentence S does not require that P understand what it would be like for *her* to determine S's truth-value, but only what it would be like for *someone* to determine its truth-value. For example, a statement made by P ascribing pain to herself may be understood by Q not in terms of Q's understanding of what it would be like for him to determine the truth-value of P's sentence, but what it would be like for P to determine the truth-value of her sentence. On this scheme, understanding a sentence S requires only an understanding of the truth-conditions of:

> 19) If someone were to be suitably placed, then they would be capable of determining S's truth-value.

Again, S's undecidability is parasitic upon the undecidability of its associated sentence of type (19) - i.e. upon the undecidability of a relevant subjunctive conditional. Thus, we need a closer examination of Dummett's arguments for the undecidability of certain subjunctive conditionals.

Before turning to that issue, we should note that the sentences we have so far been dealing with are all ones for which there is a local failure of their associated testing procedure - i.e. sentences which refer to regions of space/time which are inaccessible to some but not to others. What about sentences which refer to regions of space/time which are inaccessible *in principle*? For such sentences, the realist could not exploit the fact that there is someone for whom the sentence must be deemed decidable to argue that it must be deemed decidable for all. Furthermore, she would have difficulty in appealing to a subjunctive conditional of the form (19), as such conditionals would, of necessity, have unactualizable antecedents - it is simply not clear what the status of a subjunctive conditional with an unactualizable antecedent would be.

The best defense for the realist, it seems to me, is simply to doubt whether there are, in fact, any regions of space/time which are *in principle* inaccessible to us.[1] Dummett assumes that the distant past is so inaccessible, but at best it is inaccessible *for us*. We have seen an argument that as long as it is accessible for someone, then a sentence referring to it is decidable. Recall sentence (13). As long as we assume that swans predate humans, then the temporal region it refers to is inaccessible to all. It does not follow, however, that it is inaccessible *in principle*. At this point the realist can merely dig in her heels and insist on TU_3' - there are no grounds for denying that it is possible for there to have been a human who was capable of determining the colour of the first swan.

---

[1]What about sentence (5) referring to some region on the sun? While it is likely impossible for us to actually position ourselves on the sun, such a region may very well be accessible via scientific instruments; e.g. probes, telescopes, spectrometers, etc.

It seems to me, therefore, that the best case that the anti-realist can make for such sentences being undecidable on the basis of inaccessible space/time regions and hence problematic for the semantic realist rests on their case for the undecidability of certain subjunctive conditionals. Before turning to that issue, however, we will examine his second case on the basis of quantification over infinite totalities.

Recall that a sentence S is undecidable only under the condition that no admissible person has the capacity to determine its truth-value. What this means is that no admissible person has the capacity to determine either (i) that S is true or (ii) that S is false. If an admissible person is able to determine either of (i) or (ii), then S fails to be undecidable. Consider Dummett's example of an undecidable sentence of this class:

20) There will never be a city built on this spot.[1]

If (20) is genuinely undecidable, then no admissible person has the capacity to determine either that it is true or that it is false. In other words, both the following must hold:

a) It is not possible to determine that there will never be a city built on this spot.

b) It is not possible to determine that it is not the case that there will never be a city built on this spot.

Even to the realist whose notion of truth allows for the unrestricted applicability of bivalence, (a) is unobjectionable. Accepting bivalence, either it is the case that there will never be a city built on this spot or it is not the case that there will never be a city

---

[1]Dummett (1959a) pp. 16-17.

built on this spot. Suppose that it is not the case that no such city is ever built. What this entails is that there is a time at which a city is built on this spot. Under this supposition, as it is false that there will never be such a city built, it would not be possible to establish that no such city will be built - i.e. (a) would be acceptable. On the other hand, even if it were the case that no such city were ever built, at no time would we be in an epistemic position to be aware of that truth, and as such it would not be possible to establish that no such city will ever be built - i.e. (a) would still be acceptable. In other words, not even the realist, who asserts that either there will never be such a city built or else it is not the case that there will never be such a city built, can object to clause (a). Clause (b), on the other hand, is not unobjectionable.

If a city is built on this spot in, say, 50 years time, then it would be possible to establish that it is not the case that there will never be a city built on this spot - its possibility is shown by its (supposed) actuality. Thus, if such a city is built, then (b) is unwarranted, and (20) is not undecidable. In other words, (20) is undecidable only if (b) is warranted.

Suppose for the sake of argument that (b) is not warranted. It would then follow that (20) is not undecidable. Curiously, however, it would not follow from this that it is decidable. (20) is not undecidable because one of the conditions for it to be undecidable does not hold. It does not follow from this that the conditions for it to be decidable do hold. A sentence S is decidable only if an admissible person can determine its truth-value. What this means is that S is decidable just in case an admissible person can either (i) determine it as true or (ii) determine it as false. Mirroring the conditions for

undecidability, for (20) to be decidable exactly one of the following must hold:

c) It is possible to determine that there will never be a city built on this spot.

d) It is possible to determine that it is not the case that there will never be a city built on this spot.

but the unacceptability of (b) simply does not entail the acceptability of either (c) or of (d).

The crucial question regarding (20)'s undecidability, then, is whether or not clause (b) is acceptable. Clearly it would be unacceptable under the assumption that (20)'s falsity-conditions obtain (recognizably, as it would turn out - rendering (20) decidable). Equally clearly it would be acceptable under the assumption that (20)'s truth-conditions obtain (unrecognizably, in this case - ensuring that (20) is undecidable). Thus, it seems that *only* under the assumption that (20)'s truth-conditions obtain (unrecognizably) is (b) acceptable.

(20) is thus a queer sentence - it is decidable or undecidable depending on whether certain non-epistemic facts hold. I will call statements like it *asymmetrically undecidable*. A statement is asymmetrically undecidable if either (i) it can be determined *as* true *if* true but cannot be determined *as* false *if* false or (ii) it can be determined *as* false *if* false but cannot be determined *as* true *if* true. In other words, a statement S is undecidable *simpliciter* if both its associated (a) and (b) clauses are acceptable but is *asymmetrically* undecidable if only exactly one of those respective clauses is acceptable.

It is my contention that all sentences quantifying over an infinite domain are asymmetrically undecidable. A universally quantified sentence - $(\forall x)Fx$ - would be

falsified by a single member of the domain; i.e. by a finite portion of the domain. Thus, if false, it is recognizably false, and hence decidable. Only a determination of its truth requires a survey of the entire domain - a capacity which no admissible person has - and hence only if it is (unrecognizably) true would it be undecidable. On the other hand an existentially quantified sentence - $(\exists x)Fx$ - if true would only require a survey of some finite portion of the domain and hence would be decidable. Only a determination of its falsity requires a survey of the entire domain, and hence only if it is (unrecognizably) false would it be undecidable. Therefore, any sentence quantifying - either universally or existentially - over an infinite domain is asymmetrically undecidable.

We have seen that the manifestation argument offers undecidable sentences as posing *prima facie* difficulties for semantic realism. Would asymmetrically undecidable sentences cause similar problems? Curiously, (20) will be genuinely decidable if its truth-conditions recognizably fail to obtain and will be genuinely undecidable if its truth-conditions unrecognizably obtain. If (20) turns out to be decidable, then it can cause no problem for the semantic realist. Thus, (20) is only problematic if it turns out to be undecidable. Now the anti-realist is only in a position to present a sentence like (20) as problematic for the realist if she is able to present it as *being* genuinely undecidable. The anti-realist would be warranted in presenting it as genuinely undecidable *only if* she were warranted in supposing that its truth-conditions unrecognizably obtain. One is warranted in supposing that a set of truth-conditions obtain *only if* it is possible to recognize them as obtaining - but no one can have the capacity to recognize that a set of unrecognizable truth-conditions obtain! In other words, only if a sentence is *decidably*

undecidable can it be offered as problematic for the realist. Quite simply, asymmetrically undecidable sentences are not decidably undecidable.

Can the anti-realist acceptably eliminate the asymmetric nature of such statements? As argued above, the acceptability of a sentence like (20) as undecidable depends upon its respective (b) clause being warranted. This requires, on the anti-realist's own account, that (b) be decidable. What allows the anti-realist to make this claim? Without it the anti-realist cannot assert that (20) is undecidable.

Let us step back a bit and more closely consider why this is necessitated for the anti-realist. There are three possible epistemic attitudes one can take towards (b): either (i) it is warranted; (ii) it is not warranted; (iii) it is neither warranted nor not warranted. I suggest that the anti-realist cannot take any of these attitudes towards (b). Let me start with the second. Clearly, on the anti-realist's own grounds, if (b) is unwarranted, then (20) cannot be asserted to be undecidable. If it cannot be asserted to be undecidable, it cannot be presented as problematic for the realist.

Consider the third epistemic attitude - (b) is neither warranted nor not warranted. The anti-realist cannot adopt this attitude on pain of contradiction (letting "W(S)" abbreviate "the assertion of 'S' is warranted"):

a) $\neg(W(b) \vee \neg W(b))$	assumption
b) $\neg W(b) \wedge \neg\neg W(b)$	a - DM
c) $\neg W(b)$	b - $\wedge$ elim.
d) $W(b) \vee \neg W(b)$	c - $\vee$ intro.[1]

Line (d) contradicts the assumption. In other words, as under the assumption that (b)

---

[1]This same argument form will come back to haunt the anti-realist in the next section.

is neither warranted nor unwarranted a contradiction ensues, the anti-realist cannot be permitted it; in other words, (iii) is not open to her. Hence, in dealing with the phenomenon of asymmetrical undecidability, it is utterly crucial that the anti-realist make a case for taking attitude (i) towards (b).

The anti-realist might attempt three initial arguments for taking (b) to be warranted. First of all, she might argue that we are warranted in asserting (b) on the grounds that we are not in possession of evidence sufficient for asserting the negation of (20). However, the failure to have evidence *now* does not entail that there can be no evidence. Secondly, she might argue that, because we are obviously not warranted in asserting that it is possible to determine that it is not the case that no such city will ever be built (i.e. (b) with the external negation removed), we are by that fact warranted to assert its negation (i.e. (b) itself). However, such reasoning is in violation of the anti-realist's own intuitionistic principles - lack of evidence for some statement S is not sufficient to establish ¬S.[1] Finally, if one assumed that no such city will ever be built, one would certainly be warranted in accepting that it is not possible to establish that it is not the case that there will never be such a city built. Obviously, though, that mere assumption would not carry sufficient weight to make (b) warranted; only *establishing* that no such city will ever be built would do the job. Establishing that, however, just is to show that statement (20) is decidable, contrary to what the anti-realist is trying to show. I can see no anti-realistically acceptable way to establish (b), and hence no anti-realistically acceptable way to establish (20) - or any sentence quantifying over an

[1]Indeed, even the realist would reject such an inference.

infinite domain - as genuinely undecidable. Let us now move on to Dummett's third proposed source of undecidability.

As mentioned, a typical realist response to the question of the decidability of certain past-tense statements, such as (17) asserting Caesar's crossing of the Rubicon, is to allow past persons as admissible substitutions into one's account of decidability. One species of anti-realism will allow only extant persons as admissible. Such a position seems unreasonable, and would quickly collapse into some form of actualism. A more moderate anti-realism - more in line with ARD from §3.1.2.1 - would allow extinct (but actual-at-one-time) persons as admissible substitutions. Caesar himself, for example, was relevantly situated, and would have been able to determine (17)'s truth-value epistemically. According to this type of anti-realism, (17) would come out decidable and pose no special problem for the semantic realist.

However, we considered the special problem posed when there are not even extant persons available to ensure a past-tensed sentence's decidability - for example sentence (13) asserting the whiteness of the first swan. Even our moderate anti-realist would dismiss (13) as undecidable. Again, the realist response is to allow possible but non-actual (i.e. neither extant nor extinct) persons as admissible substitutions. As a matter of fact, we can suppose, there *was* no one available to observe the colour of the first swan, but there is no inconsistency in supposing that there *could have* been someone. The realist, then, will maintain that as it is *possible* for someone to determine (13)'s truth-value, it is decidable (even though *we* cannot *now* determine it). The move

is to ground the decidability of sentences like (13) in the truth of associated subjunctive conditionals like (19) (i.e. "If someone were to be suitably placed, then they would be capable of determining S's truth-value.").

The anti-realist, of course, will not accept this move. Dummett asserts the following principle:

C) If a sentence is true, there is something in virtue of which it is true.[1]

which he links up with decidability in the following way: A sentence is decidable just in case we can determine its truth-value, and it has a truth-value just in case there is something in virtue of which it is true. Thus, it is a condition of a sentence's decidability that there is something which would, if we knew of it, ground its truth.[2] Suppose that the opposition to the decidability of, say, (13) involves merely the availability of persons in a position to determine its truth-value - i.e. it is accepted that there *is* something in virtue of which we would, if we knew of it, accept as grounds for its truth (or falsity).[3] Under this assumption the realist suggestion of allowing possible persons as admissible seems quite reasonable - if there *is* an observable state-of-affairs involving the colour of the first swan, then a person (actual or possible) suitably placed should be able to recognize it. In other words, if the purported state-of-affairs (implicitly mentioned in

---

[1]Dummett (1976b) p. 89. C. Wright (1987) Ch. 5 suggests a stronger reading in which the quantifier is to be understood only in the present tense. Thus, all past (or future) tense statements, if true, are true in virtue only of present states of information. One price to pay, he admits, is a loss of diachronic inconsistencies (which is, he thinks, too high). (pp. 192-194).

[2]See Dummett (1976b) p. 89.

[3]Namely the state-of-affairs of the first swan being (or failing to be) white.

(C)) could be presumed to either hold or fail to hold, then the actual observation of events in those state-of-affairs would not seem to be required for the decidability of their associated sentences.

The anti-realist response would of course be that we cannot presume excluded middle to hold for such purported states-of-affairs - we cannot assume that there is/was either a state-of-affairs of there being a first swan whose colour was white or a state-of-affairs of there being a first swan whose colour was non-white. Given this response, the real problem surrounding (13) cannot merely be the unavailability of suitably situated persons, but must rather be the potential failure of excluded middle. If excluded middle fails for such purported state-of-affairs, then of course *no* person, actual or possible, could be suitably situated.

This response involves a particular ambiguity. As mentioned, *if* we can assume excluded middle - e.g. that either the first swan was white or that it was not white - then it is not unreasonable for the realist to maintain that a suitably situated possible person would be able to recognize which state-of-affair obtained, rendering, in accordance with RD, (13) decidable. But, the anti-realist will retort that we cannot assume excluded middle, and subsequently cannot assert (13)'s decidability, even allowing possible persons as acceptable substitutions in ARD. On the other hand, *if* we can assume that it is possible suitably to situate a person such that they will be able to recognize whether the state-of-affairs of there being a first swan whose colour was white obtains (or fails to obtain), then it is not unreasonable for the realist to maintain that excluded middle holds for that state-of-affairs. But, the anti-realist will retort that we cannot assume that it is

possible to so suitably situate someone. Notice the *Euthyphro* contrast:[1]

> a) We cannot assume that a possible person can be suitably situated in the relevant state-of-affairs *because* we cannot assume that there is a relevant state-of-affairs.
>
> b) We cannot assume that there is a relevant state-of-affairs *because* we cannot assume that a possible person can be suitably situated in the supposed state-of-affairs.

Which is the anti-realist argument? Consider (a). The anti-realist *assumes* the (potential) failure of excluded middle in order to deny that there is something which would, if we knew it, ground (13)'s truth, and hence to assert its undecidability. But, to assume failure of excluded middle for the relevant state-of-affairs concerning (13) *just is* to maintain that (13) is undecidable. Thus, route (a) assumes the undecidability of (13) in order to establish its undecidability. Consider (b). The anti-realist *assumes* that no one can be suitably situated in order to establish the failure of excluded middle concerning the relevant state-of-affairs. But to assume that no one can be suitably situated is to assume that there is nothing which would, if we knew of it, ground the truth of (13), and that, as we have seen, is tantamount to assuming its undecidability. In either case, then, the anti-realist argument is question-begging.

Let me try to make this clearer. The realist allows possible but non-actual persons as admissible substitutions into an acceptable account of decidability. This will render at least many (if not all) past-tensed sentences decidable. For example, (13) will be decidable just in case it is possible for a person P, suitably situated with respect to an appropriate state-of-affairs A, to determine (13)'s truth-value (which is what (19)

---

[1]The expression is borrowed from C. Wright (1992).

asserts - i.e. (13) is decidable just in case (19) is true). In addition, it is possible to so situate someone just in case A exists. We can assume that A exists just in case we can assume that (13)'s truth-conditions either obtain or fail to obtain (i.e. if we can assume excluded middle for its truth-conditions). An extreme (and unreasonable) anti-realist response is to restrict admissible personal substitutions in an acceptable account of decidability to only extant persons. A more moderate anti-realist response is to allow possible persons as acceptable personal substitutions, but to question whether there are any such possible persons by questioning whether A exists (i.e. questioning whether excluded middle holds for (13)'s truth-conditions). In other words, it will question whether, even if we allow possible persons as substitutions, we can assume that the antecedent of the associated subjunctive conditional will be satisfiable.

But, what considerations can be brought to bear for questioning whether excluded middle holds in such cases? Clearly, (13)'s undecidability cannot be presented as a reason for assuming the failure of excluded middle, for then its undecidability would be given as evidence for its undecidability. On the other hand, it might be maintained that we can assume neither that excluded middle holds for it nor that it fails for it. If we cannot assume that it holds, then we cannot assume that there is a state-of-affairs in which P can be suitably situated. As such, we would have no special reason for thinking that the antecedent of (13)'s associated subjunctive conditional is satisfiable (nor consequently that (13) is decidable). But this does not *quite* provide an argument for (13)'s undecidability. The realist can run a parallel argument: if we cannot assume that excluded middle fails for it, we cannot assume that there is no state-of-affairs in which

P can be suitably situated. As such, we would have no special reason for thinking that the antecedent of (13)'s associated subjunctive conditional is not satisfiable (nor consequently that (13) is undecidable). Granted, this does not *quite* provide an argument for (13)'s decidability either. At best it seems we are stalemated: maybe (13) is decidable, maybe it is undecidable - we just do not know.[1]

The avoid this debacle the anti-realist must, it seems, provide independent grounds for supposing the antecedents of the associated subjunctive conditionals are unsatisfiable. Unless this is done, the anti-realist would have failed to provide a sufficiently strong argument for the existence of such undecidable sentences, and thus their general arguments against semantic realism would fail. Are there any such subjunctive conditionals whose antecedents are known to be unsatisfiable? Yes: subjunctive conditionals whose antecedents are known to be contrary-to-fact - i.e. counterfactuals conditionals:

> The most obvious [violation of principle (C)] is provided by the counterfactual conditional alleged to be true even though there is nothing which, if we knew of it, we should accept as a ground for its truth...[2]

Suppose Jones is dead and never while alive faced danger, and consider the sentence:

21) Jones was brave.

Ex hypothesi it would appear that there is no such state-of-affairs of Jones facing danger

---

[1]Of course, this is already a major concession to the realist - the truth or falsity of sentence expressing (13)'s decidability would transcend our knowledge.

[2]Dummett (1976b) p. 89.

and thus there is no direct method of (21)'s verification or falsification. The anti-realist takes the fact that the antecedent of the counterfactual:

22) If Jones had (contrary to fact) faced danger, he would have acted bravely.

is unsatisfied as sufficient to reject the decidability of (21). Notice that the problem does not directly involve the availability of persons in a position to determine its truth-value; rather it involves the availability of states-of-affairs in which a suitably situated person could determine truth-value. As such, the anti-realist advocating this argument need have no complaint against the realist's bid to allow possible persons as substitution into the account of decidability; if there is no state-of-affairs in which to carry out the test, there is not even a possible person which could be appropriately situated.

There is, however, a *possible* situation, in which a person could be so situated; namely the possible - but non-actual - situation of Jones facing danger (while alive). One might argue that if, contrary to fact, Jones *had* been placed in such a situation, then he *would have* acted bravely, and thus one suitably situated in that possible situation would have the capacity to recognize the truth of (21).[1] In other words, the realist might argue that if we allow possible situations as well as possible persons as substitution, then the antecedents of such counterfactuals as (22) will be satisfiable, rendering their associated categorical sentences decidable. The realist will advocate, then, analyzing the decidability of such statements in terms of:

> 23) If it is possible suitably to situate a person P, then P would be capable of determining S's truth-value.

---

[1]Alternatively, *had* Jones been placed in that situation and not acted bravely, then one suitably placed would have the capacity to recognize its falsity.

Note that this manoeuvre is exactly analogous to the one invoked by the realist to avoid the problem posed by ordinary past-tense empirical statements; the first allows possible (but non-actual) persons to be situated in actual states-of-affairs while the second allows actual (or possible) persons to be situated so as to observe possible (but non-actual) states-of-affairs. Similarly just as one anti-realist stance towards the proposed realist solution to the past-tense problem will be to disallow possible persons as substitutions, so too will one anti-realist stance towards the proposed realist solution to the counterfactual problem be to disallow possible states-of-affairs as substitutions. But, on what basis could such a rejection be made?

There are, admittedly, substantial problems understanding the counterfactual conditional in general. These are not, however, necessarily *special* problems for the realist. There are a number of accounts on the market, notably either Lewis' or Stalnaker's, of which the realist could avail herself.[1] According to such 'possible world' accounts, (22) would be true (presently and in the actual world) just in case the *nearest* possible world (i.e. the possible world minimally differing from the actual world) in which Jones faces danger is also a world in which Jones acts bravely. Anyone suitably placed in that possible world should have the capacity to recognize whether Jones acts bravely or otherwise, and hence should have the capacity to determine (21)'s truth-value.[2] To combat this move, the anti-realist would have to offer an extended and

---

[1]Lewis (1973a) and Stalnaker (1968).

[2]There are less formal characterizations of possible worlds. Vision (1988) §7.9, for example, thinks of a possible world as a product of imagination, and our access to them is via our powers of imagination: "The [observers of possible worlds] are not superhuman

substantial rejection of any such possible world semantics of the counterfactual.[1] Dummett's counter-argument does not involve such a rejection, but rather questions whether such a possible worlds account will help the realist.

Essentially the realist locates the decidability of such categorical statements as (21) in the truth of such associated counterfactuals as (22). Dummett's main objection is that we are not in a position to regard such counterfactuals as (22) true and consequently cannot derive an argument for supposing such sentences as (21) decidable. That objection is two-pronged: (i) we have no reason to suppose that such associated counterfactuals are true, and (ii) we have reason to suppose that bivalence potentially fails for them (i.e. have reason to suppose that they are *not* true (but not that they are *false*)).

Recall Principle (C): if a sentence is true, there must be something in virtue of which it is true. Dummett has argued that categorical statements attributing dispositional properties to objects are, if true, true in virtue of the truth of their associated subjunctive conditionals. Principle (C) is thus a *reductionistic* thesis. However, he distinguishes two types of reductionism: (i) *strong* reductionism, which asserts that the *meaning* of a sentence in a disputed class reduces to or is given by the

---

observers, but beings with our powers; and we need only imagine that they observe the antecedent [of a counterfactual] fulfilled. They are then in a position to observe whether the consequent is fulfilled. What could be more ordinary and commonplace? If this doesn't provide a picture of what these powers consist in, it is difficult to see what Dummett could be requiring." (p. 213).

[1]But even this would be insufficient, for there may be *other* non-possible-world accounts, which would allow for a semantic realist interpretation of the counterfactual. The anti-realist would have to show that *no* acceptable account could be given.

meaning of sentences in its reducing class[1]; and (ii) a weaker reductionism, in which only the *truth-conditions* of a sentence in a disputed class reduces to or is given by the truth-conditions of sentences in its reducing class:

> The thesis that statements of a class M are reducible, in this sense, to statements of another class R takes the general form of saying that, for any statement A in M, there is some family $\bar{A}$ of sets of statements of R such that, for A to be true, it is necessary and sufficient that all the statements in some set belonging to $\bar{A}$ be true; a translation is guaranteed only if $\bar{A}$ itself, and all the sets it contains, are finite. In such a case we may say that any statement of M, if true, must be true in virtue of the truth of certain, possibly infinitely many, statements in R.[2]

With this notion of reductionism, Dummett is able to make a distinction between sentences which are said to be *barely true* and sentences which are not barely true. The 'bare truth' of a statement in a particular class is characterized in terms of the absence of an acceptable reducing class (i.e. a class not containing the sentence nor trivial variants of it); i.e. the analysis of its truth (that in virtue of which its truth consists) does not involve mention of a reducing class. Dummett seems to imply that a barely true statement is true in virtue of the obtaining of some state-of-affairs. On the other hand, a sentence whose analysis of its truth involves mention of a reducing class is a sentence which is not barely true. Categorical sentences attributing dispositional properties to objects are not, then, barely true. In terms of Dummett's basic constraint on a theory of meaning, knowledge of the meaning of a non-barely true sentence consists in knowledge of the meaning of some set of barely true sentences in its reducing class,

---

[1]In this context, the disputed class consists of categorical statements attributing dispositional properties to objects and the reducing class consists of the relevant subjunctive conditionals.

[2]Dummett (1976b) p. 94. See also Dummett (1969).

while knowledge of the meaning of barely true sentences consists in a capacity to use the sentence to give a report of observation:

> [If] someone is able to tell, by looking, that one tree is taller than another, then he knows what it is for a tree to be taller than another tree, and hence knows the condition that must be satisfied for the sentence 'This tree is taller than that one', to be true.[1]

This in turn entails that the state-of-affairs in virtue of which the barely true statement is true must be a recognizable one - otherwise knowledge of its meaning could not be manifested by a capacity to give a report of observation.

The core of Dummett's argument is that counterfactual conditionals cannot be barely true. If he is correct, then their truth must reduce to the truth of sentences in some class of categorical (i.e. non-counterfactual) sentences. But then the realist attempt to locate the decidability of sentences like (21) in the truth of associated counterfactuals like (22) will be unsuccessful (or at least incomplete), for the truth of the associated counterfactuals themselves must be located elsewhere.

Dummett offers three arguments against counterfactuals being barely true. First of all, the assumption that they are is self-refuting. If they are barely true, then they are true in virtue of the obtaining of some non-linguistic state-of-affairs. But, being contrary-to-fact, there are no such state-of-affairs in virtue of which they can be true. However, the realist, invoking a possible world semantics in the style of Lewis or Stalnaker[2], would surely stress that there *are* states-of-affairs in virtue of which such

---

[1]Dummett (1976b) p. 95.

[2]Especially Lewis, who reifies possible worlds.

counterfactuals could be barely true. For example, they will maintain that sentence (22) is true in virtue of states-of-affairs in the nearest possible world in which Jones faces danger. The fact that the sentence is true in virtue of states-of-affairs in some world other than the actual world is besides the point - that fact only makes (22) a *counterfactual*, not non-truth-valued.[1] Thus, Dummett's argument depends upon a prior rejection of that realist manoeuvre.

Secondly, Dummett argues that the assumption of the bare truth of counterfactuals is in tension with his basic constraint on the theory of meaning. As mentioned, according to Dummett knowledge of the meaning of barely true statements consists in a capacity to use the sentence to give a report of observation. This in turn entails that the states-of-affairs in question must be recognizable, otherwise one could not give an observational report of them. But, possible worlds other than the actual one clearly are observationally isolated from us - we quite simply do not have a capacity to recognize which states-of-affairs obtain in them:

> It is precisely for this reason that the thesis that counterfactuals cannot be barely true is so compelling, since we cannot form any conception of what a faculty for direct recognition of counterfactual reality would be like.[2]

Dummett is simply confused here. What we would need is a conception of a faculty for directly recognizing 'reality' *within* a possible world, *not* a conception of a faculty for

---

[1]Analogously, they will argue that past tense statements can be barely true in virtue of non-actual (but past) states-of-affairs; e.g. 'This tree was, at t_n, taller than that tree' may *now* be true in virtue of the state-of-affairs of this tree being taller than that tree obtaining at t_n.

[2]Dummett (1976b) p. 100.

directly recognizing 'reality' in one possible world while being located in another. Consider the following two cases. In order for *me* to understand *your* utterance "I am in pain", I need a conception of what it would be like *to be you feeling your pain*, not what it would be like for *me* to feel *your* pain. I have such a conception - it is exactly analogous to the conception I have of myself being in pain; it is with you as it is with me when I am in pain. Similarly, the realist has proposed that what I need to understand the sentence 'Caesar crossed the Rubicon' is a conception of *being at that event and witnessing it*, not a conception of witnessing that event from my current 20th century placement. Again I have such a conception - it is exactly analogous to the conception I have of observing present-tensed events. Returning to the counterfactual case, I have a conception of what it would be like to directly observe some state-of-affairs in a possible but non-actual world - it is precisely analogous to the conception I have of directly observing states-of-affairs in the actual world.

His third argument rests on the claim that the only reason one might have for supposing counterfactuals to be barely true is a prior assumption that bivalence holds for them:

> Why should anyone think that a counterfactual may be barely true? His only possible ground can be that he supposes it to be a matter of logical necessity that either that counterfactual or its opposite should be true...[1]

If (i) there are no other reasons for supposing counterfactuals to be barely true, and (ii) the assumption of bivalence for counterfactuals is unwarranted, then the claim that counterfactuals may be barely true would be unwarranted. On the other hand, if

---

[1]Dummett (1976b) p. 90.

there is reason for supposing counterfactuals to be barely true which does not rest on an assumption of bivalence, or if the assumption of bivalence for counterfactuals is reasonable, then this third argument against regarding counterfactuals as barely true will fail.

Peacocke offers two considerations which cast doubt on Dummett's claim (i).[1] Consider the sentence:

> 24) If this rock had been composed of mass m and had force f applied to it, then it would have accelerated at such-and-such a rate.

Most of us, anti-realists included, would be inclined to accept (24) as true. But, in virtue of what is it true? Normally we would say that it is true in virtue of the general physical law F=MA, but what is F=MA true in virtue of? A natural inclination would be to say that it is true in virtue of all past, present, and future - as well as possible but unactualized - states-of-affairs of accelerating massy objects. But that is only to say that it is true in virtue of all of the singular factual *and counterfactual* instances of it. Now an anti-realist might retort that this answer brings with it all of the traditional problems of induction, but as there is no generally accepted answer to the problem of induction, Peacocke rightly observes that "...the objection seems to me very serious, and until we are confident that there is some adequate answer, it ought not to be taken as at all obvious that counterfactuals cannot be barely true."[2]

Secondly, Peacocke distinguishes two distinct senses of 'true in virtue of' as used

---

[1]Peacocke (1980).

[2]Peacocke (1980) p. 63.

in Dummett's Principle (C): (i) A sentence S is true in virtue of$_1$ the truth of sentence S′; and (ii) a sentence S is true in virtue of$_2$ the obtaining of a particular state-of-affairs. 'In virtue of$_2$' seems to be the sense involved in the application of Principle (C) to barely true sentences. Consider the statement:

25) It rained in the past.

Ordinarily we would consider (25) to *now* be true in virtue of$_2$ a *past* state-of-affairs.[1] The claim is that there need not be any *current* state-of-affairs in virtue of which a past-tensed statement is barely true, there just needs to be a *past* state-of-affairs in virtue of which it is true. The analogue of this for counterfactuals is that there need not be any *actual* state-of-affairs in virtue of which a counterfactual is barely true, only *possible* state-of-affairs in virtue of which it is barely true. There seems to be no reason for denying that there are such possible states-of-affairs, and thus taking (at least some) counterfactuals to be barely true is neither obviously unreasonable nor rests solely on the assumption of bivalence.[2]

Nonetheless, Dummett casts doubt on the assumption of bivalence for counterfactuals. He characterizes that assumption in terms of it being "...a matter of logical necessity that either [a] counterfactual or its opposite should be true...", where "...the opposite of a conditional [is] that conditional which has the same antecedent and

---

[1]Dummett seems to accept this; see (1969) p. 363.

[2]It is not open to Dummett to remark that only the assumption of bivalence grounds the existence of possible state-of-affairs in virtue of which counterfactuals are barely true. As long as such purported state-of-affairs are not inconsistent, they are admissible.

the contradictory consequent...".[1] Dummett's claim, then, is that the assumption of bivalence for counterfactuals is tantamount to the assumption of the unrestricted application of:

26) $(P \mathrel{\square\!\!\rightarrow} Q) \vee (P \mathrel{\square\!\!\rightarrow} \neg Q)$

Dummett offers several arguments as to why we are not warranted to assume (26) (and thus, according to his first argument, have no ground for supposing counterfactuals to be barely true). He begins by diagnosing why one might be inclined to accept (26). We have a tendency (incorrectly, on Dummett's view) to assume unrestricted application of bivalence for most ordinary categorical statements. If we combine that tendency with the insight that some categorical statements - such as those attributing dispositional properties to objects - are reducible to associated counterfactuals, then naturally the tendency to assume unrestricted bivalence will carry over to the counterfactuals themselves:

> If, then, we assume the law of bivalence for the statements of the first kind, we are forced into granting that, for any subjunctive conditional corresponding to such a statement, either it or its opposite must be true.[2]

But, the proper attitude one should take, Dummett points out, is that if there is no

---

[1]Dummett (1976b) p. 90. See also (1991b) pp. 181-182. To avoid confusing the counterfactual conditional with the indicative one, we can borrow Lewis' symbol '$\mathrel{\square\!\!\rightarrow}$'; $P \mathrel{\square\!\!\rightarrow} Q$ is to be read as 'If it had been the case that P, it would have been the case that Q'. (Lewis (1973a) p. 1). In Dummett (1982), he distinguishes *weak* conditional bivalence: $(P \mathrel{\square\!\!\rightarrow} Q) \vee (P \mathrel{\square\!\!\rightarrow} \neg Q)$; from *strong* conditional bivalence: $(P \mathrel{\square\!\!\rightarrow} Q) \vee (P \mathrel{\square\!\!\rightarrow} \neg \Box Q)$. However, Dummett does not think the distinction affects his argument: "it makes no difference whether strong bivalence is or is not what 'bivalence' should be taken to mean when applied to subjunctive conditionals." (p. 82).

[2]Dummett, (1976b) p. 90.

reason to assume bivalence for the counterfactual statement (or rather, if there is reason to deny it), then there is no reason to assume (or rather reason to reject) bivalence for the categorical statement. Now, as mentioned, one is intuitionistically warranted to assert a disjunction only when one is warranted to assert at least one of its disjuncts. Consider the disjunction of counterfactuals associated with sentence (21):

> 27) Either if Jones had faced danger then he would have acted bravely or if Jones had faced danger then he would not have acted bravely.

Ex hypothesi we have no reason to assert either disjunct rather than the other and thus have no reason to assert the entire disjunction. Thus, given that (27) is an instance of (26), we are not warranted to assume an unrestricted application of bivalence for counterfactuals.

Dummett concedes that it is always open for the realist to reverse the order of priority and argue that the truth (or falsity) of a counterfactual like (22) in fact reduces to the truth (or falsity) of some categorical statement. (22), they might argue, offers merely a *direct* testing procedure for (21), but there may be other, more *indirect* methods. Suppose, for example, that while bravery is *manifested* by brave actions in the face of danger, that character trait can be strongly correlated with a certain 'psychic mechanism' - if Jones had this 'psychic mechanism', then he was brave, and consequently would have, had he faced danger, acted bravely. Alternatively, if he had not had this 'psychic mechanism', then he was not brave, and consequently would not have, had he faced danger, acted bravely. The realist may then go on to argue that either Jones had this 'psychic mechanism' or did not, and it is 'in virtue of$_2$' excluded middle holding for such mechanisms that (21) is bivalent. Consequently, (22)'s bivalence would be ensured

'in virtue of$_1$' (21)'s.[1]

Dummett offers three responses to this manoeuvre: the first mainly rhetorical and the others more substantial. To begin, he asserts that "only a philosophically naive person would adopt [such a view] of statements about character..."[2] This is certainly far from obvious. While talking of 'psychic mechanisms' may be part of a rudimentary philosophical psychology, talking of 'brain structures', 'genetic encoding', or even 'types of moral education' may be part of a quite sophisticated one, such as that offering a physically or environmentally deterministic account of human behaviour.

On the other hand, while the anti-realist may not be able to reject such account on the basis of its naivety, Dummett argues that they can reject it as question-begging. The realist manoeuvre assumes that excluded middle holds for whatever one takes to be the appropriate substitutions for 'psychic mechanisms', and it is this assumption which serves as that 'in virtue of$_2$' which the categorical statement is bivalent. The assumption of excluded middle, however, is *already* a realist assumption:

> In making such an assumption, we are adopting a realistic attitude towards the property or quantity in question; and it should now be apparent how it is that ... the notion of true which we take as governing our statements determines, via the principle C, how we regard reality as constituted.[3]

Dummett's counter-argument strikes me as correct - if the anti-realist is not allowed to assume the failure of bivalence in order to demonstrate the undecidability of some sentence, then neither can the realist assume bivalence (or what amounts to the

---

[1]Dummett (1963b) pp. 49-50. See also (1976b) p. 91.

[2]Dummett (1963b) p. 150.

[3]Dummett (1976b) p. 93.

same thing in this context - excluded middle) to demonstrate the decidability of some sentence - with this one proviso: excluded middle/bivalence is so entrenched in our ordinary logical practices that the burden of proof lies with the one questioning it rather than the one accepting it. It seems to me that *if* a realist could provide indirect evidence to attribute a character trait to someone other than direct manifestation of it, and *if* no reason can be given for supposing excluded middle to fail concerning that purported evidence, then sentences attributing such traits to persons could innocently be presumed either true or false and hence bivalent. Still, there are too many promissory notes in this realist argument to ensure the decidability of (21) (but still, too many to ensure its undecidability as well).

Dummett's final argument is aimed at questioning the assumption of bivalence. Assume, for the sake of argument, that bravery is manifested through behaviour, but that that behaviour is a direct causal product of some 'psychic mechanism', say a particular brain structure, which may be triggered in appropriate circumstances. Grant also the assumption of excluded middle: for any person P, P either has that brain structure or else fails to have that brain structure. Given that one either is brave or not brave 'in virtue of$_2$' either possessing or failing to possess that brain structure, then, for any person P, the following can be presumed to hold:

28) P is brave or P is not brave.

The realist 'reversal of priority' manoeuvre attempts to establish the bivalent:

29) Either if P had faced danger P would have acted bravely or if P had faced danger P would not have acted bravely.

'in virtue of$_1$' the truth of (28). According to Dummett this is a mistake; *at best* (28)

grounds the acceptability of:

> 30) If P had faced danger, then P would either have acted bravely or would not have acted bravely.

which *does not* entail (29). In other words, Dummett rejects the inference $[P \Box\!\!\rightarrow (Q \vee \neg Q) \vdash (P \Box\!\!\rightarrow Q) \vee (P \Box\!\!\rightarrow \neg Q)]$:

> What is involved here is the passage from a subjunctive conditional of the form:
> $A \Box\!\!\rightarrow (B \vee C)$
> to a disjunction of subjunctive conditionals of the form:
> $(A \Box\!\!\rightarrow B) \vee (A \Box\!\!\rightarrow C)$.
> Where the conditional is interpreted intuitionistically, this transition is, of course, invalid.[1]

He goes on to say that "the transition is not in general valid for the subjunctive conditional of natural language either",[2] and gives the following counterexample:

> For instance, we may safely agree that, if Fidel Castro were to meet President Carter, he would either insult him or speak politely to him; but it might not be determinately true, of either of those things, that he would do it, since it might depend upon some so far unspecified further condition, such as whether the meeting took place in Cuba or outside.[3]

Dummett's discussion is not *quite* on target, for he is discussing the general case of $[P \Box\!\!\rightarrow (Q \vee R) \vdash (P \Box\!\!\rightarrow Q) \vee (P \Box\!\!\rightarrow R)]$ whereas we are interested in the perhaps special case of $[P \Box\!\!\rightarrow (Q \vee \neg Q) \vdash (P \Box\!\!\rightarrow Q) \vee (P \Box\!\!\rightarrow \neg Q)]$. It is not obvious whether his counter-example is genuine regarding the latter - if we agree that if Castro were to meet Carter he would either speak politely to him or not speak politely to him[4], then it should not matter what

---

[1]Dummett (1973c) p. 244.

[2]Dummett (1973c) p. 244.

[3]Dummett (1973c) pp. 244-245.

[4]As opposed to the internal negation of speaking *impolitely* to him.

the external conditions of their meeting may be - either he will speak to him or not (and if he does not, then he is not speaking politely to him and the counter-example fails), and if he does he will either speak politely or otherwise (and again the counter-example fails).

Regardless of whether the counter-example succeeds in the special case, the realist might aim to establish the (Dummettian conception of) bivalence of counterfactuals more directly. Stalnaker, for example, offers a possible worlds semantics for the counterfactual, where the counterfactual $P \Box\!\!\rightarrow Q$ will be true in the actual world just in case Q is true in every suitable possible world. The suitability of a possible world is determined by the counterfactual's antecedent: it must be a world in which the antecedent is true and which is otherwise minimally different from the actual world.[1] Stalnaker's system thus exploits the notion of *comparative similarity* between possible worlds: worlds will be similar to the extent in which they assign the same truth-values to the same sentences. A world w will be minimally different from w′ regarding some counterfactual $P \Box\!\!\rightarrow Q$ just in case the only sentence to which they assign different truth-values is P; P is assigned the value 'false' in w (the actual world - this is why it is a counterfactual) while it is assigned the value 'true' in w′.[2] Stalnaker's view is any world minimally different from the actual world concerning some counterfactual $P \Box\!\!\rightarrow Q$ will be a world in which either Q is true or Q is false, and thus *conditional excluded middle*

---

[1]Stalnaker (1968) and (1980).

[2]This is not quite accurate - the differing truth-values of P across the worlds will necessarily force differing truth-values for other sentences. The main idea, though, is that such differences must be kept to the bare minimum, whatever they turn out to be.

drops out:

CEM) $(P \mathbin{\Box\!\!\rightarrow} Q) \vee (P \mathbin{\Box\!\!\rightarrow} \neg Q)$[1]

Lewis points out a serious problem with (CEM) - it rests on the assumption that, for any counterfactual, there will exist a *single nearest* possible world; i.e. that there can be only one possible world differing from the actual world solely in assigning a different truth-value to the antecedent (and thus, in that world, either the consequent holds or not).[2] Consider the counterfactual:[3]

31) If Bizet and Verdi were compatriots, then Bizet would be Italian.

There are, Lewis points out, *at least* two equally close possible worlds which would render the antecedent of (31) true - namely one in which the borders of France are extended to include parts of Italy and one where the borders of Italy are extended to include parts of France. We cannot say, then, that in the nearest possible world concerning (31), either its consequent is true or false, and thus are unable to generate (CEM) concerning it. Stronger than this: (CEM) will fail for (31); it is *not* the case that one disjunct of:

32) Either if Bizet and Verdi were compatriots then Bizet would be Italian or if

---

[1]Actually, in Stalnaker's system (CEM) drops out from the axiom: (SA) $\Diamond P \rightarrow [\neg(P \mathbin{\Box\!\!\rightarrow} Q) \equiv (P \mathbin{\Box\!\!\rightarrow} \neg Q)]$. I will be discussing the axiom later, but one should note that it coincides with Dummett's view that the negation of a counterfactual can be identified with its opposite.

[2]"Stalnaker's theory depends for its success [on the stronger assumption] that there never are two equally close closest ϕ-worlds to *i*, but rather (if ϕ is true at any world accessible from *i*) there is exactly *one* closest ϕ-world. Otherwise there would be no such thing as *the* closest ϕ-world to *i*..." Lewis (1973a) p. 77. See also Lewis (1973b).

[3]The example is originally taken from Quine (1950).

Bizet and Verdi were compatriots then Bizet would not be Italian.

is true to the exclusion of the other.[1]

The real problem concerning (CEM) lies not with the uniqueness assumption needed to generate it, but rather with whether it even captures the intuitive notion of *bivalence*. Both Stalnaker and Dummett have proposed that the negation of a counterfactual should be identified with its opposite; i.e. the negation of $P\Box\rightarrow Q$ should be expressed as $P\Box\rightarrow\neg Q$. Many philosophers have argued this to be a mistake. Peacocke, for instance, sees an ambiguity in the term 'negation' when applied to the counterfactual. Regarding a counterfactual $P\Box\rightarrow Q$, he distinguishes between its *internal* negation expressed as $P\Box\rightarrow\neg Q$ and its *external* negation expressed as $\neg(P\Box\rightarrow Q)$, and notes that they are not equivalent.[2] He furthermore insists that *only* the external negation can be read as the genuine negation involved in the intuitive notion of bivalence; *given* that the general form of bivalence is expressed as $P\vee\neg P$, the bivalence of counterfactuals should properly be expressed as:

CBV) $(P\Box\rightarrow Q)\vee\neg(P\Box\rightarrow Q)$

and not (CEM) as proposed by either Stalnaker or Dummett.[3] Counter-examples of the

---

[1]Stalnaker attempts to respond by invoking a theory of supervaluations in the manner developed by Van Fraassen (Stalnaker (1980)). The damage, however, has I think been done.

[2]Peacocke (1980) p. 60.

[3]Williamson (1988) distinguishes between *strong bivalence* $((P\Box\rightarrow Q)\vee(P\Box\rightarrow\neg Q))$ and *weak bivalence* $((P\Box\rightarrow Q)\vee\neg(P\Box\rightarrow Q))$, arguing that only weak bivalence is required to resist Dummett's rejection of the decidability of counterfactuals. He furthermore argues, as we shall see, for the weak bivalence of counterfactuals.

sort encountered to (CEM) will not, then, be counter-examples to (CBV).

Williamson offers the following argument against the identification of $\neg(P\Box\!\!\rightarrow Q)$ with $P\Box\!\!\rightarrow\neg Q$:

a)	$\neg[(P\Box\!\!\rightarrow Q)\vee(P\Box\!\!\rightarrow R)]$	assumption
b)	$\neg(P\Box\!\!\rightarrow Q)\wedge\neg(P\Box\!\!\rightarrow R)$	a - DeMorgan's
c)	$(P\Box\!\!\rightarrow\neg Q)\wedge(P\Box\!\!\rightarrow\neg R)$	b - problematic hypothesis
d)	$P\Box\!\!\rightarrow\neg(Q\vee R)$	c[1]
e)	$\neg[P\Box\!\!\rightarrow(Q\vee R)]$	d - problematic hypothesis
f)	$\neg[(P\Box\!\!\rightarrow Q)\vee(P\Box\!\!\rightarrow R)]\rightarrow\neg[P\Box\!\!\rightarrow(Q\vee R)]$	a,e - conditional proof
g)	$[P\Box\!\!\rightarrow(Q\vee R)]\rightarrow[(P\Box\!\!\rightarrow Q)\vee(P\Box\!\!\rightarrow R)]$	f - contraposition

Thus: $P\Box\!\!\rightarrow(Q\vee R)\vdash(P\Box\!\!\rightarrow Q)\vee(P\Box\!\!\rightarrow R)$, which is worrisome as shown by Dummett's Castro meeting Carter counter-example. The only suspect move, Williamson claims, is the original hypothesis identifying the negation of a counterfactual with its opposite. Thus, there is good reason to resist it, and (CBV) rather than (CEM) should be regarded as expressing the genuine bivalence of counterfactuals.

As Dummett assumes (CEM) to express the bivalence of counterfactuals, he has little to say regarding (CBV). However, a Dummettian position would question how (CBV) is to be understood - in particular how are we to manifest our understanding of an externally negated counterfactual (which is not itself a counterfactual) as opposed to an internally negated one (which is itself a counterfactual)? To give a general account of the meaning of externally negated counterfactuals, Lewis introduces the symbol: $P\Diamond\!\!\rightarrow Q$ (which is to be read as "If it had been the case that P, then it *might* have been the case

---

[1]Williamson expects this inference to be uncontroversial; it rests upon the recognition that $\neg Q$ and $\neg R$ entail $\neg(Q\vee R)$ and the elementary assumption: if $Q_1,...,Q_n$ logically entails R, then $P\Box\!\!\rightarrow Q_1,...,P\Box\!\!\rightarrow Q_n$ logically entails $P\Box\!\!\rightarrow R$. (Williamson (1988) pp. 408-412).

that Q") which contrasts with the ordinary 'would' conditional $P\Box\!\!\rightarrow Q$ (which reads "If it had been the case that P, then it *would* have been the case that Q"). Lewis furthermore proposes that they are interdefinable: $P\Diamond\!\!\rightarrow Q =_{df} \neg(P\Box\!\!\rightarrow\neg Q)$ and $P\Box\!\!\rightarrow Q =_{df} \neg(P\Diamond\!\!\rightarrow\neg Q)$. Thus, the negation of a normal 'would' counterfactual is equivalent to $P\Diamond\!\!\rightarrow\neg Q$: "It is not the case that if it had been the case that P then it would have been the case that Q" is tantamount to "If it had been the case that P, it might not have been the case that Q".[1] Understood this way, (CBV) could be read as:

CBV′) $(P\Box\!\!\rightarrow Q)\vee(P\Diamond\!\!\rightarrow\neg Q)$

and the Dummettian worry of how we are to understand (CBV) would appear to be circumvented.

Lewis's paraphrase will also, he maintains, allow (CBV) to avoid the original counter-example to (CEM). The original argument against (CEM) was that, under Stalnaker's system at any rate, sentence (31) would come out both true and false (as there is a nearest possible world in which the antecedent and consequent both hold *and* a nearest possible world in which the antecedent holds while the consequent fails). Consider, however, Lewis's proposed paraphrase of the external negation of (31):

> 33) If Bizet and Verdi had been compatriots, then Bizet might not have been Italian.

As there is apparently no contradiction in asserting both (31) and (33), the fact that there are two equally close possible worlds does not provide an argument against

[1]Lewis (1973a). Williamson (1988) follows Lewis on this point.

(CBV).[1]

On the other hand, Williamson provides an elegant counter to the claim that the bivalence of counterfactuals is unwarranted. Consider the sentences "There is nothing but a gold sphere" (abbreviated "G") and "There is nothing but a silver sphere" (abbreviated "S"). Williamson begins with the intuitive claim that the actual world is neutral towards the counterfactuals "Had there been nothing but a gold or a silver sphere, it would have been gold" and "Had there been nothing but a gold or silver sphere, it would have been silver"; i.e. $(G \vee S) \Box\!\!\rightarrow G$ and $(G \vee S) \Box\!\!\rightarrow S$ respectively. In other words, we have no more reason for thinking that the actual world is such as to make the one over the other true as the reverse; if one is true of the actual world we must also take the other to be true of the actual world. To capture this notion he proposes the principle that, for any compound statement A made up entirely of G, S and the logical operators, A is true (in the actual world) iff A(G/S) is true (in the actual world) where 'A(G/S)' is the sentence formed by replacing all occurrences of 'G' in 'A' with 'S' and simultaneously replacing all occurrences of 'S' in 'A' with 'G'. Given further assumptions: $\neg\Diamond(G \wedge S)$ (call this assumption I) and $\Diamond(G \vee S)$ (call this assumption II); (CEM) seems in a bad way - each of two mutually exclusive states-of-affairs stand or fall together.[2]

---

[1]I admit that Lewis's paraphrase makes me uneasy. If (31) and (33) are both assertible, then it is utterly unclear to me how (33) could be an adequate paraphrase of a sentence expressing the *negation* of (31). At worst, however, we would merely need another account of how to understand the negation of a counterfactual.

[2]This counter-example is structurally identical to the Bizet and Verdi one.

Here is his argument[1]; it rests upon four other assumptions regarding the properties of the counterfactual which he takes to be uncontroversial:

III) If $Q_1,...,Q_n$ logically entails R, then $P\Box\!\!\rightarrow Q_1,...,P\Box\!\!\rightarrow Q_n$ logically entails $P\Box\!\!\rightarrow R$.[2]
IV) If $(P \equiv Q)$, then $((P\Box\!\!\rightarrow R) \equiv (Q\Box\!\!\rightarrow R))$.
V) $P\Box\!\!\rightarrow P$ is a logical truth.
VI) If $(P\Box\!\!\rightarrow Q)$, then if P is possible Q is possible.

	a)	$(G \vee S)\Box\!\!\rightarrow G$	assume
	b)	$(S \vee G)\Box\!\!\rightarrow S$	a - (G/S)
	c)	$(G \vee S)\Box\!\!\rightarrow S$	b - IV
	d)	$(G \vee S)\Box\!\!\rightarrow (G \wedge S)$	a,c - III
	e)	$\Diamond(G \vee S) \rightarrow \Diamond(G \wedge S)$	d - VI
	f)	$\Diamond(G \wedge S)$	e - MP & II
	g)	$\Diamond(G \wedge S) \wedge \neg\Diamond(G \wedge S)$	f - I & $\wedge$ intro
$\therefore$	h)	$\neg[(G \vee S)\Box\!\!\rightarrow G]$	a,g RAA

By similar reasoning, and by the assumption that A is true if and only if A(G/S) is true (and given commutivity of disjunctions), $\neg[(G \vee S)\Box\!\!\rightarrow S]$ can be established. Williamson carries on the argument:

	i)	$(G \vee S)\Box\!\!\rightarrow \neg G$	assume
	j)	$(G \vee S)\Box\!\!\rightarrow (G \vee S)$	V
	k)	$(G \vee S)\Box\!\!\rightarrow S$	i,j - DS
	l)	$(G \vee S)\Box\!\!\rightarrow S \wedge \neg[(G \vee S)\Box\!\!\rightarrow S]$	k - $\wedge$ intro[3]
$\therefore$	m)	$\neg[(G \vee S)\Box\!\!\rightarrow \neg G]$	i,l RAA

And again by similar reasoning, as well as replacing "G" with "S" in (m), $\neg[(G \vee S)\Box\!\!\rightarrow \neg S]$ can be established. Thus, the following are all true:

i) $\neg[(G \vee S)\Box\!\!\rightarrow G]$
ii) $\neg[(G \vee S)\Box\!\!\rightarrow S]$
iii) $\neg[(G \vee S)\Box\!\!\rightarrow \neg G)]$
iv) $\neg[(G \vee S)\Box\!\!\rightarrow \neg S)]$

---

[1]I have made some slight modifications to it to make it clearer.

[2]Recall that this assumption was used in his rejection of $[(P\Box\!\!\rightarrow \neg Q) \equiv \neg(P\Box\!\!\rightarrow Q)]$.

[3]Recall that $\neg[(G \vee S)\Box\!\!\rightarrow S]$ has already been established.

and hence bivalent. Not only does this yield strong reason to suppose bivalence is warranted for counterfactuals (at least for counterfactuals of this sort - the sort which initially caused trouble for their apparent bivalence), it also provides strong reasons for rejecting Stalnaker's and Dummett's identification of the negation of a counterfactual with its opposite. If $\neg(P \boxrightarrow Q)$ were equivalent to $P \boxrightarrow \neg Q$, then given the derived truth of both (i) and (iii), $(G \vee S) \boxrightarrow G$ and its purported negation $(G \vee S) \boxrightarrow \neg G$ would both come out true.

So, where does all this leave us? As we saw, Dummett's case for the undecidability of past-tense statements rested on his case for the undecidability of subjunctive conditionals. The realist can make a strong case for the decidability of subjunctive conditionals by extending the admissibility of persons whose capacities would determine a sentence's truth-value to possible but non-actual persons. No compelling anti-realist objections were raised against such an extension.

However, it was objected that the proposed extension does no good unless one is warranted to assume that such persons can be suitably placed, and unless one assumes a universal applicability of excluded middle, the realist cannot guarantee situations in which to suitably place her non-actual persons, and thus cannot guarantee the decidability of the sentences in question. In response to this, the realist similarly advocated admitting possible but non-actual testing situations into one's acceptable notion of decidability. Again, no compelling anti-realist argument was raised against such an extension.

However, counterfactual conditionals make implicit reference to *impossible* testing

situations, and thus the realist proposed extension would appear to be of no avail for this class of sentences. The realist response is to evoke a 'possible worlds' semantics for counterfactual conditionals, and thus locate suitable testing situations on possible but non-actual worlds. It was admitted that there are some substantial problems with such a semantics, but they are not special problems for the realist (nor, it should be noted in passing, has the anti-realist given any conclusive reasons to suppose that they will not be surmounted). A strong case was made for assuming bivalence and hence derivatively decidability to hold for counterfactual conditionals. All in all, counterfactuals do not present a compelling case for the existence of undecidables.

The anti-realist will not likely be impressed with the previous manoeuvring. They will argue that as long as there exists at least *one* undecidable sentence, a realist semantics is untenable: at best I have shown that Dummett's three candidates fail to generate the manifestation argument. There is no reason, they will continue, to suppose that Dummett's three candidates exhaust the field.

There are, it would seem, other *formally* undecidable sentences, such as that involved in Gödel's Incompleteness Theorem, or the Continuum Hypothesis, or the Liar's Sentence for that matter.[1] Thus, it would appear that the manifestation argument

---

[1]Gödel demonstrated that any consistent formal system powerful enough to express arithmetic must be incomplete by showing that there must exist at least one true arithmetical sentence not provable in the system. (For a somewhat simplified discussion, see Boolos and Jeffrey (1989) Ch. 15). The Continuum Hypothesis asserts that every set of real numbers either is enumerable or has the same cardinal number as the set of all reals (the continuum). Cohen demonstrated that, while intuitively true, it cannot be derived from the axioms of set theory. (Boolos and Jeffrey (1989) p. 212). Tarski

goes through, despite my efforts to stave it off.

However, it is worthwhile to reflect on the significant difference between these formally undecidable sentences and the original candidates presented Dummett. They are all undecidable by purely formal results - i.e. there is something about the various systems in which they are expressed which precludes them from being proved *within the system*. For example, Tarski's solution to the paradox of the Liar was to distinguish meta-language from object-language, and maintain that the truth-predicate of, say, the object-language, could only be defined in its meta-language. In other words, what distinguishes the formally undecidable sentences from Dummett's candidates is that the latter were deemed undecidable because of epistemological shortcomings *in us* - in our evidence-gathering powers - while the former are deemed undecidable because of the

---

demonstrated that no formal system can include its own truth-predicate; for example, any 'proof' of the Liar's Sentence ("This sentence is false") would yield its falsity, and any 'disproof' of it would yield its truth. Thus we cannot, on pain of contradiction, determine the truth-value of the Liar's Sentence. (Tarski (1944)).

Nor do we have to go to mathematics or the semantic paradoxes to find such examples. Edgington (1985) points out that any sentence of the form "S and no one at any time has any evidence that S" is logically impossible to verify, as any evidence which would verify one conjunct would falsify the other. She uses the sentence as a counter-example to Appiah's (1986) claim that there exist no sentences which are logically impossible to verify. However, Edgington's sentence is, I take it, an informal expression of the Gödel sentence (realists will in fact take it to indicate that there are sentences whose truth transcend their verification) and does not contest Appiah's more specific claim that there are no sentences *of the type Dummett presents* (e.g. concerning the past) which are logically impossible to verify.

On an aside, C. Wright (1987) distinguishes between sentences whose content guarantees their undecidability and those which do not. Examples of the latter include Dummett's candidates, and examples of the former include a sentence involving a claim of a reversed spectrum and "Everything is uniformly increasing in size". His view is that most would resist a realist interpretation of the former without feeling an urge towards a full-blown anti-realism.

(interesting and often surprising) logical properties of the systems in which they are expressed. Not even a "hypothetical being whose intellectual capacities and powers of observation may exceed our own" would be able to prove the Gödel sentence within the formal system.

The manifestation argument is intended to combat the realist construal of truth - i.e. the view that the truth-value of a sentence may transcend our capacities to determine it. After all, the core of the anti-realist argument is that the realist construal of truth is incompatible with an adequate theory of meaning - i.e. with facts about how we actually *acquire* and *use* our language. Thus, it would seem that only undecidable sentences of the sort sought by Dummett are relevant in that context. That being the case, the existence of formally undecidable sentences poses no *prima facie* problems for a realist semantics. I feel justified, then, in maintaining that, as all candidates for undecidable sentences *of the required type* presented by the anti-realist have been adequately dealt with, the manifestation argument fails to present insurmountable problems for a realist semantics.

On the other hand, even if there is in fact no significant difference between the type of undecidability displayed by the formally undecidable sentences and Dummett's original candidates (which I deny), Dummett himself has given a way out for the realist. As mentioned, Dummett tends to favour carrying on the debate in a local as opposed to a global fashion. In other words, whether one is a realist or an anti-realist concerning one area of discourse depends on whether one accepts a realist or an anti-realist construal of truth for sentences in that discourse. Accepting a realist construal of truth

for, say, sentences in the past tense does not necessitate accepting a realist construal of truth for, say, sentences in the future-tense - one may be a realist concerning the past but an anti-realist concerning the future. At best, formally undecidable sentences suggest one ought to be an anti-realist concerning mathematics (and other formal systems). The usual anchor for the realist, however, is empirical statements concerning some (supposed) mind-independent reality. Mathematical reality might not be mind independent, as the Platonist supposes, but that should not induce one to accept that more ordinary empirical reality - the world in which we eat and sleep and dream dreams - is not mind-independent. Thus, any serious realist should, it seems to me, insist that the manifestation argument present *empirically* undecidable sentences, as opposed to *formally* undecidable one, and that has simply not been done.

On a side note, the Gödel's Incompleteness Theorem is interesting. On the surface, it seems to assert that there exists a statement "expressible in the system but not provable in it, which not only is true but can be recognized by us to be true".[1] In other words, it seems to assert that there exists a true sentence which transcends our capacity to determine its truth-value, and would thus seem to validate a realist construal of truth. Truth, it would seem, must be a non-epistemic notion. However, this is not the main challenge that Dummett sees the theorem presenting - rather he takes it to provide *prima facie* evidence against the identification of meaning with use.[2]

The problem is this - the theorem proves that mathematics must, if consistent, be

---

[1]Dummett (1963a) p. 186.

[2]Dummett (1963a).

incomplete - i.e. there is no formal model which can completely characterize all of mathematics. In particular, says Dummett, the concept of 'natural number' is not one which can be characterized by any formal system. But, what does it mean to say that Gödel's theorem is true? What it must mean, says Dummett, is that we have some intuitive, but non-formally expressible *intended mathematical structure* in mind, and say that it is true *in that structure*. But, given that we cannot completely characterize that structure, there is no guarantee that the structure I have in mind when I consider Gödel's Theorem true is the same as the structure you have in mind when you consider it true. Thus, we may all attach slightly different meanings to 'natural number' (by reference to our potentially different intended mathematical structures) even though we all use the expression in the same way. Thus, meaning would seem to potentially transcend use:

> We all of us have the concept of 'natural number'; but no finite description of our use of arithmetical statements constitutes a full account of our possession of this concept, and this is shown by the fact that we shall always be able, by appeal to our intuitive grasp of the concept, to recognise as true some statement whose truth cannot be derived from that description of the use of such statements.[1]

Dummett's 'solution' to this problem is the same as that to the apparent difficulty for an epistemic notion of truth - deny that any definite sense can be attached to the claim that the Gödel sentence is true. To say that the Gödel sentence is true is to say that it is true in some model M, but:

---

[1]Dummett (1963a) p. 190. He gives the following analogy. We can never be sure that the colour to which we all give the name 'blue' has a common phenomenal 'feel', and thus cannot be sure that we all attach the same meaning to it. Nonetheless, we all use the word 'blue' in the same way. Thus, it would seem to follow that the meaning of an expression may transcend its use. (Dummett (1963a) P. 187).

> There is no way in which we can be 'given' a model save by being given a description of that model. If we cannot be given a complete characterization of a model for number theory, then there is not any other way in which, in the absence of such a complete description, we could nevertheless somehow gain a complete conception of its structure.[1]

In other words, Dummett asserts that the Gödel sentence may be unprovable in the system, but cannot be considered true. Thus, no candidate for a true but unprovable sentence has been provided. On the other hand, the meaning of the Gödel sentence is completely determined by its use. We *may* pretend that we attach different meanings to the sentence by supposing it to be true relative to some intuitive intended model, but that supposition is just an illusion - one of those "errors of thought to which the human mind seems naturally prone".[2]

Whether or not we should find Dummett's response to the Gödel problem convincing, it points to an interesting and powerful general realist response to the anti-realist challenge. Dummett has correctly realized that the Gödel challenge contains three elements: the anti-realist presumption that truth is coextensive with provability; the identification of meaning with use; and the existence of true but undecidable sentences. Thus, Gödel's Theorem can be seen as presenting two distinct challenges: (i) *if* one accepts the identification of meaning with use *and* one accepts the existence of true but undecidable sentences *then* one must reject the anti-realist presumption regarding truth; or (ii) *if* one accepts the anti-realist presumption regarding truth *and* one accepts the

---

[1]Dummett (1963a) p. 191.

[2]Dummett (1963b) p. 374. Note that this response mirrors the one I provided to McGinn in §2.2.1.1.

existence of true but undecidable sentences *then* one must reject the identification of meaning with use. His answer to both is to reject the claim that there are true but undecidable sentences.

The anti-realist challenge to the realist mirrors the Gödel one to the anti-realist; it contains three inconsistent elements as well: the realist presumption that truth may transcend provability; an identification of meaning with use; and the existence of undecidable sentences. Thus, there are at least three ways of resolving the inconsistency: (i) accept the existence of undecidables and the identification of meaning with use and reject the realist presumption regarding truth; (ii) accept the realist presumption regarding truth and the identification of meaning with use and reject the existence of undecidables; or (iii) accept the realist presumption regarding truth and the existence of undecidables and reject the identification of meaning with use.

Either of (ii) or (iii) will allow for realism. My strategy has been to support (ii). Alternatively, the realist might choose to uphold (iii). Such a strategy would not necessarily be *ad hoc* - there are some serious doubts about whether a theory of meaning in Dummett's sense is possible anyway; Dummett himself presents some serious difficulties in attempting to formalize the concept of 'use' to generate a genuine use-theory of meaning.[1]

Be that as it may, my core argument is still to grant Dummett his main premises concerning the harmonization of a theory of meaning (in his sense) with an adequate

---

[1]Dummett (1963a) acknowledges that 'use' may simply be too vague a notion to adequately capture in a formalized theory of meaning.

theory of understanding (in his sense) and to still deny that his conclusion of the rejection of realism follows. To do this, I have argued that no sufficiently strong case has been presented for the existence of undecidables of the type needed to generate the manifestation argument. That argument consisted of meeting the anti-realist's claims one by one, and as such has delivered inconclusive results: the fact that no eligible candidate for an appropriate undecidable sentence has been established is not sufficient to demonstrate that none in fact exist. Thus, a more general argument is needed. In the next section I will argue that no anti-realist, on pain of contradiction, can assert that there exists even a single undecidable sentence.

3.2 The Non-Assertibility of Undecidability

As mentioned, the cornerstone of the manifestation argument is the supposed existence of undecidable sentences. A sentence is undecidable just in case we are not capable of determining its truth-value. According to the epistemically constrained anti-realist conception of truth, truth is a property conferrable on a sentence just in case it is capable of being verified. Thus, the assertion that a particular sentence is undecidable - i.e. neither verifiable nor falsifiable - is tantamount to the assertion that it lacks a truth-value. It would seem, then, that the central plank in the anti-realist's argument is that there exist sentences which are neither true nor false - i.e. sentences for which bivalence fails.

Can we sum this up into a defining thesis for both realism and anti-realism?[1] As Dummett says:

> Realism consists in the belief that for any statement there must be something in virtue of which either it or its negation is true: it is only on the basis of this belief that we can justify the idea that truth and falsity play an essential role in the notion of the meaning of a statement, that the general form of an explanation of meaning is a statement of the truth-conditions.[2]

Thirteen years later he says:

> We may, in fact, characterize realism concerning a given class of statements as the assumption that each statement of that class is determinately true or false.[3]

And more recently:

> Integral to any given version of realism [is] the principle of bivalence for statements of the disputed class.[4]

Thus, according to Dummett, an integral feature of a realist conception of truth (and hence of realism *per se*) is the acceptance of bivalence. On the other hand, from considerations of manifestation and acquisition, unrecognizable truth-conditions can play no role in our understanding of sentences - the only truth-conditions which can play such a role are conditions understood as being epistemically constrained; i.e. recognizable

---

[1]I urge caution here. While I have argued that it is a mistake to suppose that acceptance of bivalence is *constitutive* of realism, nonetheless I have argued for a realism which denies that there are sentences (of the sort needed to generate the manifestation argument) with unrecognizable truth-conditions. If there are no sentences with unrecognizable truth-conditions, then there are no sentences for which bivalence fails. Contrapositively, if there are sentences for which bivalence fails, then there are sentences which have unrecognizable truth-conditions. Thus, in the following, for ease of presentation, I will talk *as if* acceptance of bivalence is constitutive of realism.

[2]Dummett (1959a) p. 14.

[3]Dummett (1976b) p. 93. See also Dummett (1973c) p. 228.

[4]Dummett (1991b) p. 325.

conditions. But once one denies that a set of truth-conditions may obtain independently of our recognitional capacities, then it would seem that one is bound to deny that every sentence is either determinately true or false independently of our capacity to determine which. In other words realism, according to Dummett, is committed to bivalence while anti-realism - centred on a commitment to the existence of undecidable sentences - would seem to be committed to a denial of bivalence.

In §2.1 the semantic principle of bivalence was expressed as:

BV) $(\forall S)$('S' is true or 'S' is false)

Dummett furthermore stresses that, while BV is distinct from the logical law of excluded middle, "once we have lost any reason to assume [bivalence] we have no reason, either, to maintain the law of excluded middle."[1] Thus, the anti-realist rejection of BV will coincide with a rejection of:

LEM) $(\forall S)(S \vee \neg S)$

What is it to reject LEM? On a natural reading it would be to assert that not all sentences are either true or false:

A) $\neg(\forall P)(P \vee \neg P)$

From this and the claim that there exist some sentences which are either true or false (i.e. effectively decidable sentences), it would seem that the anti-realist would be warranted to assert that there exists some sentence which is neither true nor false:

B) $(\exists P)\neg$('P' is true $\vee$ 'P' is false)

---

[1]Dummett (1991b) p. 9. Note that excluded middle can be derived from BV with the assumption of disquotation and the identification of "'S' is false" with "$\neg$S". In the following, I will treat BV and excluded middle as standing or falling together.

which, as I have been arguing, is precisely what the anti-realist needs to support the acquisition and manifestation arguments. Translating (B) into its object-language counterpart yields:

B′) $(\exists P)\neg(P \vee \neg P)$

However, there is a problem with the move from (A) to (B′) - it is not sanctioned intuitionistically.[1] Nonetheless, there are good reasons for supposing that any anti-realist, Dummett in particular, would be content to accept it. By linking up undecidability to the failure of bivalence in the way Dummett does, being committed to the existence of undecidable sentences thereby commits him to instances of (B). He moreover recognizes this fact:

> We thus arrive at the following position. We are entitled to say that a statement P must be either true or false, that there must be something in virtue of which either it is true or it is false, only when P is a statement of such a kind that we could in a finite time bring ourselves into a position in which we were justified either in asserting or in denying P; that is, when P is an effectively decidable statement. This limitation is not trivial: there is an immense range of statements which, like 'Jones was brave', are concealed conditionals, or which, like 'A city will never be built here', contain - explicitly or implicitly - an unlimited generalization, and which therefore fail the test.[2]

---

[1]Tobias Chapman has pointed out that intuitionism need not contain the negation of "P∨¬P" - it merely requires that bivalence is not *provable*. He furthermore suggests that this claim can be made good by adopting a multi-valued logic such as Łukaciewiz's (if "P" is indeterminate then "¬P" would be indeterminate, and consequently so would "P∨¬P"). While I agree that intuitionism *per se* need not accept (A), there are reasons for thinking that Dummettian anti-realism would. In the first place, Dummett clearly favours a two-valued logic (see, in particular, Dummett (1959a) p. 14). In the second place, the manifestation argument turns on the *existence* of sentences for which bivalence fails; i.e. it requires that it can be shown/proven that there are such sentences. Such a demonstration would itself be intuitionistically tantamount to a proof of (A). This latter point will be more fully developed in the main text.

[2]Dummett (1959a) pp. 16-17.

> [W]e shall conclude that it may be the case that the statement 'He was brave', is neither true nor false.[1]

> ...an anti-realist view of statements about the past, the view, namely, that a statement about the past, if true, can be true only in virtue of what is or will be the case, and that therefore there may be statements about the past which are neither true nor false.[2]

Also, it is only under the assumption of (B) that either the manifestation or acquisition arguments pose any problem for semantic realism. For these reasons, then, it is not unreasonable to take claim (B) as the natural anti-realist attitude towards bivalence, and indeed as one of the central theses in their attack on semantic realism.

Now, can an anti-realist correctly assert (B)? It would seem that, on pain of contradiction, he cannot. The inferences used in the following argument are all ones that are intuitionistically acceptable, yet from (B), it can be shown that a contradiction quickly ensues:

a) $(\exists P)\neg(P \vee \neg P)$	B′
b) $\neg(p \vee \neg p)$	a - $\exists$ instan.
c) $\neg p \wedge \neg\neg p$	b - DM
d) $\neg p$	c - $\wedge$ elim.
e) $p \vee \neg p$	d - $\vee$ intro.[3]

Line (e) contradicts line (b), each of which follow directly from the line above it by application of a single accepted rule of inference. Thus, line (a) alone is responsible for the inconsistency. What are the consequences of this argument for anti-realism?

The immediate consequence is that it is contradictory to assert that there exists

---

[1]Dummett (1963b) p. 149.

[2]Dummett (1963b) p. 153.

[3]Prawitz (1980) recognizes a version of this proof.

a sentence for which bivalence fails. Classically, this plays right into the realist's hands - for if it is contradictory to assert that there exists a sentence for which bivalence fails, it would follow that there is no sentence for which bivalence fails, and thus it would hold for every sentence. However, that reasoning relies upon two intuitionistically unacceptable assumptions: (i) that $\neg(\forall x)Fx$ entails $(\exists x)\neg Fx$ (so that the entire argument can be seen to rest on (A) as premise) and (ii) the validity of DNE. The argument can be represented as:

!) $\neg(\forall P)(P \vee \neg P)$	A
a) $(\exists P)\neg(P \vee \neg P)$	! - ass. (i)
⋮	
f) $(p \vee \neg p) \wedge \neg(p \vee \neg p)$	b,e - $\wedge$ intro.
g) $\neg\neg(\forall P)(P \vee \neg P)$	A,f - RAA
h) $(\forall P)(P \vee \neg P)$	g - ass. (ii)/DNE

The move from (!) to (a) is simply not acceptable to the anti-realist, and as such the contradiction derived at (f) cannot sufficiently establish the double negation of BV at (g). Moreover, *even if* that move were acceptable, the move from (g) to (h) is not acceptable to an anti-realist as the purported validity of DNE presupposes a realist notion of truth.

So, it is a consequence of the main argument that it is contradictory to assert that there exists a sentence which is neither true nor false but *not* that an unrestricted applicability of bivalence is warranted - and thus that a realist notion of truth is sanctioned. In other words, the argument can in no way be taken as a vindication of realism.

Nor, it seems, can it be taken as an argument *against* the rejection of bivalence expressed in (A). However, a rejection of bivalence unaccompanied by an assertion of

(B) is impotent as an argument against semantic realism - both the acquisition and manifestation arguments, if they are to have any teeth, require a rejection of bivalence in the form of (B) and not merely in its intuitionistically distinct (A). Therefore, the rejection of bivalence, if it is to characterize the main *anti*-realist thesis, must take the form of (B). Assuming that (B) represents the *rejection* of bivalence, the main argument does show that that rejection is inconsistent.

Faced with this consequence, anti-realists may deny that (B) captures their attitude towards bivalence. But then, what *could* their attitude be? On the one hand, the main argument shows that the rejection of bivalence is inconsistent; on the other hand, as they argue, the acceptance of bivalence is incompatible with any adequate theory of meaning. Therefore, the correct anti-realist *attitude* towards bivalence; i.e. the central thesis of anti-realism; should be that bivalence itself is *undecidable*.[1]

Can an anti-realist then correctly assert their attitude towards bivalence? To do so they would have to make clear what it would mean to *assert* a sentence as undecidable - and this is the stumbling block. From the reasoning leading up to this suggestion, it seems obvious that to assert that some sentence is undecidable would be to assert that neither it nor its negation holds. Specifically, to assert that bivalence is undecidable would be to assert:

(C) $\neg[(\forall P)(P \vee \neg P) \vee \neg(\forall P)(P \vee \neg P)]$

Understanding undecidables in this way will not help - such an assertion is

---

[1]This seems to be Brouwer's attitude: "And it likewise remains uncertain whether the more general mathematical problem: 'Does the *principium tertii exclusi* hold in mathematics without exception?' is solvable." Brouwer (1908) p. 110.

identical in form to premise (b) and hence generates a similar contradiction. In fact, under this interpretation, *any* undecidable sentence would be inconsistent (and hence false). The main argument shows that no proof of the failure of bivalence is possible (i.e. consistent), *including* the failure of the bivalence of bivalence.

McDowell (1976) §2 argues that the proper anti-realist attitude towards bivalence should be a refusal to assert it combined with a refusal to reject it (though he hints, in a footnote, that bivalence may need to be *rejected* for decidably undecidable sentences).[1] Thus, the anti-realist need not countenance counter-examples to BV which generate the contradiction. Secondly, he argues that the intuitionistically sanctioned double negation of BV (which is not, it must be remembered, equivalent to BV) ensures that we need countenance no more than the standard two truth-values without being committed to a realist semantics. However, the objection to the manifestation argument still stands: the anti-realist is unable, on pain of contradiction, to present a genuine sentence of type U, and thus is unentitled to the argument's key premise.

Weir (1986) suggests that we (intuitionistically) interpret the assertion that a sentence is undecidable as an assertion that the sentence is not neither true nor false (i.e. as the double negation of its excluded middle). This, he claims, will appease the realist who regards it as equivalent to expressing its excluded middle as well as the anti-realist who does not regard them as equivalent. However, Weir's suggestion is unacceptable, as under it every decidable sentence would entail its own undecidability:

a) Q assumption

---

[1]See also Rasmussen and Ravnkilde (1982) §2.

b) $Q \vee \neg Q$ a - $\vee$ intro.
c) $\neg\neg(Q \vee \neg Q)$ b - DNI[1]

The moral I am tempted to draw is that the issue of bivalence is (again) a red herring. Dummett is simply mistaken to take one's attitude towards bivalence as *the* chief bone of contention between realism and anti-realism. The real difference is surely this: the one resists while the other insists that there are undecidable sentences. Put in these terms, the bone of contention is over the admissibility of (B) - i.e. the existence of a true sentence which we can neither verify (in principle) nor falsify (in principle). The main argument, as seen, shows that an assertion of (B) is inconsistent, and thus gives strong reason to doubt that the acquisition and manifestation arguments pose any serious problems for semantic realism. In other words, we can carry on the proof to yield a result fatal to the force the acquisition and manifestation arguments have against semantic realism:

f) $(p \vee \neg p) \wedge \neg(p \vee \neg p)$ b,e - $\wedge$ intro.
g′) $\neg(\exists P)\neg(P \vee \neg P)$ a,f - RAA

If the anti-realist is simply not in a position to assert that there exists a sentence which is neither verifiable nor falsifiable - i.e. which is undecidable and hence has unrecognizable truth-conditions - then semantic realism is not in jeopardy.

---

[1](b) and (c) are both intuitionistically sanctioned.

4.0 RESPONSES TO THE POSITIVE PROGRAMME

The negative position of Dummett's anti-realism is that a realist semantics, with its non-epistemic notion of truth, is unable to harmonize with an adequate account of understanding; we could neither, consistently with such an account of understanding, acquire such a central concept nor manifest our understanding of some clearly intelligible expressions. As was argued, the argument from acquisition is parasitic upon the argument from manifestation. Thus, the core of the negative anti-realist attack is that there are certain clearly intelligible sentences for which we could not manifest our understanding *if* understanding were modeled on a realist semantics.

It was argued that, at the very least, the range of such sentences is considerably smaller than the anti-realist would have us suppose, thus undercutting the number of classes of sentences for which we might be tempted to seek an anti-realist interpretation. An argument was raised to the effect that the range of such sentences, *even from an anti-realist perspective*, was empty - on pain of contradiction, there could be no sentence which the anti-realist could present as problematic for the anti-realist. It would do the anti-realist no good to insist that, nonetheless, there *may be* such sentences even though she cannot recognize them as such, for that would be to suggest that truth outstrips recognition.

Thus, there seems to be strong reasons for doubting that a semantic realism is in any serious difficulty. In this section, however, we will ignore the preceding results and assume that the anti-realist has made good on his attack on realism. The question remains, does an anti-realist theory of meaning constitute an acceptable semantics? In

other words, is it the case that, supposing that a realist theory of meaning ultimately fails, only an anti-realist theory of meaning, with its epistemically constrained notion of truth, is capable of harmonizing with an adequate account of understanding?

I think not. In the first place, an anti-realist semantics harmonizes no better with an adequate account of understanding than a realist semantics does - undecidables are just as problematic for the anti-realist as for the realist. Secondly, a realist account can, contrary to Dummett's insistence, meet the manifestability constraint - i.e. an anti-realist theory of meaning is not the *only* such account. Finally, and most importantly, a case can be made for supposing that an anti-realist semantics fails to harmonize with another basic constraint on any adequate theory of meaning - that it account for the compositional nature of language. That deficiency can only be remedied, it will be argued, by admitting a realist notion of truth.

4.1 Manifestability and Undecidability

Recall that according to a realist account, the meaning of a sentence S is given by its truth-conditions, expressed by "'S' is true iff P" where P refers to some state-of-affairs whose obtaining would be both necessary and sufficient for the truth of S (and whose non-obtaining would be both necessary and sufficient for the falsity of S). To understand a sentence S, then, is to understand its associated T-sentence. One understands an associated T-sentence if one is capable of correctly assenting to S when appropriately situated with respect to the state-of-affairs P (or dissenting from S when one is situated so as to recognize that P does not obtain).

This account founders, Dummett argues, over undecidable sentences. An undecidable sentence is one which has only unrecognizable truth-conditions; i.e. one for which we cannot determine its truth-value. Recall one of Dummett's favourite examples - sentence (20): "Jones was brave" where Jones is now dead and never, while alive, faced danger. According to the semantic realist, understanding that sentence consists in grasping something like: "'Jones was brave' is true iff Jones was brave", which we attribute to someone if (and only if) they are such that when favourably situated *viz-a-viz* the state-of-affairs of Jones being brave assent to 'Jones was brave'. However, because the categorical sentence reduces, according to Dummett, to a counterfactual conditional, there is no situation which one could be placed in under which they would recognize that the state-of-affairs of Jones being brave obtains, hence one could not manifest their understanding of the sentence. Either such a sentence is unintelligible - which is unacceptable - or the realist theory of meaning delivers an incorrect account of understanding.

Does semantic anti-realism fare any better in the face of such sentences?[1] Dummett's proposal is that we replace a truth-conditional semantics by a verification-conditional one. On his account, to know the sense of a sentence would be to know

---

[1]Moriconi and Napoli (1988) §4 present an argument similar to the one developed here. Part of their aim, however, is to demonstrate that a realist construal of truth (in terms of recognition-transcendence and bivalence) is not responsible for semantic realism's failing to satisfy manifestation constraints (as they argue semantic anti-realism, which eschews such a construal, similarly fails to satisfy them). Their implication, I take it, is that the issue of the correct construal of truth is a red herring in characterizing the realism/anti-realism debate: "the full manifestation of meaning [makes reference not to] an undecidable predicate (be it truth or assertibility) but to the decidable relation 'construction c is a proof of the statement S'." (p. 378).

under what conditions an assertion of it would be verified. How are we to understand verification-conditions? It would seem that we should think of them along the lines of truth-conditions, replacing the T-schema with a V-schema:

(V) The assertion 'S' is verifiable iff C.

How are we to understand C in this case? Dummett asserts that:

> The meaning of a logical operator is given by specifying what is to count as a proof of a mathematical statement in which it is the principal operator, where it is taken as already known what counts as a proof of any of the constituent sentences (any of the instances, where the operator is a quantifier)...
> The intuitionistic explanations of the logical constants provide a prototype for a theory of meaning in which truth and falsity are not the central notions. The fundamental idea is that a grasp of the meaning of a mathematical sentence [consists in] an ability to recognize, for any mathematical construction, whether or not it constitutes a proof of the statement; an assertion of such a statement is to be construed, not as a claim that it is true, but as a claim that a proof of it exists or can be constructed.[1]

He then expands that notion to cover non-mathematical sentences:

> Such a theory of meaning generalizes readily to the non-mathematical case. Proof is the sole means which exists in mathematics for establishing a statement as true: the required general notion is, therefore, that of verification. On this account, an understanding of a statement consists in a capacity to recognize whatever is counted as verifying it, i.e. as conclusively establishing it as true.[2]

In other words, Dummett maintains that we ought to understand C in terms of an effective procedure which, if carried out, would warrant the assertion. We can thus replace (V) with:

(V′) The assertion 'S' is verifiable iff there is an effective procedure which, if

---

[1]Dummett (1976b) p. 109.

[2]Dummett (1976b) pp. 110-111.

carried out, would warrant the assertion.[1]

For ordinary observation sentences, such as:

34) Snow is white.

that procedure consists in simple observation:

V_{34}) The assertion 'Snow is white' is verifiable iff someone with normal vision when placed in front of snow will observe that the snow is white.

Hence, according to Dummett, we can correctly ascribe the implicit knowledge of the meaning of 'Snow is white' to someone who, having normal vision and being placed in the presence of snow, will assent to the sentence 'Snow is white'.[2] For mathematical sentences, such as:

35) 2+2=4

that procedure consists in calculation:

V_{35}) The assertion '2+2=4' is verifiable iff the operation of applying the function 'x+y' yields the value '4' for the arguments <2,2>.

Hence, according to Dummett, we can correctly ascribe the implicit knowledge of the meaning of '2+2=4' to someone who, when in the presence of the operation of applying the function 'x+y' to the arguments <2,2> will assent to the sentence '2+2=4'.

Now in these cases there is no noticeable difference between semantic realism and semantic anti-realism. The truth-conditions and the verification-conditions of such

---

[1]There is confusion whether the warrant in such cases should be conclusive or defeasible. While this is an important problem, it is not one that I need to consider here.

[2]Recall Dummett's claim that one knows the meaning of 'this tree is taller than that tree' if one can tell, by looking, that this tree is taller than that tree. (1976b) p. 95.

ordinary observation sentences and elementary sentences of arithmetic more or less coincide. However, let the assertion be our problematic undecidable (20) (concerning Jones' bravery). According to Dummett, grasping the sense of (20) will consist in grasping:

> (V_{20}) The assertion 'Jones was brave' is verifiable iff there is an effective procedure which, if carried out, would warrant the assertion.

What is the effective procedure in this case? (20) is thought to be undecidable precisely because there is *no* effective procedure which we could carry out which would warrant it. Thus, there is no situation such that if one were in that situation they would recognize that the assertion is warranted. Therefore, we are not able to attribute a grasp of the verification-conditions of (20) to someone, and subsequently could not attribute an understanding of (20) to them. There would be no reason, then, to attribute an understanding of (20) to anyone: (20) must be *unintelligible*. That, however, is to return to an undesirable verificationism.

The problem is that if a sentence genuinely has unrecognizable *truth*-conditions then it would seem to have unrecognizable *verification*-conditions as well. Well, not quite; by definition, a verification-condition must be recognizable. To be more precise, if a sentence has unrecognizable truth-conditions, then it would seem to have *no* verification-conditions.

The anti-realist will of course object. Consider our standard past-tense statement (17) ("Caesar crossed the Rubicon") whose T-sentence will be something like:

T_{17}) 'Caesar crossed the Rubicon' is true iff Caesar crossed the Rubicon.

On the other hand, (17)'s associated V-sentence will be something like:

(V_{17}) The assertion 'Caesar crossed the Rubicon' is verifiable iff the operation of consulting memories and/or historical records would warrant the assertion.

Thus, someone who, when situated appropriately as regards relevant memories and/or historical records, assented to 'Caesar crossed the Rubicon' would be one who has the knowledge needed to grasp the sense of the assertion. Verification-conditions for (at least some) past tense statements, as opposed to their truth-conditions, are not, it seems, unrecognizable.[1] Therefore, past tense statements pose no threat to semantic anti-realism. In general then, according to Dummett's proposed verification-conditional semantics, understanding the sense of some assertion S will consist in grasping the conditions under which the assertion is verified; or, in other words, in grasping that sentence's associated V-sentence; and we can attribute a grasp of the conditions under which an assertion is verified to someone who, when presented with the effective operation in question, assents to the sentence.

However, the most that the excursion into past tense sentences shows is that the general claim that *any* sentence with unrecognizable truth-conditions lacks verification-conditions is mistaken. It nonetheless remains the case that considerations which render (20)'s truth-conditions unrecognizable also entail that it has no verification-conditions. The apparent distinction between sentences like (20) and sentences like (17) is grounded, I suspect, in an ambiguity in the notion of a truth-condition. If we conceive of truth-conditions in realist terms, then sentences (20) and (17) both qualify as

---

[1]I am here (and in the following) ignoring the realist position of admitting non-actual (i.e. extinct or possible) persons to secure the recognizability of truth-conditions for past tense statements argued for in §3.1.2. Keep in mind that, in this section, I am assuming that the anti-realist has made good on the negative programme.

undecidable in virtue of their truth-conditions being unrecognizable. However, if we conceive of truth-conditions in anti-realist terms - i.e. as identified with verification-conditions - then only sentence (20) qualifies as undecidable in virtue of its truth-conditions being unrecognizable. Sentence (17), in so far as its verification-conditions - i.e. its anti-realist truth-conditions - are recognizable, must be considered as decidable.

There is, therefore, a distinction between sentences like (20) and sentences like (17). The former are undecidable from either perspective whereas the latter are only undecidable from that of the realist: according to a realist conception of truth, past tense statements are undecidable, whereas according to an anti-realist conception of truth they are not. Thus, semantic realism can still be seen to founder on *what it takes to be* undecidable sentences. This is, I think, a bit quick. The realist denies neither that there are truth-conditions nor subsequently truth-values for such statements. What they typically *do* deny is that we can *conclusively determine* what those truth-values are. But, the realist will respond that our inability to conclusively determine the truth-*values* of such sentences should not be taken as an inability to understand what the truth-*conditions* of such sentences are. It is only the anti-realist for whom truth-conditions cannot be epistemically independent of truth-values; for the anti-realist, to know the truth-conditions for a statement S is to be capable, in principle, of determining what S's truth-value is. The argument, then, against a realist construal of past tense statements involves a curious blend of realist and anti-realist notions:

> a) We cannot conclusively determine the truth-values of past tense statements.
> b) Therefore, we cannot grasp the truth-conditions of past tense statements.
> c) Therefore, our understanding of the sense of a past tense statement cannot consist in a grasp of its truth-conditions.

Premise (a) may be acceptable according to a realist notion of truth (certainly to a skeptical realist), but is unacceptable according to an anti-realist notion of truth - if truth is identified with verifiability, then as we can be warranted to assert past tense statements, we can conclusively determine their truth-value. Premise (b), on the other hand, is acceptable according to an anti-realist notion of truth but unacceptable according to a realist notion of truth - if knowledge of truth-conditions can be epistemically independent of knowledge of truth-values, then (b) simply does not follow from (a). Thus, the anti-realist argument against semantic realism on the basis of past tense statements involves an equivocation on the notion of truth.

Where does this leave us? It seems that the manifestation argument can be generalized to: for any theory of meaning M for a language L, if L contains a sentence S which is undecidable according to M, then, as no one can manifest an understanding of S, M is inadequate. Thus generalized, semantic anti-realism is not immune from the argument - at least not as far as sentences like (20) are concerned. Sentences which are undecidable from either perspective cause as many problems for an anti-realist theory of meaning as they do for a realist theory - or more precisely, if they are fatal for a realist theory then they are also fatal for an anti-realist theory.

Dummett's official response to this problem is woefully inadequate:

> It is not necessary that we should have any means of deciding the truth or falsity of the statement, only that we be capable of recognizing when its truth has been established. The advantage of this conception is that the condition for a statement's being verified, unlike the condition for its truth under the assumption of bivalence, is one which we must be credited with the capacity for effectively

recognizing when it obtains...[1]

What Dummett is claiming is that all we require to understand an undecidable

---

[1]Dummett (1976b) p. 111. Cooper (1978) argues that the capacity to recognize a proof is neither necessary nor sufficient for manifesting understanding. It fails to be sufficient "because one may be able to recognize a proof or derivation of a statement, when presented, without being able to understand it ... owing to the complexity of the derivation." It fails to be necessary "because one may be able to understand a mathematical statement and yet be unable to recognize a proof of it when presented ... due to the sheer complexity and difficulty of what is presented to us as a proof." (p. 173). It seems to me that my beginning logic students understand 'every sentence is either true or false' long before they learn how to derive theorems (and so be able to recognize a proof as something more than a dizzying collection of symbols).

See also Tennant (1981), (1984), (1985) and (1987) p. 119. In the latter he interprets 'undecidable' in two ways: (i) a sentence is undecidable just in case it is impossible to either prove or refute it; and (ii) a sentence is undecidable just in case we do not, at present, possess either a proof or a disproof of it. He then claims that "the weaker reading" is "the one involved in my claim that 'the anti-realist can admit the possibility of definitely meaningful but undecidable sentences'." (p. 119, in (1984) he distinguishes them as *pro tempore* undecidable as opposed to *tout court* undecidable). However, in §3.1.2.1 it was argued that if a sentence is ever decidable then it is always decidable. If a sentence S is not now either provable nor refutable, but will be at some future time, then S is now decidable. Thus, Tennant's case for the existence of such undecidable sentences depends upon their never (in fact) being either proven or refuted. In such a case, the undecidability of S will reduce to the undecidability of 'S will never be proven', which, involving quantification over a (potentially) infinite domain, is only asymmetrically undecidable. As such, it is insufficient to ground the undecidability of S.

In addition, Weir (1983) and (1985) question why Tennant's claim (that manifestability constraints are satisfied by the ability to recognize a proof rather than by the ability to construct a proof) should, if accepted, provide "a warrant for attributing grasp of meaning rather than merely for refuting a denial of meaningfulness." ((1985) p. 69). That is, why is it taken to support anti-realism rather than refute (the more radical) verificationism? (See Tennant's reply in (1985)). More importantly, Weir (1986) points out that to say that a sentence is meaningful just in case we *would* or *could* recognize evidence warranting its assertion if presented with it is to locate the meaningfulness of the sentence in the *truth* of an associated subjunctive conditional, which Dummett regards as generally undecidable. On the other hand, if the anti-realist is allowed to appeal to dispositional states (expressed by subjunctive conditionals) to ground the meaningfulness of undecidable sentences, there is no reason why the realist could not avail herself of the same manoeuvre. See also Demopoulos (1982) for much the same argument.

sentence is possession of an (unexerciseable) capacity to recognize a (non-existent) proof. The fact that such a capacity cannot, in any way, be manifested, seems not to worry him. On the other hand, he indicts the realist for being unable to manifest a capacity to recognize unrecognizable truth-conditions![1]

Perhaps, then, anti-realists should base their argument entirely on sentences of the second sort - i.e. sentences which are deemed undecidable only from a realist, not an anti-realist, perspective. In other words, they will argue that the requirement of manifestability for grasp of meaning forces the epistemic interdependence of knowledge of truth-conditions and knowledge of truth-values needed for the acceptability of premise (b) above. Thus, for independent reasons, they will insist that the realist find (b) acceptable, and hence the conclusion will follow. What the realist needs, then, is a way to block the inference from the requirement of manifestability to (b). If truth-conditions can be made intelligible in a way other than by a capacity to recognize when such conditions obtain or fail to obtain, then the anti-realist argument will fail.

---

[1]George (1987) remarks that "we do understand a given undecidable sentence, because we can effectively determine whether any given construction is a proof of it - even though there is no effective procedure that we could in principle apply to yield either a proof of the sentence or a refutation of it." (p. 404). He locates the required capacity, then, in an ability to recognize that any proof one could actually be presented with (i.e. recognize) fails to constitute a proof of the sentence. But, if this were sufficient, the realist could locate their required capacity in an ability to recognize that any state-of-affairs that one can actually recognize as obtaining underdetermines the truth-conditions for the sentence. As Weir (1986) says: "we possess the capacity [to recognize the truth-value of undecided sentences] if we could recognize the truth-value in favourable circumstances, circumstances we may not be able to get ourselves into at will." (p. 470). George criticizes Weir for understanding 'favourable circumstances' as involving possession of superhuman powers, but §3.1.2.1 argued that this need not be the case.

Before considering such options, we should recognize that the anti-realist cannot legitimately ignore sentences like (20). Just as a single sentence with unrecognizable realist truth-conditions is all that is required for the inadmissibility of a global semantic realism, so too would a single sentence lacking anti-realist truth-conditions reveal the inadmissibility of a global semantic anti-realism. Sentences like (20), it seems to me, destroy any hope for an anti-realist semantics - at least, as long as they also destroy any hope for a realist semantics - and hence the positive programme fails. At best, the anti-realist could raise only the negative programme.

The anti-realist may resurrect the positive programme by maintaining that all that is lost is the prospect for a *global* theory of meaning - local theories for isolated classes of sentences are not in jeopardy. For example, they might argue that sentences about the past require an anti-realist interpretation. I am content to let this possibility stand - we only need to observe that, as a general response to the problem of undecidables, it is just as open to the realist as it is to the anti-realist.

4.2 Alternative Accounts of Manifestability

Dummett's manifestability requirement says that grasp of meaning must *ultimately* consist in a capacity to manifest one's grasp by being able to associate correctly an assertion with a recognizable state-of-affairs which either verifies it or falsifies it. What the requirement does *not* say is that grasp of meaning must *exclusively* consist in such a capacity. There may be manifestable capacities, ultimately founded on the first kind, which would also allow ascription of the knowledge needed to grasp the sense of an

assertion.[1] Take, for example, the sentence:

36) The Queen of England pays taxes.

Following Frege, the sense of (36) is determined by the senses of its constituents. For simplicity, we can suppose (36) to have two constituents - an object expression:

37) The Queen of England

and a function expression:

38) x pays taxes.

Consider the following sentences:

39) The Queen of Monaco was Grace Kelly.

40) The people of England eat greasy breakfasts.

41) People in Canada pay taxes.

Take someone P whom we have determined, via Dummett's proposal, to have grasped the senses of (39)-(41). Given that the sense of a sentence is a function of the senses

---

[1]The core of the following argument derives from some remarks made by Appiah (1986) (criticised by Edgington (1985)). He argues that it is only a form of semantic scepticism which precludes one from accepting other means for the attribution of a grasp of sense to someone. McGinn (1976) also observes that at best Dummett's argument forces the dilemma that "either it is, after all, a mistake, an illusion, to suppose ourselves (or others) capable of conceiving a recognition transcendent reality, or there must be some way of manifesting such a conception in use otherwise than by the exercise of a capacity to conduct a verification procedure." (pp. 29-30). His 'other way' is by manifesting a capacity to interpret the utterances of others. In other words, success in communication is a sufficient manifestation: "[This] way of locating knowledge of speech interpretation serves, unambitiously but satisfactorily, to relate conceptions of transcendent states of affairs to a practical linguistic capacity, to actual use." (p. 30; Howich (1982) agrees). Loar (1987) argues that a semantic realist, who invokes a holistic theory of meaning, can meet the manifestation constraints (of course, Dummett rejects all holistic theories of meaning - see his reply to Loar in Dummett (1987)). What Loar means by 'holistic', however, is very similar to what I mean by 'compositional' in the following.

of its constituents, the fact that P grasps the senses of (39)-(41) warrants our attributing to him a grasp of the following two object and one function expressions:

42) The Queen of ___

43) The ___ of England

44) ___ pay taxes

and also that P understands how the senses of constituent expressions can be conjoined to form the sense of a sentence. In such a case then, we are perfectly warranted to attribute to P a grasp of the sense of (36), even though P has not revealed his grasp by manifesting a capacity to assent to (36) when appropriately situated *viz-a-viz* the state-of-affairs in which the Queen of England pays her taxes. Nonetheless, our attribution of the knowledge of the sense of (36) is ultimately achieved via a connection to such a manifestation.

In other words, if the sense of a sentence is a function of the senses of its constituents, then P's manifestation of the grasp of the constituents of a sentence should suffice to attribute to P a grasp of the sentence itself. Conversely, P's manifestation of the grasp of a compound sentence in which some simple sentence is a constituent should suffice to attribute to P a grasp of that simple sentence. Consider:

45) If the Queen of England pays taxes, then someone in England pays taxes.

It seems at least plausible that acceptance of (45) indicates an understanding of (36), and someone could accept (45) without having to manifest their understanding in the manner Dummett proposes. Dummett will retort that one can accept (45) on purely logical grounds, with no understanding of the senses of either the antecedent or the constituent.

For example, P can accept:

46) If a snark is a boojum, then something is a boojum.

without any understanding of either the antecedent or the consequent. Fair enough, but it is less plausible to suppose that one will accept:

47) If the Queen of England pays taxes, then England's monarchy has undergone a profound change.

solely on the basis of recognizing the logical connection between antecedent and consequent. Dummett's reply will be to ask for the basis on which we can attribute a grasp of (47) to someone. The answer, it seems to me, would be something like this: P knows that, at least historically, England's monarch was exempt from certain obligations held by others living on the Island of Great Britain, and that this privilege was thought to be an important element of the institution. That P knows these historical facts presupposes that he understands the senses of the sentences expressing these facts.[1] I can agree that we attribute to P a grasp of those sentences ultimately by his manifestation of the required capacity, but it is his grasp of *those* sentences which warrant our attribution of a grasp of (47) to him, and hence of our attribution of a grasp of (36) to him. Again, the only point I wish to make is that, while Dummett's manifestability requirement might be taken as the *ultimate* basis for attributing grasp of sense, it need not be the *only* basis for doing so. P's independent grasp of the senses of the constituents of some sentence S or her independent grasp of the sense of a compound sentence containing S as a constituent may, in some circumstances, suffice to

[1]In other words, one can manifest an understanding of a sentence by demonstrating knowledge of some of its non-trivial consequences.

attribute grasp of S to P.

If this is at least plausible, then such an account should be possible for our problematic (17). If P is able to demonstrate a grasp of the sense of 'Caesar' by its presence in other sentences, and similarly is able to demonstrate a grasp of the senses of 'the Rubicon' and what it is to cross something, then we should be able to attribute to P a grasp of the sense of (17). For example, through P's manifesting an understanding of the following sentences:

48) Caesar was the first Roman to invade Britain.

49) Columbus crossed the Atlantic.

50) The most famous river in Italy is the Rubicon.

we are warranted to attribute to P an understanding of the senses of 'Caesar', 'x crossed y' and 'the Rubicon'. Furthermore, P's understanding of these sentences reveal that he is able to grasp how the senses of the constituents of a sentence contribute to the senses of the sentences made up of them. We have all the evidence we need, then, to attribute to P a grasp of (17). That grasp can perhaps be expressed in the following form:

> 51) The object designated as 'Caesar' in the sentence 'Caesar was the first Roman to invade Britain' performed the action designated as 'x crossed y' in the sentence 'Columbus crossed the Atlantic' to the object designated as 'the Rubicon' in the sentence 'The most famous river in Italy is the Rubicon'.

That understanding can be made more general:

> 52) The object designated by 'Caesar' in sentences containing that object-expression performed the action designated by 'x crossed y' in sentences containing that function-expression to the object designated by 'the Rubicon' in sentences containing that object-expression.

Finally, accepting the schema <'B' refers to B>, we can derive the following:

53) The object designated by 'Caesar' in sentences containing that object-expression = Caesar.

54) The function designated by 'x crossed y' in sentences containing that function-expression = x crossed y.

55) The object designated by 'the Rubicon' in sentences containing that object-expression = the Rubicon.

(51) reduces, then, to:

56) Caesar crossed the Rubicon.

which is the sentence used on the right hand side of (17)'s associated T-sentence; i.e. its truth-conditions. If this approach is plausible, then we seem to have vindicated the semantic realist's claim that a grasp of (17) consists in a grasp of its truth-conditions.

Tennant (1981) offers much the same defense for the apparent problem posed by undecidable sentences for *anti*-realism discussed in §4.1. Understanding an undecidable sentence does not require, Tennant says, being capable of producing evidence which would warrant it, but only being capable of *recognizing* such evidence if presented with it. Of course, there is no such evidence for undecidable sentences. Tennant remarks that one manifests such knowledge by demonstrating a grasp of compositionality:

> We have no right to insist that grasp of meaning be confirmed sentence by sentence, that we present for inspection proofs or disproofs of each and every sentence for which we raise the question whether meaning has been grasped. The intuitionist has compositional capacities like those of the classicist. He can understand individual words and operators and grasp sentential pedigree. And this allows him to grasp the meanings of new sentences, including Goldbach's conjecture. Moreover, the 'basic grasps' involved can be ascertained by investigating his general ability to infer conclusions, reduce proofs to canonical proofs, find proofs of simple theorems, etc. etc.[1]

---

[1]Tennant (1981) p. 117.

But, if the anti-realist is allowed to attribute the capacity to recognize evidence which *would* warrant the assertion of an undecidable sentence to a person in virtue of their displaying a capacity to recognize evidence which *does* warrant the assertion of a decidable sentence, there is no reason to prohibit the realist from attributing the capacity to recognize the obtaining of an undecidable sentence's truth-conditions to a person in virtue of their displaying a capacity to recognize when the truth-conditions of decidable sentences obtain. The anti-realist then, must find some other objection to the proposal.

They might attempt to maintain that the proposal is trivial: it maintains that a grasp of 'Caesar crossed the Rubicon' consists in a grasp of 'Caesar crossed the Rubicon'. This response, however, ignores the appeal to compositionality in the process of demonstrating of in what the grasp of (17) consists. The grasp of 'Caesar crossed the Rubicon' *does* consist in a grasp of 'Caesar crossed the Rubicon' (how could any theory deny this obvious truism?), but the proposal suggests *how* a grasp of (17) can be manifested in a grasp of its constituents. It further suggests that a grasp of a sentence's constituents (combined in the right structure) can *be* a grasp of that sentence's truth-conditions. Moreover, it at least seems to conform to Dummett's insistence that grasp of meaning must ultimately consist in a certain kind of manifestation.

A second, more substantial, response would be that such a proposal eliminates any motivation for semantic realism. The proposal says that we can understand S through a grasp of S's truth-conditions, but that a grasp of S's truth-conditions depend upon a grasp of the verification-conditions of sentences other than S. If we cannot ultimately eliminate an appeal to verification-conditions, there is no real reason to seek to avoid

them for sentences like (17). In other words, if we cannot eliminate an appeal to verification-conditions *everywhere*, there is no real reason to eliminate it *anywhere*.

My reply to this is, admittedly, tenuous. It takes seriously Dummett's observation that, on the level of decidable sentences, there is no substantial difference between a truth-conditional and a verification-conditional semantics:

> There is no substantial disagreement between the two models of meaning so long as we are dealing only with decidable statements: the crucial divergence occurs when we consider ones which are not effectively decidable.[1]

---

[1]Dummett (1973c) p. 231. See also (1973c) pp. 224-225 and (1963b) p. 155. C. Wright (1987) accepts this in the introduction, but seems to reject it in Ch. 7. His argument seems to be this: according to the manifestation constraint, knowledge of the meaning of some sentence S consists in a capacity to determine when (some of) the criteria which warrant an assertion of S obtain - i.e. essentially Dummett's testing procedure. However, where S is an empirical (but perhaps decidable) sentence, criteria warranting its assertion (i.e. criteria for ascribing knowledge of truth-conditions, according to the realist) will always be defeasible. But, (suppose) genuine knowledge is never defeasible. Thus, knowledge of truth-conditions is always underdetermined by (and hence cannot be identified with) such criteria. C. Wright's point, then, is that as all we can manifest is knowledge of such criteria, knowledge of truth-conditions can never be manifested, including those of decidable sentences.

However, C. Wright confuses knowledge of what a sentence's truth-conditions *are* with knowledge of when a sentence's truth-conditions *obtain*. One can know what a sentence's truth-conditions are even when they make mistakes about when they obtain (a child who believes that Santa Claus lives at the North Pole does not necessarily misunderstand the sentence 'Santa Claus lives at the North Pole'). The realist with skeptical sympathies will agree that a manifestable knowledge of a sentence's truth-conditions will always underdetermine (and hence fail to be identified with) knowledge of its truth-conditions. In addition, if C. Wright continues to insist on identifying knowledge of sentential meaning with knowledge of when criteria warranting assertion obtain, then his anti-realism will undoubtedly founder on genuinely undecidable sentences.

In a similar vein, Vision (1988) §7.4 attempts a realist argument from considerations of defeasibility: truth-conditions are never, while (empirical) verification-conditions are always, defeasible. Therefore, truth-conditions (of even decidable sentences) cannot be equivalent to (but must transcend?) verification-conditions. Vision, however, gives no particular reason why an anti-realist would accept that truth-conditions are never defeasible.

Let us suppose that this claim is correct, and that it entails that a truth-conditional theory of meaning is adequate for decidable sentences. Dummett's argument is that, while a truth-conditional theory of meaning is adequate for decidable sentences, it is inadequate when extended to non-decidable sentences, and hence is inadequate in general. My proposal above took this form: for a non-decidable sentence S, a grasp of its sense can be attributed to anyone who manifests an ability to grasp the senses of S's constituents in sentences other than S. Let Q, R, and T be the sentences other than S whose grasp by a person P warrants the attribution of a grasp of the sense of S to P. The upshot of the present response is that as grasp of the senses of Q, R, and T consist in a grasp of their verification-conditions, we cannot ultimately eliminate appeal to verification-conditions in our ascription of the grasp of S to P, hence there is no reason why we should opt for a truth-conditional analysis over a verification-conditional analysis for S.[1]

Suppose, however, that Q, R, and T are decidable sentences. In such a case, a truth-conditional analysis of them is adequate, and as it is a grasp of their senses which warrants an attribution of the grasp of the sense of S to P, an account of what P's grasp of S consists in need not make reference to verification-conditions. In such a case, we can eliminate appeal to verification-conditions *everywhere*, and hence the objection would fail.

---

[1]Tennant (1987) suggests that such a manoeuvre is already to give up semantic realism - i.e. the view that a realist construal of truth is the central concept in a theory of meaning (pp. 128-129). However, once it is recognized that a realist construal of truth *accommodates* the restrictive anti-realist sense, this objection fails.

This reply, while tenuous, is not implausible. It is intuitive to suppose that we learn our language originally by coming to grasp the senses of observation sentences, and we extend the range of our sentences by extending outward from these basic sentences. Observation sentences are certainly decidable, hence a truth-conditional analysis of them should be adequate. If observation sentences form the core from which the senses of all other sentences can be traced, then, according to my reply, a truth-conditional analysis should be possible for all sentences in the language (at least for all of the problematic past tense sentences).[1]

In my case for (17), however, I made appeal to (48) and (49), both of which are in the past tense and hence are considered undecidable under a realist construal. Perhaps we were injudicious in our choice of sentences other than (17) to test whether a attribution of the grasp of (17) to P was warranted. What we need are decidable sentences in the present tense which would warrant an attribution of the senses of 'Caesar' and 'x crossed y' to P. Take 'x crossed y' first. A grasp of, say:

57) Fred is crossing the room.

coupled with an understanding the past 'flowing' backwards from the present should suffice in attributing a grasp of that function to P. The problem is obvious: on account of what can P grasp the notion of the past flowing backwards from the present? What

[1]The Dummettian view is that an extension of the realist analysis to sentences further away from the observation sentences is beyond all reasonable limits. There is no longer a sufficient analogy between, say, a sentence involving quantification over an infinite domain and a sentence involving quantification over a finite domain. As was argued, it is far from clear that Dummett's observations in this matter have the force they appear to.

provides the basis of the transition from a grasp of the present tense to a grasp of the past tense?[1]

Perhaps the problem is a pseudo-one. After all, Dummett insists that such knowledge is implicit, not explicit. All we need to ascribe a grasp of the past tense to someone is to position them, say, in front of a table with a book on it, and see if they assent to the sentence:

58) The book is on the table.

and then remove the book, and see if they assent to the sentence:

59) The book was on the table.

This will provide, it seems to me, evidence both that P grasps the sense of the past tense *and* that the sense of (59) derives from the sense of (58), which is a decidable sentence (and hence one for which a truth-conditional analysis is adequate).

Thus, if (and it is a big 'if') (i) the senses of all non-decidable sentences can ultimately be traced to the senses of decidable sentences, and (ii) decidable sentences are ones for which a truth-conditional theory of meaning are adequate (i.e. we need make no reference to verification-conditions to attribute a grasp of them to someone), and (iii) a grasp of the sense of a sentence S can consist in a grasp of the senses of the

---

[1]This problem is analogous to that involving truth-value links. It certainly is part of our linguistic practice that acceptance of 'x is F' at t_n warrants acceptance of 'x was F' at $t_{m>n}$. The question is, what grounds this practice? C. Wright (1987) Ch. 5 seems to argue that as expressions of truth-value links are themselves tensed statements, we cannot appeal to knowledge of them to ground knowledge of the meanings of past (or future) tensed statements with unrecognizable truth-conditions. Such a response fails, however, if we can (and do) in fact manifest knowledge of truth-value links. Both Dummett and C. Wright, as mentioned, accept that such links are a significant aspect of our linguistic practices.

constituents of S derived from a grasp of the senses of sentences other than S containing those constituents (with the proper structure), and (iv) the sentences other than S in question are all decidable ones, then a truth-conditional theory of meaning is possible for such problematic sentences as (17).

So, where are we? The realist can accept Dummett's claim that grasp of meaning must ultimately consist in use - i.e. in a certain kind of capacity. Semantic realism, which takes a grasp of the meaning of a sentence to consist in a grasp of its truth-conditions is not obviously at odds with this constraint, and I have provided an outline of how semantic realism can be made consistent with this constraint. That outline is schematic, and needs a great deal of work, but at least as a research project is validated - it is far from clear that an anti-realist theory of meaning is the only one consistent with an adequate account of understanding.

4.3 Semantics and Compositionality

The main claim of Dummett's positive programme is that the only admissible notion of truth consistent with an adequate theory of meaning is that of warranted assertibility. What I intend to show now is that this claim is unacceptable - there must be conditions for the correctness of an assertion other than those of warranted assertibility. This in itself will not constitute an argument *for* semantic realism but only an argument *against* semantic anti-realism. To draw the stronger conclusion it must be shown that these other conditions must be conceived in terms of realist truth-conditions.

The base argument is primarily Brandom's.[1] It starts from the recognition of the role of compositionality in a theory of meaning. Frege, to whom the importance of compositionality can be attributed, mentions at least two compositional theses: the sense of an atomic sentence is determined by the senses of its constituent words and its internal structure; and the sense of a compound sentence is determined by the senses of its constituent sentences and the contribution of the logical constants employed. Only the latter thesis concerns us here.

Thus, the sense of some arbitrary compound sentence, "A*B", is determined exhaustively by the senses of "A" and of "B" and of how the connective "*" internally relates them. Moreover, we can assume that such connectives are logical constants - i.e. they play the same role in any sentence which contains them. In other words, we can take their contribution for granted. Let me introduce some technical notions to make this more precise. We can think of an arbitrary compound sentence as consisting of one component sentence and one sentential predicate. For example, a conjunction "A∧B" can be thought of as consisting of the atomic sentence "A" and the sentential predicate "conjoins with B".[2] Similarly, disjunction, negation, and the conditional can be rendered respectively as "alternates with B", "is not the case", and "implies B". We can thus express any arbitrary compound sentence as "P(a)" where "a" names an atomic sentence

---

[1]Brandom (1976). I modify Brandom's symbolism, which tends to be cumbersome. The thrust of his argument is that sentential meaning is underdetermined by assertibility-conditions. See also Appiah (1986) Ch. 7 for a discussion of this argument.

[2]I am not using the expression 'atomic sentence' in any absolute sense. Any constituent sentence of a compound can be regarded as atomic relative to that compound.

and "P" names a sentential predicate.[1] Now, as with the logical constants, we can take the contribution of the sentential predicate in determining the sense of the compound for granted. According to the thesis, the sense of "P(a)" (represented as "S{P(a)}") can be thought of as the value of a function $\int$ which take as argument the sense of "a" (represented as "S{a}"). We can thus render the thesis of compositionality as:

(TC) $(\forall P)(\exists \int)(\forall a)(S\{P(a)\} = \int\{S(a)\})$

Understanding semantic realism as the thesis that the sense of a sentence consists in its truth-conditions:

(R) $S\{a\} = T\{a\}$

and semantic anti-realism as the thesis that the sense of a sentence consists in its assertibility-conditions:

(AR) $S\{a\} = A\{a\}$

we can derive specific formulations of (TC) for both realism and anti-realism:

(TCR) $(\forall P)(\exists \int)(\forall a)(T\{P(a)\} = \int(T\{a\}))$
(TCA) $(\forall P)(\exists \int)(\forall a)(A\{P(a)\} = \int(A\{a\}))$

Modifying an expression from Brandom, call any language for which (TC) holds - i.e. any language for which we can explicate the sense of a compound sentence by a consideration of its composition - *compositionally explicable.* Therefore, if the thesis of compositionality is required for any adequate theory of meaning for a language L, then L must be compositionally explicable. Similarly call any language for which (TCR) holds *alethically explicable* and any language for which (TCA) holds *assertorically explicable.*

---

[1]In what follows, I use lowercase English letters to stand for sentences and uppercase English letters to stand for sentential predicates.

Thus, given the necessity of compositionality for any adequate theory of meaning, any language for which a realist theory of meaning is acceptable must be alethically explicable and any language for which an anti-realist theory of meaning is acceptable must be assertorically explicable.

Now, the thesis of compositionality, in either of its two forms, entails a semantic version of Leibniz's Law. If it is the *sense* of an atomic sentence which exclusively contributes to the sense of a compound sentence in which it is embedded, then any other atomic sentence alike in sense will contribute no more and no less to the sense of the compound as did the original atomic sentence. We can thus establish the following semantic principle of substitutivity:

(PS) $(S\{a\} = S\{b\}) \rightarrow (S\{P(a)\} = S\{P(b)\})$

The realist and anti-realist versions of (PS) are, respectively:

(PSR) $(T\{a\} = T\{b\}) \rightarrow (T\{P(a)\} = T\{P(b)\})$
(PSA) $(A\{a\} = A\{b\}) \rightarrow (A\{P(a)\} = A\{P(b)\})$

Therefore, as (TCR) and (TCA) entail (PSR) and (PSA) respectively, we can conclude (PSR) must hold for any language for which a realist theory of meaning is adequate and that (PSA) must hold for any language for which an anti-realist theory of meaning is adequate.[1]

Brandom notes that (PSR) and (PSA) both fail for natural languages, such as English. Under the assumption that 'George Gardiner' and 'Mark's dad' both denote the

---

[1]I will shortly weaken this claim - a realist theory of meaning is acceptable for a language in which (PSR) fails *only if an explanation consistent with a realist theory of meaning can be given*, and the same holds for an anti-realist theory and (PSA).

same object, the truth-conditions for:

60) George Gardiner is a retired social worker.

and the truth-conditions for:

61) Mark's dad is a retired social worker.

are the same, yet the truth-conditions for:

62) Ernie believes that George Gardiner is a retired social worker.

differs from that of:

63) Ernie believes that Mark's dad is a retired social worker.

Therefore, English, as a natural language, is not alethically explicable, and hence it would seem that a realist theory of meaning is inadequate. Similarly, the assertibility-conditions for:

64) I will marry Jane.

and:

65) I predict that I will marry Jane.

are the same, yet the assertibility-conditions for:

66) If I will marry Jane, then I will not be a bachelor.

and:

67) If I predict that I will marry Jane, then I will not be a bachelor.

differ.[1] Thus, neither is English assertorically explicable, and hence it would seem that

---

[1]Owing to some uncertainty about the acceptability of Brandom's example (in turn modified from that of Dummett (Dummett (1973a) p. 450), my example is a modification of his. Later on I offer an argument to the effect that *any* empirical sentence will lead to the same result. See also Dummett (1990) and (1991b) Ch. 7.

an anti-realist theory of meaning is likewise inadequate.[1]

Now an anti-realist might argue that there is an interpretation of (65) which would allow (67) to have the same assertibility-conditions as (66): namely if we interpret it as:

65′) I *correctly* predict that I will marry Jane.

then its embedding will produce:

67′) If I correctly predict that I will marry Jane, then I will not be a bachelor.

which does not differ in its assertibility-conditions from (66). This will ensure that (66) and (67) have the same assertibility-conditions, but at the price of precluding (64) and (65) from having the same such conditions. The warrant of the assertion of (64) depends only upon conditions which are *defeasible* - the state of information which warrants its assertion at one time may fail, with the addition of new evidence, to warrant that

---

[1]A stronger (mirror-image) counter-example can be raised. Those even mildly sympathetic to Descartes' third skeptical argument in the *Meditations* will concede that:

a) It is exactly as if a malignant demon is creating the illusion of a chair in front of me and there is no chair in front of me.

is assertible in the same experiential condition of having a visual and tactile sensation of a chair in front of one as:

b) It is exactly as if a malignant demon is not creating the illusion of a chair in front of me and there is a chair in front of me.

Yet, their component sentences - "A malignant demon is creating the illusion of a chair in front of me and there is no chair in front of me" and "A malignant demon is not creating the illusion of a chair in front of me and there is a chair in front of me" - cannot, on pain of contradiction, share the same assertibility-conditions. Thus, the senses of (a) and (b) cannot (solely) be a function of the assertibility-conditions of their component sentences.

It is interesting to compare this problem with Quine's Underdetermination Thesis (discussed in the Putnam chapter) that, associated with any empirical theory, there exists another which is empirically equivalent yet cognitively inequivalent. If the underdetermination thesis is true (and entails the existence of such theories), then for any empirical sentence S there exists another, S′, which shares assertibility-conditions but not meaningfulness. Thus, assertibility-conditions must underdetermine meaning.

assertion at other times. The warrant for asserting (65′), however, by smuggling in a notion of correctness stronger than that of defeasible warrant, depends upon evidence which could not be overturned by new evidence.[1] If I assert 'I will marry Jane' on June 1, and on June 2 Jane is killed in a car accident, then my assertion, while false, could nonetheless have been warranted. However, if I assert 'I correctly predict that I will marry Jane' on June 1, and she dies the next day, then my assertion, in addition to being false, was unwarranted: the only state of information which would warrant its assertion, unlike that which would warrant (64), would be that which precluded any possible further counter-evidence. Thus, (64) and (65′) would not have the same assertibility-conditions, and hence the fact that (66) and (67′) do have the same assertibility-conditions provides no evidence for supposing English to be assertorically explicable.

Thus, according to this argument, because semantic realism in the form of (R) and semantic anti-realism in the form of (AR) are respectively committed to (TCR) and (TCA), which respectively entail (PSR) and (PSA), both of which fail for natural languages, neither (R) nor (AR) can be the core expressions of an adequate meaning-theory for a natural language. Brandom's response is that this is worrisome only under the assumption that exactly one of (R) or (AR) must be correct.

Consider an alternative meaning-theory:

---

[1]When I first developed this argument, I assumed the problem with (65′) was that it smuggled in a realist notion of truth. This need not be the case, as all that it requires is a distinction between a *defeasible* assertion and an *indefeasible* one. There is no compelling reason, at this point, to identify conditions for an indefeasible assertion with realist truth-conditions, and there is reason to refrain from doing so - the truth-conditions of my marrying Jane and my predicting, even correctly, that I will do so differ.

(R′) $S\{a\} = (T\{a\} \wedge X\{a\})$

where 'X' denotes some condition *other* than truth-conditions. As (R′) continues to take truth-conditions as *a* central notion, it qualifies as a form of semantic realism. The associated compositionality thesis would be:

(TCR′) $(\forall P)(\exists \int)(\forall a)((T\{P(a)\} \wedge X\{P(a)\}) = \int(T\{a\} \wedge X\{a\}))$

which would generate a substitutivity thesis in the form:

(PSR′) $((T\{a\} \wedge X\{a\}) = (T\{b\} \wedge X\{b\})) \rightarrow ((T\{P(a)\} \wedge X\{P(a)\}) = (T\{P(b)\} \wedge X\{P(b)\}))$

We have no reason to suppose that (PSR′) fails for sentences (60)-(63), for the fact that (60) and (61) share the same truth-conditions is no guarantee that they share the same X conditions, whatever those turn out to be. Similarly, consider an alternative anti-realist theory:

(AR′) $S\{a\} = (A\{a\} \wedge Y\{a\})$[1]

which would entail the following compositionality thesis:

(TCA′) $(\forall P)(\exists \int)(\forall a)((A\{P(a)\} \wedge Y\{P(a)\}) = \int(A\{a\} \wedge Y\{a\}))$

and hence the following substitutivity principle:

(PSA′) $((A\{a\} \wedge Y\{a\}) = (A\{b\} \wedge Y\{b\})) \rightarrow (A\{P(a)\} \wedge Y\{P(a)\}) = (A\{P(b)\} \wedge Y\{P(b)\}))$

We likewise have no reason to suppose that (PSA′) fails for sentences (64)-(67), for there is no guarantee that sentences (64) and (65) share the same Y conditions.

---

[1]Dummett is aware of the deficiencies of (AR). In (1991b) Ch. 2 he introduces a distinction between *assertoric content* and *ingredient content.* Assertoric content is determined by the A-conditions, while ingredient content would, presumably, be determined by the Y-conditions, whatever they turn out to be. Sentences (64) and (65), in Dummettian language, share the same assertoric but not ingredient contents.

Thus, modifications of the basic meaning theories provides a promising solution to the dilemma.

What modifications, specifically, to the basic meaning theories are required? That is, what are the X and Y conditions which the semantic realist and the semantic anti-realist respectively need in order to preserve compositional explicability? Frege, in dealing with the issue, focused on how the sentential contexts in the problematic compound sentences affected the senses of the constituent sentences. The solution I want to offer follows along the same line.

The truth-values of sentences (62) and (63) do not depend upon the truth-values of their constituent sentences; e.g. the truth of sentence (62) is compatible with either the truth or the falsity of sentence (60). This is due to the nature of belief - i.e. to the doxastic context in which (60) is embedded. But, due to the thesis of compositionality, some aspect of sentence (60) *must* contribute to the sense of sentence (62). Sentence (62) is correctly assertible - or true - just in case Ernie in fact believes that sentence (60) is true. Now, Ernie will believe that sentence (60) is true just in case he is in possession of evidence which would warrant its assertion. In other words, (62)'s correctness depends upon whether Ernie recognizes that the *assertibility*-conditions for (60) obtain. Thus, it seems not unreasonable to suppose that it is (60)'s *assertibility*-conditions, as opposed to its *truth*-conditions, which contribute to the sense of sentence (62). Therefore, there is good reason to suppose that the X conditions required by the

semantic realist just are assertibility-conditions.[1]

One may be worried about the apparent sleight of hand - Ernie might very well grasp the conditions under which (60) is true, the conditions under which it is assertible, the conditions under which (61) is true, and the conditions under which it is assertible, and *nonetheless* believe that George Gardiner is a retired social worker but not that Mark's dad is one. Quite right, but only if sentences (60) and (61) have different assertibility-conditions and hence, on this view, have different meanings. To say that Ernie believes that George Gardiner is a retired social worker is to say that Ernie recognizes that the conditions which would warrant an assertion of that belief obtain. To say that he does not believe that Mark's dad is a retired social worker is to say that he does not recognize that the conditions which would warrant an assertion of such a belief obtain; to demonstrate that the assertibility-conditions for a pair of sentences differ it is sufficient to show that one set of conditions may be recognized as obtaining while the other not.

Thus, it seems to me that the moral to draw from sentences (60)-(63) is not that

---

[1]It may seem that I have been assuming that Ernie displays an unusually high degree of rationality. Actually, I have been assuming that, associated with each sentence, there are some standard set of assertibility-conditions. That assumption is, I admit, somewhat suspect. It may be the case that Ernie believes sentence (60) for deviant reasons - e.g. he gets his beliefs anew each morning from a random belief generating machine (or from reading the Tarot cards). In that case, Ernie (tacitly) assigns a different set of assertibility-conditions to sentences from others (for him, "S" is assertible just in case "S" has been produced that morning by the random belief generating machine). This does not, in itself, challenge the proposed meaning theory; all that follows is that Ernie assigns a different meaning to (60) than others. If we know (60)'s truth-conditions, and we know the assertibility-conditions Ernie assigns to it, then we should be able to recover the meaning that Ernie assigns to it. For simplicity, however, assume that, associated with each empirical sentence, there is some standard set of assertibility-conditions.

they refute a theory of meaning which conjoins truth and assertibility-conditions, but rather that sentences (60) and (61), contrary to our original supposition, are not alike in meaning. The truth-conditions of (60) and (61) coincide, but their assertibility-conditions do not.[1]

Can there be a pair of sentences alike in both truth-conditions and assertibility-conditions for which the relevant principle of substitutivity fails? It seems to me that, for any pair of sentences offered as a purported counter-example, it would be at least as reasonable to conclude that they do not share *both* truth *and* assertibility-conditions as it would be to conclude that they constitute a genuine counter-example: i.e. it is at least as reasonable to suppose that such a pair of sentences reveal failure of synonymy rather than failure of substitutivity. Perhaps that is enough for my purposes.

We seem justified, then, to take as a more adequate formulation of the basic semantic realist's theory of meaning the following principle:

(R″) $S\{a\} = (T\{a\} \wedge A\{a\})$

A serious objection to (R″) may be raised: once assertibility-conditions are required by the semantic realist, a major concession to the semantic anti-realist has been made - perhaps so much so that there is no longer any strong motivation to preserve realist sentiments. If we *need* something other than truth-conditions, then perhaps we do not need truth-conditions at all. If appeal to truth-conditions can be entirely avoided then there is no longer any reason to remain a realist of either the semantic or the

---

[1]For example, Ernie's overhearing me utter 'My dad is a retired social worker' may be sufficient to warrant his assertion 'Mark's dad is a retired social worker' but *not* sufficient to warrant his assertion 'George Gardiner is a retired social worker'.

metaphysical variety.

For example, in the context of preserving compositional explicability in the face of sentences (60)-(63), it was appeal to assertibility-conditions which did the lion's share. Perhaps it did the only share; if we identify the meaning of sentences (60) and (61) exclusively with their assertibility-conditions, and we suppose that they share the same such conditions, then sentences (62) and (63) cannot violate substitutivity. Compositional explicability would be preserved, at least for these sentences, without any mention of truth-conditions at all.

However, to identify the meaning of a sentence exclusively with its assertibility-conditions is to resurrect (AR). As seen, (AR) is incompatible with compositional explicability. Assertibility-conditions *alone* cannot do the job. Thus, the acceptability of eliminating appeal to truth-conditions in a theory of meaning will depend upon the success of determining what the semantic anti-realist's required T conditions are.

Recall that sentences (64)-(67) reveal that conditions other than assertibility-conditions are required to ensure that English is compositionally explicable. As we saw in sentences (62)-(63), there was something about the contexts of the sentences which accounted for the failure of substitutivity *salva veritate*[1] - compound sentences containing doxastic contexts are ones whose constituent sentences contribute their assertibility-conditions as opposed to (or in addition to) their truth-conditions to the sense of the larger sentence. A parallel anti-realist position would be to suppose that there is something about the context in sentences (66)-(67) which accounts for the failure of

---

[1]And hence *salva significatione*.

substitutivity *salva assertione*; i.e. that constituent sentences in hypothetical contexts[1] contribute something other than (or in addition to) their assertibility-conditions to the sense of the larger sentence.

Whether or not the assertibility-conditions for sentence (64) obtain is irrelevant to whether the assertibility-conditions for sentence (66) obtain. In other words, (66)'s warranted assertibility is consistent with either there being evidence which would warrant the assertion of (64) or there not being evidence which would warrant the assertion of (64). As Dummett points out, sentence (66) should not be construed as asserting that if there is evidence which would warrant the assertion that I will marry Jane, then there is evidence which would warrant the assertion that I will no longer be a bachelor; rather it should be construed as asserting that if *it is true* that I will marry Jane, then *it will be true* that I will no longer be a bachelor.[2] In other words, understanding the antecedent and the consequent in terms of their being true, as opposed to their being assertible, is crucial to understanding the entire conditional.[3] Thus, it would not be unreasonable to suppose that sentence (64), as antecedent to sentence (66), contributes its truth-conditions, as opposed to (or in addition to) its assertibility-conditions, to the sense of

---

[1]Or contexts relating phenomena to noumena.

[2]Similarly, 'It is exactly as if a malignant demon is creating the illusion of a chair in front of me and there is no chair in front of me' such not be interpreted as 'It is exactly as if *it is assertible that* a malignant demon ...' but rather as 'It is exactly as if *it is true* that a malignant demon ...'.

[3]Cf. Dummett (1973a) Chapter 13 and Dummett (1991a). It is interesting to note that Dummett had not quite realized this yet in 1963: "a conditional statement, 'If A, then B', means in effect, 'If we had evidence that A, then we should have evidence that B'". ((1963b) p. 371).

(66). We could then construe the basic meaning-theory of semantic anti-realism as:

(AR″) S{a} = (T{a} ∧ A{a})

In other words, under this supposition, (R″) and (AR″) would be equivalent: there would be no difference at all between our new and improved anti-realist and realist theories of meaning.[1] Given that the metaphysical issue we started with is at stake, and realists and anti-realists are fundamentally opposed over that, it would seem that something has gone wrong.

There are at least two initial ways to incorporate the insights above and retain the differences between realism and anti-realism. The first proceeds from a distinction between various sorts of sentential contexts, while the second proceeds from a conception of truth-conditions weaker than that envisioned by the realist.

To begin with, we were led to modify (R) and (AR) from a consideration of how certain sentential contexts appear to affect compositionality. Embedding sentences into doxastic contexts suggests that the assertibility-conditions of the constituent sentence plays a role in determining the sense of the larger sentence, and embedding sentences into hypothetical contexts suggests that the truth-conditions of the constituent sentence plays a role in determining the sense of the larger sentence.

Let me introduce the following terminology. Contexts which allow for substitution of sentences *salva veritate* will be said to be *alethically transparent* whereas those which do not will be said to be *alethically opaque*. Contexts which allow for substitution of

---

[1]Brandom (1976) asserts that the auxiliary Y-conditions must be realist truth-conditions. Appiah (1976) criticises the scantness of his argumentation on this point.

sentences *salva assertione* will be said to be *assertorically transparent* whereas those which do not will be said to be *assertorically opaque*. Thus, doxastic contexts are alethically opaque, and hypothetical contexts are assertorically opaque.

Meaning theories (R), (R″), (AR), and (AR″) all assume that meaning is univocal across sentential contexts; i.e. that a sentence contributes the same thing - namely the type of meaning that it has - to the determination of the meaning of any compound sentence of which it is a constituent. Let us envisage a bipartite meaning-theory which denies this assumption: what a sentence contributes to the meaning of a compound sentence of which it is a constituent is a function of *both* the type of meaning that it has *and* the type of sentential context involved in the compound sentence. Roughly, according to a realist bipartite meaning-theory, a sentence embedded in an alethically transparent context contributes its truth-conditions whereas when embedded in an alethically opaque context contributes its assertibility-conditions; according to an anti-realist bipartite theory, a sentence embedded in an assertorically transparent context contributes its assertibility-conditions whereas when embedded in an assertorically opaque context contributes its truth-conditions.[1]

It would seem that such a bipartite theory, in the face of the inadequacy of (AR), would be desirable to an anti-realist. The contribution of truth-conditions towards the meaning of sentences would be limited only to those classes of sentences involving assertorically opaque contexts. For all other classes of sentences, (AR) would, it seems,

---

[1]Such bipartite theories are, it seems to me, reminiscent of Frege's semantics. Sentences in transparent contexts bear their customary senses whereas when embedded in oblique contexts they bear their indirect senses. See Frege (1892).

be perfectly adequate. Thus, the anti-realist would not need to abandon their basic principles across the board.

However, such bipartite theories are not acceptable. Besides the obvious problem of determining, of any given context, whether it is alethically/assertorically opaque or transparent, it leaves ambiguous the kind of meaning any given atomic sentence has.

Take sentence (60). What type of sentential context does it involve? We can think of the sentence as equivalent to:

68) ϕ(George Gardiner is a retired social worker).

where the sentential predicate "ϕ" is a 'dummy' - it can be thought of as a function which maps the sentence on to itself. The context determined by such a predicate must surely be alethically transparent. Being involved in an alethically transparent context, sentence (60) contributes its truth-conditions *alone* to the determination of its sense. In other words, the sense of an atomic sentence, viewed from a realist perspective, consists exclusively in its truth-conditions. It may *have* both truth-conditions and assertibility-conditions, but its sense is to be identified exclusively with its truth-conditions.

Now reconsider sentence (62). Given that doxastic contexts are alethically opaque, the sentence embedded in (62) - i.e. 'George Gardiner is a retired social worker' - contributes its *assertibility-conditions*, not its truth-conditions, to the sense of (62). But, according to the thesis of compositionality, the sense of a compound sentence is a function of the *senses* of its constituent sentences. Under the realist bipartite meaning-theory, the sense of an atomic sentence is not to be identified with its assertibility-conditions, hence if such a sentence contributes its assertibility-conditions exclusively to

the sense of a compound sentence in which it is embedded, then such compound sentences are not compositionally explicable.

It is no good to maintain that, in such a context, the sense of a sentence is to be identified with its assertibility-conditions, for then the sense of a sentence embedded in a compound sentence will not coincide with the sense of its syntactic counterpart outside of that context. E.g. if the sense of 'George Gardiner is a retired social worker' consists exclusively in its assertibility-conditions in sentence (62) but exclusively in its truth-conditions in sentence (60), then it is simply incorrect to say that sentence (60) is the sentence embedded in sentence (62). Such a suggestion would likewise violate compositional explicability - the sense of a compound sentence is not a function of the senses of its constituent sentences, as its constituent sentences *have* no sense outside of the compound sentence. In a sense, compound sentences could not have constituent *sentences* - i.e. self-contained meaningful linguistic expressions.

Let me try to reinforce this. Contrast the following to sentence (62):

69) Either George Gardiner is a retired social worker or snow is white.

Assuming that ordinary truth-functional contexts are alethically transparent, is sentence (60) the sentence embedded in sentence (62) or is it the sentence embedded in sentence (69)? It cannot be both, for if it is the sentence embedded in sentence (62), then its sense consists exclusively in its assertibility-conditions. If its sense consists exclusively in its assertibility-conditions, then it does not contribute its sense (or any part of it) to the sense of sentence (69); i.e. it cannot be the sentence embedded in (69). If it is the sentence embedded in sentence (69), then its sense consists exclusively in its

truth-conditions. If its sense consists exclusively in its truth-conditions, then it does not contribute its sense (or any part of it) to the sense of sentence (62); i.e. it cannot be the sentence embedded in sentence (62). Suppose it is the sentence embedded in sentence (62), then sentence (69) reveals that English is not compositionally explicable, and the bipartite theory is inadequate. Suppose it is the sentence embedded in sentence (69), then sentence (62) reveals that English is not compositionally explicable, and the bipartite theory is inadequate. Therefore, the bipartite theory is inadequate. Parallel considerations will demonstrate that an anti-realist bipartite theory is also inadequate.

The moral to draw, I suggest, is that if we are forced to acknowledge conditions other than those of truth or assertibility for realist and anti-realist meaning theories respectively as contributing towards a sentence's meaning *in any context*, then we must acknowledge them *in all contexts*.[1]

There is an additional, perhaps more serious, problem with this proposal; namely, the class of assertorically opaque contexts will cut across all sentences. Assertibility-conditions, although significantly unlike truth-conditions, nonetheless obey a form of disquotation. That is, in terms of assertibility-conditions, every sentence S must share the same assertibility-conditions as the sentence formed by enclosing S in quotation marks and appending it to the sentential predicate 'is assertible'. For example, the

---

[1]Such a bipartite theory is envisioned by Brandom. He argues that only sentential contexts which he calls *truth inducing sentential contexts* (or TISC's) require conditions other than assertibility-conditions (which he takes to be truth-conditions) for the determination of the senses of sentences in those contexts. He appears, however, to go part way to accepting my conclusion of the inadequacy of such a bipartite theory in that he believes that *every* sentential compounding device is a TISC. However, this response leaves the status of atomic sentences unclear.

assertibility-conditions for:

70) There are exactly nine planets in our solar system.

must be the same as the assertibility-conditions for:

71) 'There are exactly nine planets in our solar system' is assertible.

There is, therefore, a general method of constructing, for any sentence S, a second sentence sharing the same assertibility-conditions: namely '"S" is assertible'. Secondly, it is a law of logic, whether classical or intuitionistic, that every sentence entails itself. Thus from (70) we can derive:

72) If there are exactly nine planets in our solar system, then there are exactly nine planets in our solar system.

(71), though sharing the same assertibility-conditions as (70), cannot be substituted *salve assertione* for it in the antecedent position in (72). Sentence (72) is assertible in all possible circumstances - including the possible state-of-affairs in which there is, contrary to our present evidence, a tenth planet, whereas its counterpart:

73) If 'there are exactly nine planets in our solar system' is assertible, then there are exactly nine planets in our solar system.

is not; it would not be assertible in that possible circumstance precisely because it would be false in that circumstance. This consequence is perfectly general. For every sentence S there is associated the following tetrad of sentences: (i) S; (ii) 'S' is assertible; (iii) If S then S; (iv) If 'S' is assertible then S; where (i) and (ii) share the same assertibility-conditions while (iii) and (iv) do not. Thus there is, associated with every sentence, an assertorically opaque context.

The consequence this has for the anti-realist is this: given that we must understand

the antecedents of conditional statements in terms of their truth-conditions (as hypothetical contexts are assertorically opaque ones), and every sentence can serve as the antecedent in a true and assertible conditional, every sentence must have a set of truth-conditions which, at least partially, contributes to its meaning. Therefore, there are no classes of sentences for which a theory of meaning making no reference to truth-conditions can be adequate.

What then do we say about sentences in transparent contexts? It seems to me that a realist should hold that in an alethically transparent context, *our* determination of a sentence's truth-conditions will *suffice* for our determination of its meaning but not *exhaust* it. In such contexts, their assertibility-conditions are semantically necessary but epistemically superfluous. Similarly, an anti-realist should hold that in an assertorically transparent context determination of assertibility-conditions will suffice for, but not exhaust, determination of sense. Thus, the fourfold classification of sentential contexts should only be taken as indicating what needs to be involved in our determination of meaning, not as what is involved in meaning *per se*.[1]

The second proposal for distinguishing realism from anti-realism proceeds by making a distinction between truth-conditions as envisioned by realists and those whose acceptance is forced upon the anti-realist. This seems to be Dummett's favoured approach:

I do not want to deny that, from an anti-realist standpoint, we need a notion of

---

[1]Perhaps it would be correct to say that in transparent contexts - whether alethic or assertoric - assertibility and truth-conditions coincide; i.e. to determine the one just is to determine the other.

> truth broader than would result from an equation of A with 'EA'[1], and broader even than entitlement to assert A. This is most clearly apparent when we reflect upon the meaning of conditional statements... Nevertheless, it cannot be consistent to admit a purely realist conception of truth, as attaching to statements just in virtue of how things are in the world (as a matter of fact) quite independently of whether we have any means of knowing them...[2]

What is this notion of truth which is broader than warrant to assert but narrower than recognition-transcendence? Dummett is vague on this point, but he does offer a suggestion in the same letter. Following the above quote he mentions "I have tried to arrive at such a notion in the formulation I gave above (for which I do not claim much merit, or propose to stand by very firmly)." He presents the 'above formulation' as:

> Let us suppose the relevant general notion of truth to be that a statement A is true if we possess or shall come to possess an effective procedure which would, if carried out, lead to a canonical means of establishing A. And let us understand '¬A' to mean that we are or shall come to be in possession of a demonstration that it is impossible to establish A canonically. We could then express the thesis of anti-realism as: $A \rightarrow \neg\neg EA$.[3]

Now, given that we need to interpret the negation constructively, the epistemic functor 'E' is superfluous. Thus, I take his suggestion to be that the double negation (again, understood constructively) of a sentence expresses the truth of the sentence in

---

[1]'E' is Griffin's epistemic functor; 'EA' is read as 'A is canonically established'. (Griffin (1993)).

[2]Dummett (1991a). See also (1990) p. 14: "[Our] linguistic practice cannot be fully described in terms of the notion of justifiability, and that, in achieving a mastery of it, we appear compelled to adopt the conception that to most of the informative sentences of our language ... are associated determinate conditions for their truth that obtain independently of our knowledge or abilities." He does not, however, draw the realist conclusion - we only 'appear' to be so compelled. The necessitated concept of truth must still pass muster in an adequate theory of meaning (i.e. be fully manifestable); it is possible that "we are under [the] illusion that we have acquired a genuine concept or have mastered a coherent linguistic practice." (p. 15).

[3]Dummett (1991a).

the required restricted sense. On the surface this appears adequate. Take sentence (70) for example. It is assertible only under those conditions in which we are in possession of evidence which would warrant its assertion. However, that evidence may be misleading in that there may be future evidence which would overturn its present assertion. But, the assertibility-conditions for:

> 74) It is impossible to disprove that there are exactly nine planets in our solar system.

are more broad than the assertibility-conditions for (70). Nonetheless, (74) does not express a possibly recognitionally transcendent state of affairs.

However, Dummett's suggestion fails. Recall our candidate undecidable sentence (20): "Jones was brave". If (20) is genuinely undecidable, then it is impossible to disprove (for if it were possible to disprove, it would be false and hence not undecidable). Thus, from the undecidability of (20) we can derive its double negation, and hence its *truth* in this restricted sense. In other words, under Dummett's suggestion, all undecidable sentences must (inconsistently) entail their truth. It is fortunate for Dummett that he does not claim much merit for his suggestion.

A consideration of undecidables give rise to another problem - one which casts doubt on whether there can be a notion of truth somewhere between the notion of warranted assertibility and recognition transcendence.[1] At the core of the anti-realist's position is the notion that truth cannot outstrip provability. That is, for any sentence S, if S is true then it must be possible, at least in principle, to verify it:

---

[1]The problem posed for the anti-realist by undecidables in antecedent positions was suggested to me by some remarks in Edgington (1985) 33-52.

> This line of though is related to a second regulative principle governing the notion of truth[1]: If a sentence is true, it must be in principle possible to know that it is true.[2]

In other words, the following is one of the basic theses of anti-realism:

K) $(\forall S)(S \rightarrow$ 'S' is verifiable)

Take some supposed undecidable statement, say (20). Substituting it into (K) yields:

K_{20}) If Jones was brave, then 'Jones was brave' is verifiable.

Anti-realists should, it seems to me, accept (K_{20}) as both true and assertible. However, given that (K_{20}) involves an assertorically opaque context, under the current theory we are required to understand the meaning of its antecedent - 'Jones was brave' - in terms of its truth-conditions. Yet, by supposition, it is undecidable and hence lacks truth-conditions.

Anti-realists are forced to say one of two things. Either they must maintain that undecidables have no truth-conditions or else accept that they do.[3] If they maintain that they have none, then they must abandon the view that sentences in assertorically opaque contexts must be understood in terms of their truth-conditions. But then the anti-realist

---

[1]The first is Principle C, discussed in §3.1.2.2.

[2]Dummett (1976b) p. 99.

[3]Tennant (1981) allows undecidable sentences to have truth-conditions, it is just that we are unable to determine whether they obtain. In other words, for Tennant, undecidable sentences have determinate truth-conditions (which yield their meaningfulness) but lack determinate truth-values. (Tennant's position is criticized in Weir (1983)). Dummett (1959a), on the other hand, denies that they can have truth-conditions, although he softens that claim in (1978) by denying that they can have *realist* (i.e. bivalent) truth-conditions.

owes a new solution to the original problem of how to secure the compositional explicability of natural languages. If they accept that purported undecidables have truth-conditions, then there is no special reason why we cannot grasp the meaning of undecidables by grasping their truth-conditions (with the possible addition of their assertibility-conditions). That is, Dummett's central argument against realist theories of meaning from consideration of undecidables fails. Furthermore, if the purported undecidables are genuine, then it must be the case that we are forever doomed to fail to recognize that their truth-conditions obtain when they do. In other words, the truth-conditions of undecidables, supposing that they have some, must forever be recognitionally transcendent. If there must be some truth-conditions which are recognitionally transcendent, then a realist conception of truth must be admissible.

Thus, I contend that there is good reason to suppose that, whether one is a semantic realist or a semantic anti-realist, as long as one maintains compositionality in one's meaning-theory, one must accept that truth-conditions and assertibility-conditions jointly contribute to sentence meaning. Furthermore, we should think of truth-conditions in the way the realist proposes - as being recognitionally transcendent. Thus, a realist conception of truth is not only vindicated, it may be necessitated.

SECTION II: PUTNAM

1.0 PORTRAITS: METAPHYSICAL AND INTERNAL REALISMS

1.1 Putnam's Metaphysical Realism

Historically, there have been two Putnams. Putnam the Elder - the pre-*Meaning and the Moral Sciences* Putnam - espoused a form of metaphysical realism (then conceived as opposed to verificationism). Reference, he felt, was a correspondence relation (along the lines of Tarski's satisfaction relations[1] but faithful to causal constraints[2]) between our language and extra-linguistic reality[3] which determined the meanings, and consequently the truth-values, of our sentences:

> The essence of the relation is that language and thought do asymptotically correspond to reality, to some extent at least. A theory of reference is a theory of the correspondence in question.[4]

Putnam the Younger, on the other hand, thinks that metaphysical realism is *incoherent*. I have virtually no interest in Putnam the Elder - it is Putnam the Younger, and his rejection of metaphysical realism, which I find of interest. Putnam the Younger (hereafter simply 'Putnam') characterizes metaphysical realism as the position which adheres to the following tenets:

---

[1]Tarski (1931) and (1944).

[2]See, among others, Putnam (1975).

[3]Putnam the Elder actually views truth as a triadic, not dyadic, relation: "...it is very important that a true sentence is *not* one which bears a certain relation to extra-linguistic facts, but one which bears a certain relation to extra-linguistic facts *and to the rest of language*..." Putnam (1960) p. 82.

[4]Putnam (1974) p. 290.

MR_1: The world consists in a fixed totality of mind-independent objects.[1]

MR_2: Truth involves a correspondence relation between linguistic items and objects in the world.[2]

MR_3: There is exactly one true and complete theory/description of the way the world is.[3]

MR_4: Truth is radically non-epistemic.[4]

Berkeley has traditionally been taken as representative of the view known as *idealism*. At its core is the thesis that all existents are either mental entities or entirely dependent upon the mental. Today, Berkeley's idealism has been replaced by Goodman's *irrealism*. Irrealism is, however, idealism 'with a human face'.[5] Goodman replaces Berkeley's reality as mental construction with reality as symbolic (for our purposes, though not exhaustively for Goodman's, *linguistic*) construction: "We can have words without a world but no world without words or other symbols."[6] As he says:

> [In] my view what there is consists in what we make ... Irrealism does not hold that everything or even anything is irreal, but sees the world melting into versions

---

[1]Putnam (1981b) p. 49, (1981d) and (1987b).

[2]For example: "What the metaphysical realist holds is that we can think and talk about things as they are, independently of our minds, and that we can do this by virtue of a 'correspondence' relation between terms in our language and some sorts of mind-independent entities." Putnam (1981d) p. 205. See also Putnam (1976a), (1981b) p. 49, and (1987b).

[3]Putnam (1976a), (1981b) p. 49, (1987a), and (1987b).

[4]Putnam (1976b), (1981b), and (1987b).

[5]Or, if you like, idealism without Berkeley's God.

[6]Goodman (1978) p. 6.

> and versions making worlds, finds ontology evanescent, and inquires into what makes a version right and a world well-built.[1]

On Goodman's view, we *make* the world by various mental processes of conceptualizing: by 'composition and decomposition', 'weighting', 'ordering', 'deletion and supplementation', and by 'deformation'.[2] The obvious question to ask is "from *what* do we 'make' worlds?", and Goodman's answer is that we make worlds from other worlds:

> The many stuffs - matter, energy, waves, phenomena - that worlds are made of are made along with the worlds. But made from what? Not from nothing, after all, but *from other worlds*. Worldmaking as we know it always starts from worlds already on hand: the making is a remaking.[3]

But what of the origins of worlds - or of worldmaking? On the one hand he says that the search for such origins is "best left to theology"[4], but on the other he takes a bolder stand and maintains that there is nothing but worlds of our own making from which we make other worlds:

> But what is *it* that is so organized? When we strip off as layers of convention all differences among ways of describing *it*, what is left? The onion is peeled down to its empty core.[5]

MR_1 is, at its core, anti-idealistic (or what amounts to the same thing (for our purposes), anti-irrealistic), and has, I believe, led most self-proclaimed realists to their

---

[1]Goodman (1984) p. 29.

[2]Goodman (1978) pp. 7-17.

[3]Goodman (1978) p. 6.

[4]Goodman (1978) p. 7.

[5]Goodman (1978) p. 118.

position.[1] Metaphysical realism is thus an anti-idealistic doctrine in this sense: many (if not most) of the things which exist and hence constitute the world are *not* dependent upon the mental (or the symbolic, or the linguistic). This sense can be succinctly expressed in counterfactual form by reflecting upon Goodman's dictum that while we can have words without worlds, we cannot have worlds without words:

MR_{1a}: Even if there had not been words, there would still be a world.[2]

This helps to clarify the metaphysical realist's claim that the world is populated with mind-independent objects, but not the claim that it consists of a *fixed totality* of such objects. To say that the world consists of a fixed totality of objects would be to say that there was exactly one collection of objects which constituted the population of the world (at least at any given time). To use Putnam's example[3] envisage a 'version' of a world W (call it V_1) which consists of three objects: x_1, x_2, and x_3; and a second 'version' - that of the Polish Logician (call it V_2) - which admits objects as sums of other objects consisting of seven objects: x_1, x_2, x_3, x_1+x_2, x_1+x_3, x_2+x_3, and $x_1+x_2+x_3$. Which of V_1 or

---

[1]Devitt, for example, characterizes realism solely as the position which holds that "Tokens of most current common-sense and scientific physical types objectively exist independently of the mental." (1991) p. 23. Field echoes this sentiment: "A unique and mind-independent world is enough [for metaphysical realism]." (1982) p. 554. See also C. Wright (1987) Intro.

[2]Lepore and Loewer (1988) reinforce this counterfactual interpretation of idealism. As we shall see, they provide Putnam with an argument to the effect that internal realism is not idealistic in this sense.

[3]From Putnam (1987a) pp. 32-40 and (1987c).

V_2 constitutes the population of W?[1] The metaphysical realist, according to Putnam, would opt for one over the other - for it is the *world*, which is independent of our conceptualizations, which determines its own population. The sense of this type of realism is a *realism concerning classification*; it maintains that ontological categorization is a mind-independent feature built into the structure of the world itself. The fixed totality of objects, on this view, is a fixed totality of what Putnam calls *self-identifying objects*.[2]

The traditional opponent of *this* sort of realism has been *nominalism*.[3] Once again, Nelson Goodman serves as our modern-day nominalist. He denies that there are fixed and mind-independent categories which serve to classify and categorize mind-independent objects. There is a clear sense, he argues, in which we 'made' the category *constellation* - we *decided* that it would be convenient to group together and label

---

[1]Or a Quinean example: W_1 which is populated with rabbits, W_2 which is populated with undetached rabbit parts, and W_3 which is populated with *both* rabbits *and* undetached rabbit parts. Which of W_1-W_3 is the *actual* world?

[2]Putnam (1981b) and (1981d) (he attributes the term to David Wiggins). As he says, according to the metaphysical realist, the world contains "...Self-Identifying Objects, for this is just what it means to say that the *world*, and not thinkers, sort things into kinds." (1981b) p. 53. I take the phrase to mean that the identity-conditions which serve to individuate objects (as intended in the Quinean slogan 'no entity without identity' (see Quine (1957) and (1966)) are located in the objects themselves and not in our determination of them. Perhaps a better phrase would have been 'self-individuating objects'.

[3]It is, after all, a realism concerning universals. The mereological case, for example, can be construed in a number of ways involving universals: does the (realistically construed) universal 'object' genuinely include mereological sums among its instantiations?; is 'mereological sum' a genuine ontological category (i.e. a universal)?; etc.

particular stars, and it was this mind-dependent process which created the genus *constellation* and its various species (e.g. *Big Dipper*). Goodman's nominalism is a generalized extension of this basic insight - just as we make the category *constellation* on the basis of our mental selection, so too do we make the category *star* on the basis of our mental selection. But to say that we make the category *star* is tantamount, he claims, to saying that we make the stars themselves (and indeed everything, for everything falls under *some* category):

> Now as we thus made constellations by picking out and putting together certain stars rather than others, so we make stars by drawing certain boundaries rather than others. Nothing dictates whether the skies shall be marked off into constellations or other objects. We have to make what we find, be it the Great Dipper, Sirius, food, fuel, or a stereo-system.[1]

Metaphysical realism, again as perceived by Putnam, stands opposed to this sort of nominalism; it is thus an *anti-nominalistic* doctrine.[2] We can succinctly represent this strand using Putnam's terminology:

MR_{1b}: The world is populated with self-identifying objects.

MR_2, MR_3, and MR_4 are intended, by Putnam, to capture the notion of truth and its associated notion of reference as evoked by the metaphysical realist. Basically,

---

[1]Goodman (1984) p. 36. See also Goodman (1980). Notice that, for Goodman at any rate, nominalism entails idealism. Many think Goodman's argument involves an obvious non-sequitur: just because we make the labels it does not follow that we make the things labelled. This problem will be discussed later, but it is important to keep Goodman's idealism in mind - labels (i.e. linguistic symbols) exhaust what there is.

[2]It was Hacking (1983) which first clearly showed me the distinction between the anti-idealistic and the anti-nominalistic aspects of Putnam's metaphysical realism.

Putnam takes that notion to be a model-theoretic adaptation of the basic account provided by Tarski.[1] We are to think of a theory T as an interpreted formal system[2] consisting of two 'parts': (i) a syntactic part characterized as the set of sentences (theorems) derived from some basic set (axioms) in accordance with some particular rules of inference, and (ii) a semantic part characterized in terms of a model M = <D,I> in which I (the interpretation) is a function assigning a unique member of a set of objects[3] D (the domain) to each unique term of T. We can introduce truth of theory model-theoretically (exploiting Tarski's semantic conception) in the following way: a theory T is true (in a model M) just in case all of T's sentences come out true under M. To illustrate, consider a theory T consisting of the single sentence:

1) The Eiffel Tower is in Paris.

and the model M_1 consisting of the domain {The Eiffel Tower, Paris, 1, 2, the spatial relation *is in*, the mathematical relation *is greater than*} and the following interpretation I_1:

a_1) The Eiffel Tower is assigned to 'The Eiffel Tower'
b_1) Paris is assigned to 'Paris'
c_1) the spatial relation *is in* is assigned to 'is in'

We can say, then, that (1) (and hence T) is true *relative* to M_1 - in other words,

---

[1]Tarski (1931) and (1944). See Putnam (1976a), (1976b), (1976c) and (1978) for his explication of Tarski's semantic conception of truth.

[2]The following is a condensed account. See Landini (1987) for a fuller account in the context of Putnam's model-theoretic argument.

[3]I use 'object' untechnically to refer indiscriminately to objects, properties, and relations. An interpretation will assign an object to names, properties to 1-place predicates, and n-place relations to n-place predicates.

it is true-in-M_1. Consider, however, an alternative model M_2 consisting of the same domain and the following interpretation I_2:

a_2) 1 is assigned to 'The Eiffel Tower'
b_2) 2 is assigned to 'Paris'
c_2) the mathematical relation *is greater than* is assigned to 'is in'

(1) (and hence T) is *false* relative to M_2; i.e. it is false-in-M_2. In this example, given that M_1 and M_2 share the same domain, we can see that it is the differing interpretations I_1 and I_2 which alone are responsible for (1) 'passing' from truth to falsity.[1] If we forget that 'true' and 'false' are only abbreviations for 'true-in-M_n' and 'false-in-M_n' we might be misled into thinking that the model-theoretic account of truth leads us into the embarrassing position of supposing (1) to be both true and false. The model-theoretic account would only be problematic if it led us to suppose that some single sentence P were both true-in-M_n and false-in-M_n, but of course any model which led to such a result would be rejected as inadmissible - there are thus at least some *theoretical* constraints on the admissibility of a model.

Now, is (1) *really* true or is it *really* false? At this point the question is misplaced - we as yet have no notion of 'real truth' or truth *simpliciter*. According to Putnam, the metaphysical realist recruits MR_1 at just this point: a sentence is true just in case it accurately describes the nature and geography of the fixed totality of mind-independent objects; in other words, a sentence is true just in case it accurately describes some portion of the world. It does this, according to MR_2, in virtue of its terms standing in

[1]Talking about 'passing from truth to falsity' can be misleading - to say that S 'passes from truth to falsity' is to say only that S is true-in-M_1 and false-in-M_2.

the correspondence relation of reference to objects in the world. It is fairly easy to amalgamate these realist ontological intuitions with the model-theoretic account. Truth is still to be understood as truth-in-M, but the only admissible model is one which takes objects and relations in the world as domain and whose interpretation assigns only those objects and relations in the world to which the sentential terms in fact stand in the correspondence relation of reference. An interpretation which assigns the 'correct' objects to terms is an 'intended' one, and only models whose interpretations are intended are admissible. Thus, assuming the following:

i) 'The Eiffel Tower' refers to The Eiffel Tower
ii) 'Paris' refers to Paris
iii) 'is in' refers to the spatial relation *is in*
iv) 'The Eiffel Tower' does not refer to 1
v) 'Paris' does not refer to 2
vi) 'is in' does not refer to the mathematical relation *is greater than*

I_1 is intended while I_2 is unintended and thus M_1 is admissible while M_2 is inadmissible. On these assumptions (1) is true under an admissible model and hence is true *simpliciter*, or *really* true.

Now given MR_{1b}, there is only one way that the world is divided up. Hence, for any two equally well supported theories differing in ontology, at most one of them can be true; i.e. only one of their differing ontologies can correctly describe the way the world carves itself up. Thus, understanding truth in terms of MR_2 and metaphysics in terms of MR_1, it would seem that the metaphysical realist is committed to MR_3. Putnam is fond of expressing this by saying that the metaphysical realist is committed to there being a unique and privileged position from which an accurate description of the world

would be given - a God's Eye View.[1]

Moreover, given that truth is primarily to be understood by appeal to the notion of reference, if there can at most be one true and complete theory of the way the world is, there can at most be one genuinely referential relation between any term and particular objects in the world. Put in Quinean terms, MR_{1b} entails that the world either is populated in part with rabbits or with undetached rabbit parts (but not both, understanding them to be ontologically distinct). If 'rabbit' refers at all, it will refer to one to the exclusion of the other. Suppose the world is populated with rabbits, and as such 'rabbit' refers to them. In that case, whatever relation 'rabbit' has to undetached rabbit parts, that relation cannot be the reference relation (where the reference relation is a satisfaction relation in an admissible model). Thus, MR_3 entails:

> MR_{3a}: The correspondence relation between words and objects which constitutes reference is unique.

Finally, because truth is defined as a semantic relation whose primary relata are mind-independent objects, it is primarily a metaphysical notion. On this understanding, Putnam contends, the metaphysical realist is committed to the claim that even a theory which is epistemically ideal might turn out to be false - the terms of a theory as highly confirmed as can be might fail to refer to actual entities in the world. Take, for example, a theory which postulates the existence of mind-independent material objects to explain the regularities in our phenomenal field. No matter how well confirmed that

---

[1]See for example Putnam (1986) and (1987b). In (1981b) p. 49 he labels metaphysical realism the *externalist* perspective "because its favourite point of view is a God's Eye point of view."

theory may be, Putnam's metaphysical realist says, it may be false in that its terms may fail to refer to actual items in the world - it *may* be the case that there are no tables and chairs for 'tables' and 'chairs' to refer to, but only the mind-influencing activity of some malevolent demon. Put in model-theoretic terms, the model M under which the material object theory would be true consists of an interpretation which maps its terms onto (unknown to us) non-existent items and hence is (unknown to us) inadmissible. The theory may be true-in-M, but as M is inadmissible, it fails to be true *simpliciter.*

Of course, this result is expected given MR_2 - truth *simpliciter* depends *exclusively* upon whether certain non-epistemic facts concerning the relation of language to the world obtain. Truth *simpliciter* is here being characterized in terms of truth under only admissible models. Thus, we should expect the metaphysical realist to assert that the admissibility of a model is a non-epistemic matter - it depends *exclusively* on whether certain non-epistemic facts obtain regarding both the objects in its domain (i.e. they must exist in the world) and its interpretation (the satisfaction relation it determines must, in fact, be the reference relation). Thus, it does not seem unreasonable for Putnam to maintain that the metaphysical realist is committed to the radical non-epistemic nature of truth: i.e. to MR_4.

1.2 Internal Realism

Putnam characterizes his alternative to metaphysical realism - internal realism - in the following synopsis:

> The perspective I shall defend has no unambiguous name. It is a late arrival in the history of philosophy, and even today it keeps being confused with other

> points of view of a quite different sort. I shall refer to it as the *internalist* perspective, because it is characteristic of this view to hold that *what objects does the world consist of?* is a question that it only makes sense to ask *within* a theory or description. Many 'internalist' philosophers, though not all, hold further that there is more than one 'true' theory or description of the world. 'Truth', in an internalist view, is some sort of (idealized) rational acceptability - some sort of ideal coherence of our beliefs with each other and with our experiences *as those experiences are themselves represented in our belief system* - and not correspondence with mind-independent or discourse-independent 'states of affairs'. There is no God's Eye point of view that we can know or usefully imagine; there are only the various points of view of actual persons reflecting various interests and purposes that their descriptions and theories subserve.[1]

In a nutshell, it denies MR_1 and MR_4, and offers in their place:

> IR_1: The world consists of theory-dependent objects.
>
> IR_2: Truth is (idealized) justification.

It accepts, as I shall point out, MR_2, but only after the ontological sense of its 'objects in the world' component is sufficiently modified according to IR_1. On the other hand, again modifying the sense of 'the world' along the lines of IR_1, it replaces MR_3 with:

> IR_3: There are a plurality of complete and true theories/descriptions of the way the world is.

and consequently replaces MR_{3a} with:

> IR_{3a}: There are a plurality of genuine reference relations between words and objects in the world.

Goodman's irrealism, as we have seen, is committed to there being no reality outside of various 'versions' - i.e. various symbol systems conventionally 'created' by

---

[1]Putnam (1981b) pp. 49-50.

humans to serve various human interests. A consequence of this view is the truth of various counterfactuals like "If we had not developed star-theory, there would be no stars"[1]; the truth of these counterfactuals being tantamount to an idealistic anti-realist conception of ontology.

Internal realism is idealistic in one sense: the nature and existence of objects is not entirely independent of human theorizing and describing. It is not, however, idealistic in the stronger sense; Putnam is quite insistent that we did not make the stars:

> One perfectly good answer to Goodman's rhetorical question "Can you tell me something that we didn't make?" is that we didn't make Sirius a star. Not only didn't we make Sirius a star in the sense in which a carpenter makes a table, *we didn't make it a star.* Our ancestors and our contemporaries (including astrophysicists), in shaping and creating our language, created the concept *star*, with its partly conventional boundaries, with its partly indeterminate boundaries, and so on. And that concept *applies* to Sirius. The fact that the concept *star* has conventional elements doesn't mean that *we* make it the case that the concept applies to any particular thing, in the way in which we made it the case that the concept "Big Dipper" applies to a particular group of stars. The concept *bachelor* is far more strongly conventional than the concept *star*, and that concept applies to Joseph Ullian, but our linguistic practices didn't make Joe a bachelor...[2]

Putnam's metaphor for his basic ontological commitment is that "the mind and the world jointly make up the mind and the world".[3] In other words, he does not endorse an anti-realism of the Goodmanian idealistic variety (it is partially this which allows

---

[1]It would be a mistake, it seems to me, to think of 'star-theory' mentioned in the counterfactual entirely in astrophysical terms - I don't think Goodman would deny that there were stars in the time of Aristotle (he might, however, be willing to say that Aristotle's stars are not the same as ours). In any event, a more charitable reading of the counterfactual might be something like "If star-talk wasn't part of the description of our world-version, there would be no stars."

[2]Putnam (1992) pp. 114-115.

[3]Putnam (1981b) p. xi.

internal realism to be a *realism*), but he does seem to endorse an anti-realism of the nominalistic sort. We do (via our theorizing), he more or less says, 'make up' the general ontological categories, but then it is the nature of the objects themselves which sorts them into those various categories. More or less ... but this is a bit misleading as well. There is, he suggests, a conventional aspect to objects (i.e. an aspect determined by human theorizing) as well as a non-conventional aspect (i.e. an aspect determined by the world), but it is a mistake to think that these different aspects can be clearly delineated: "To try to divide the world into a part that is independent of us and a part that is contributed by us is an old temptation, but giving in to it leads to disaster every time."[1] The ontological anti-metaphysical-realism of internal realism, then, amounts primarily to a rejection of MR_{1b}:

IR_{1a}: There are no self-identifying objects.[2]

Brown (1988) argues that internal realism *is* a form of idealism. His argument issues, I believe, from a failure to distinguish idealism from nominalism in the manner suggested by Hacking (1983) and Lepore and Loewer (1988). In other words, he equates idealism not with the strong view that objects are entirely mind-dependent (thought to be captured by the truth of various counterfactuals) but rather with the weaker view that

---

[1]Putnam (1992) p. 58.

[2]See primarily Putnam (1981d) for his argument against there being self-identifying objects. In a nutshell, his argument is that the thesis of self-identifying objects is bound up with MR_3 (the "belief in one true theory requires a *ready-made* world: the world itself has to have a 'built-in' structure since otherwise theories with different structures might correctly 'copy' the world (from different perspectives) and truth would lose its absolute (non-perspectival) character." p. 211). See §2.3 for an extended discussion of his argument.

there are mind-dependent aspects to objects. If the preceding has been a correct account of internal realism then, according to it, there is always a mind-dependent aspect to any object - namely its description in terms of falling under a general category. Now Brown recognizes that there is an element of human choice in the ontological categories that we in fact employ, but he resists the nominalist conclusion that they are therefore 'made up' by us. Instead he proposes that the world *itself* divides up in many distinct and non-overlapping ways. Human choice merely decides *which* of these mind-independent divisions we find convenient to use. He admits, however, that the price of maintaining the anti-nominalistic components of metaphysical realism is to reject MR_3 (and consequently MR_{3a}). While this suggestion is interesting, it is really unhelpful in terms of offering a defense of metaphysical realism as characterized by Putnam: if MR_3 has to go, it makes no real difference whether it goes while retaining MR_{1b} or not. In Brown's defense, he does argue that losing MR_{3a} - the determinacy of reference - is not as serious for the metaphysical realist as Putnam supposes. In any event, for the purpose of arriving at an adequate understanding of how Putnam intends internal realism, there is a clear sense in which it is nominalistic but not idealistic.

Putnam notes that "the first clear indication that a coherent alternative to both the correspondence theory and the pure disquotational theory might be available came from the writings of Michael Dummett."[1] There is no need to go over Dummett's views on truth; it is sufficient to point out that for Dummett truth is primarily an epistemic

---

[1]Putnam (1983b) p. xvi.

notion; truth-conditions are justification-conditions. Putnam admits to a modified acceptance of this basic (though somewhat inaccurate) Dummettian position: "Whereas Dummett identifies truth with justification, I treat truth as an *idealization* of justification."[1] The advantage of viewing truth in this way, Putnam claims, is that it allows a retention of three intuitive components to the notion of truth: (i) a statement may be (ordinarily or even highly) justified but not true, (ii) truth is stable over time, and (iii) truth does not admit of degree.[2] While ordinary justification changes over time (people 3,000 years ago were justified in believing the earth was flat whereas we are not) and admits of greater and lesser degrees, idealized justification does not.[3]

Recall that idealism has been characterized as the position asserting the truth of various counterfactuals of the form "If there were no O-talk, then there would be no Os". By construing truth in terms of idealized justification, Lepore and Loewer (1988) maintain that internal realism is able to avoid this idealist tendency:

> [It] is no consequence of IR that counterfactuals like "If we had not constructed the theory of electrons, then there would be no electrons" are true. In fact, on Putnam's, but perhaps not on Dummett's, account the counterfactual "Even if we had not constructed the theory of electrons there would be electrons" is justified. So we have reason to think it true. In general, whenever S is justified, so is "Even if I had not thought of S it would be justified."[4]

Their argument, however, is somewhat cryptic. It appears to be something like the

---

[1]Putnam (1976c) p. 84. See also (1976c), (1982), (1983b), (1986), and (1987a).

[2]Putnam (1981b) pp. 55-56.

[3]See C. Wright (1992) Ch. 2 for a strong but anti-realistically sympathetic criticism of Putnam's elucidation of the truth-predicate.

[4]Lepore and Loewer (1988) p. 470.

following: anti-idealism is the view that such counterfactuals as "Even if we had not constructed star-theory, there would be stars" are true. Given that truth is (idealized) justification, anti-idealism is tantamount to the view that those counterfactuals are (ideally) justified. Consider the consequent of such a counterfactual: "There are stars". Assuming that *this* sentence is ideally justified, it follows that it is true. From any sentence P, given the propositional calculus, one is able to derive $Q \rightarrow P$. Thus, given that "There are stars" is true, it follows that, for any statement Q, "If Q then there are stars" is true. Let Q be: "We have not constructed star-theory". Then, "If we have not constructed star-theory, there would be stars" (or in more colloquial English: "Even if we had not constructed star-theory, there would still be stars") is true.[1] Generalizing, as we have good reason to suppose that the various counterfactuals are true, we have good reason to reject idealism.

There are two curious points about this argument. In the first place, it does not depend upon Putnam's identification of truth with idealized justification - contrary to Lepore and Loewer's claim, it works just as well with an identification of truth with warranted assertibility - nothing is lost if we deleted 'idealized' from every occurrence in the argument. In fact, it will work for any conception of truth which allows the consequent and the inference $P \vdash Q \rightarrow P$. Secondly, the argument turns on treating counterfactuals exactly as if they were straightforward material implications - it is only understanding them in such a way that warrants the inference. But counterfactuals do

---

[1]As they say: "In general, whenever S is justified, so is 'Even if I had not thought of S it would be justified'." (Lepore and Loewer (1988) p. 470).

not display the same logical properties as ordinary conditionals. The argument just seems to be a logical sleight of hand.[1]

Regardless of whether the anti-idealistic stance can be derived from Putnam's basic conception of truth, it remains the case that Putnam does not intend internal realism to be a form of idealism. That much seems clear from his rejection of Goodman's irrealism. To reiterate his slogan, the mind and the world jointly make up the mind and the world.

Now while it is somewhat unclear what Putnam means by 'idealized justification', it is clear that he does not intend it to mean anything like 'the justification an omniscient God would have from her God's Eye View' given his rejection of MR_3. I suggest we understand an idealization of justification procedures along the lines suggested for understanding undecidables discussed in the Dummett section - i.e. as an extension of our current justification procedures. A sentence would be ideally justified, then, if it were something like 'as rationally acceptable as is humanly possible.'

According to the rejection of MR_3, it is possible for there to be more than one

---

[1]It is perhaps possible to repair their argument. Jill Leblanc has pointed out that the argument shifts back and forth between the indicative "there are stars" and the subjunctive "there would be stars". If we read the basic denial of idealism as a hybrid subjunctive/indicative conditional, such as "Even if we had not constructed star-theory, there are stars", then the argument seems to go through. The hybrid conditional is, however, an odd construction - it is not clear how it should be interpreted, and consequently not clear whether it can be treated as a material conditional for implicational purposes. Secondly, given that idealism has been characterized as the position which accepts the truth of various *counterfactuals*, an anti-idealism should, one would expect, be a position which rejects the truth of such *counterfactuals*.

true and complete theory or description of the world. A complete theory, in Putnam's sense, seems to be a general or overarching theory which includes all particular theories. Thus, a complete theory of the world would include astronomical theory, paleontological theory, fluid mechanical theory, etc. and a *true* complete theory would include only true component theories. It is Putnam's view that rationality or justification are *themselves* theoretical - i.e. a complete theory would *include* an account of the conditions under which a sentence is rationally acceptable. There is no reason to think, and, he argues, every reason not to, that standards of rationality will remain constant across various true and complete theories. In other words, rational acceptability is itself a theory-relative notion. Thus, what counts as ideally justified will vary from theory to theory, and thus what counts as true will vary from theory to theory as well. In other words, internal realism, in addition to holding an epistemic conception of truth, holds a relativistic one as well:

> The "internal realism" I have defended has both a positive and a negative side. Internal realism denies that there is a fact of the matter as to which of the conceptual schemes that serve us so well - the conceptual scheme of commonsense objects, with their vague identity conditions and their dispositional and counterfactual properties, or the scientific-philosophical scheme of fundamental particles and their "aggregations" (that is, their mereological sums) - is "really true". Each of these schemes contains, in its present form, bits that will turn out to be "wrong" in one way or another - bits that are right and wrong *by the standards appropriate to the scheme itself* - but the question "which kind of 'true' is really Truth" is one that internal realism rejects.[1]

His chief argument for this is of the interest-relativity of explanation. Willie Sutton, a famous bank robber, was asked why he robbed banks. His answer was

---

[1]Putnam (1987c) p. 96.

"Because that is where the money is."[1] Whether it is reasonable to see Sutton's answer as constituting an explanation of his actions will depend upon what our interests are. If the interrogator is a psychologist studying criminal behaviour (i.e. if she means to ask "Why do you rob banks *at all*?"), then the answer will be unexplanatory. On the other hand, if the interrogator is an apprentice robber (i.e. if he means to ask "Why do you rob banks instead of, say, trains?") the answer will be explanatory.

To broaden the example, we tend to think of explanations largely in causal terms; e.g. if we ask for an explanation of why the boiler exploded we are generally asking for the cause of the boiler's exploding. Now Putnam points out that there is a conceptual distinction between the notion of a total cause and our ordinary notion of a cause.[2] The total cause of the boiler's exploding include the sticking of the valve, the corrosion of the boiler's body, the *absence* of additional reinforcement, the *absence* of holes in the body, the internal pressure of the water, the altitude of the boiler, etc.: if *any* of those conditions had been different, it is very likely that the boiler would not have exploded. However, it is part of our normal notion of a cause, Putnam argues, that we select *one* (or a few) of the elements of the total cause and relegate the rest to 'background conditions'; we speak of the sticking of the valve as *the* cause. In this case, we would not find it helpful to be told that the fact that there were no holes in the body of the boiler explains why it exploded. The point is that what we find *rationally acceptable* as a cause or an explanation depends upon what our particular interests are. Thus there is always

---

[1]Putnam (1976a) p. 44.

[2]Putnam (1984a) and (1992).

a conventional, normative, non-mind-independent aspect to any notion of rational acceptability. Thus, there is no such absolute notion, and thus no single account for any true and complete theory to incorporate. Equating truth with idealized justification, and admitting notions of justification to be interest- or theory-relative, it follows that truth itself is interest- or theory-relative.[1]

However, Putnam is insistent in denying that the relativistic aspects of internal realism generate the three traditional problems of relativism: (i) being incoherent[2], (ii) being incapable of distinguishing between truth and belief[3], and (iii) allowing that 'anything goes'.[4]

Since at least the time of Protagoras and Plato, it has been common to argue that the fundamental relativistic thesis, more or less expressible as "All truths are true relative to particular conceptual schemes", must be a general truth which is not *itself* true

---

[1]All of this fits in nicely with his rejection of MR_{3a} and his modified acceptance of MR_2 - if truth involves a correspondence relation, but there is no unique correspondence relation, we should not expect truth to be unique. More precisely, if particular theories *themselves* select particular correspondence relations for the reference relation (i.e. reference is theory-relative), then truth itself will be theory-relative.

[2]As we shall see, Putnam presents at least three such 'incoherence' arguments.

[3]"[The] relativist cannot, in the end, make any sense of the distinction between *being right* and *thinking he is right*..." (Putnam (1981b) p. 122).

[4]"Internalism is not a facile relativism that says 'Anything goes'." (Putnam (1981b) p. 54). When dismissing the self-refutation counter-argument, he equates (traditional) relativism with the view that "*no* point of view is more justified or right than any other." ((1981b) p. 119). On an aside, it is a common attitude that Putnam's internal realism is an attempt to find a 'middle position' between the 'extremes' of relativism and metaphysical realism (see, for example, Throop and Doran (1991)). Harman (1982) argues that the two positions are not at extremes, finding both metaphysical realism and (moral) relativism attractive. See Putnam (1982) for a response.

relative to a particular conceptual scheme.[1] Thus, the thesis of relativism, if true, would imply its own falsity.[2] Putnam's counter is that it is a mistake to view the thesis of relativism as a general or meta-theoretic pronouncement upon all ordinary (object-level) truths. That would be to resurrect, he argues, a God's Eye View - i.e. a privileged position external to any particular theoretical or conceptual scheme we happen to occupy from which it would be capable of surveying those actual schemes - it would be say that "from a God's-Eye View, there is no God's-Eye View"[3]:

> Relativism, just as much as Realism, assumes that one can stand within one's language and outside of it at the same time. In the case of Realism this is not an immediate contradiction, since the whole content of Realism lies in the claim that it makes sense to think of a God's-Eye View (or better, a "View from Nowhere"); but in the case of Relativism it constitutes a self-refutation.[4]

But, with the rejection of MR_3 comes a rejection of there being a God's Eye View - there is simply no position from which to self-refutingly assert the thesis of relativism. Nonetheless it *can* be part of a particular conceptual scheme that there can be *other* conceptual schemes which admit of different correspondence relations and hence different truths. That truth, though, is a truth *within* a theory: "The important point to

---

[1]See Preston (1992) for an illuminating classification of relativisms, as well as Putnam's position on this issue.

[2]Garfinkel's 'one-liner': "Alan Garfinkel has put the point very wittily. In talking to his California students he once said, aping their locutions: 'You may not be coming from where I'm coming from, but I know relativism isn't *true for me*'... If any point of view is as good as any other, then why isn't *the point of view that relativism is false* as good as any other?" (Putnam (1981b) p. 119). See Johnson (1991) for a relativist defense. Putnam himself does not place much weight on this argument.

[3]Putnam (1987b) p. 25.

[4]Putnam (1987b) p. 23.

notice is that if all is relative, then the relative is relative too."[1]

In "Why Reason Can't Be Naturalized",[2] Putnam offers an argument to the effect that (cultural) relativism not only is self-refuting but is also *incoherent.* (Cultural) relativism finds it origin in the recognition of other cultures with norms, practices, and beliefs distinct from our own. At its heart is a claim about truth and meaning: "'P' is true", uttered by a member of a culture C, means "'P' is true relative to the norms and standards of C". Let R.R. and Karl be members of American and German culture respectively. When Karl asserts "Schnee ist weiss", what he means (according to the relativist schema) is that snow is white relative to the norms of German culture. R.R., however, *cannot* so understand Karl's utterance - given the schema, he can only interpret Karl's utterance as: "'Schnee ist weiss' is true relative to the norms of German culture" is true relative to the norms of American culture.

In general, if R.R. understands *every* utterance *p* that *he* uses as meaning 'it is

---

[1]Putnam (1981b) p. 120. See also Putnam (1981a), (1981d), and (1987b). Two asides: Nicholas Griffin points out that this leaves open the possibility that 'metaphysical realism is true-according-to-metaphysical-realism' is a truth *within* internal realism (except that Putnam maintains that metaphysical realism is *unintelligible*) but that metaphysical realism does not repay the compliment. Moser (1990) sees a dilemma facing the internal realist. Putnam's way out of the first relativist problematic is to deny that there is a 'God's Eye View' from which to issue the relativist thesis (and thus, it does not apply to itself). It seems, then, that the argument rests upon the claim "There is no 'God's Eye View'", or, as Moser would put it "any claim of how things are (theory-independently) is unintelligible". But, is *this* a claim about how things are (theory-independently)? If it is, and Putnam's proposed escape from the first relativist problematic is successful, then the claim (and hence Putnam's defense) is unintelligible by its own lights. If it is not, then what sort of claim is it? (It cannot, for example, be a meta-theoretical claim concerning all lower-level theories, for, according to Putnam, there is no such over-arching meta-theory.)

[2]Putnam (1981c).

> true by the norms of American culture that *p*', then he must understand his own hermeneutical utterances, the utterances he uses to interpret others, the same way, no matter how many qualifiers of the 'according to the norms of German culture' type or however many footnotes, glosses, commentaries on the cultural differences, or whatever, he accompanies them by.[1]

Consequently, Karl's culture - or any other - becomes a mere "logical construction out of the procedures and practices of American culture".[2] The (cultural) relativist, then, cannot make any genuine sense of there *being* other cultures, as distinct and independent from her own, and thus the relativist ends up denying the position from which she began.[3]

There is a third, somewhat cryptic, argument for the incoherence of relativism. Take the relativist's construal of truth: 'P' is true (when uttered in a culture C) just in case 'P' conforms to C's norms and beliefs. It is, Putnam maintains, a *reductionistic* thesis in that it attempts to reduce the *normative* notion of truth to a *non-normative* notion of agreement with one's culture. But, *thinking*, as a human propensity, essentially involves attempting to 'get things right' and to avoid 'getting things wrong' (both normative notions) - i.e. it aims at truth and the avoidance of error. If 'truth' loses its normative content (as proposed by the relativist), then thinking (a propensity which essentially involves normativity) would reduce to "making noises in counterpoint or

---

[1]Putnam (1981c) p. 237.

[2]Putnam (1981c) pp. 237-238. The position is, Putnam maintains, strictly analogous to that of the solipsist. See Solomon (1990) for a critique of Putnam's argument.

[3]Putnam is a little vague on how his internal realism is able to escape this defect, but it would seem that by castigating a 'God's Eye View' there is no position from which to accept the relativist's problematic meaning schema.

chorus" - i.e. humans would merely be noise emitting animals as opposed to rational thinking creatures.[1] Consequently, relativism (or any position) would cease to be a *rational thesis.* Putnam's internal realism avoids this consequence by accepting a normative conception of truth - truth is an idealization of justification. The relativism enters not with a reductive notion of truth but with the recognition that standards of justification are themselves *internal* to points-of-view.[2]

Regarding the second traditional worry, it is commonplace to view relativism as the thesis that a sentence P is true only relative to a conceptual scheme C. In particular, according to traditional relativism, P is true *for* some person S just in case P can be derived from S's conceptual scheme C. The problem is that a conceptual scheme appears to be nothing other than a set of beliefs - each person's 'web' of belief constitutes their conceptual scheme. Now, if P can be derived from a person S's conceptual scheme C (i.e. P is true), then (assuming an artificial degree of rationality), S will believe that-P. Secondly, if S believes that-P, then obviously the belief that-P is contained in C, and C will trivially entail P. Thus, P is true-relative-to-C just in case S believes that-P. There can be no distinction, given relativism, between P being true and believing that-P.[3]

---

[1]Putnam (1981c) p. 235. Thus, not only does he find relativism incoherent, he also finds it dangerous.

[2]See Throop (1989), Throop and Doran (1991), and Johnson (1991) for commentaries on this argument, as well as Putnam's (1991) responses.

[3]This argument has been couched in what Preston (1992) calls *subjective* relativism. It can also be run in terms of a cultural relativism - simply replace S with some society R.

In order to combat this problem, we need to reflect that, according to internal realism, truth is identified with *idealized* justification. Ordinarily (again assuming an artificial degree of rationality), one will believe that-P if one feels oneself to be justified in believing it. Understanding this in terms of the 'acceptable' relativism of internal realism, one will believe that-P if one feels themselves to be justified *in accordance with the standards of justification internal to one's conceptual scheme* in believing it. But, those internal standards themselves can be idealized - P will be true on Putnam's account only if it can be so ideally justified. Thus, there is a clear sense in which one can believe some statement on the basis of ordinary standards of justification without that statement being ideally justified, *even if* standards of justification are internal to conceptual schemes.[1]

Finally, relativism will only have an 'anything goes' consequence if any conceptual scheme is just as good as any other. Putnam denies this. A conceptual scheme is at least in part a reflection of our interests, but (i) we are not free to choose our interests at will, and (ii) interests themselves are subject to normative criticism - criticism which is *internal* to a conceptual scheme:

> Every culture has norms which are vague, norms which are unreasonable, norms which dictate inconsistent beliefs... Our task is not to *apply* cultural norms, as if

---

[1]Preston (1992) p. 65 sums this up nicely: "How can the objective relativist make sense of the distinction between being right and thinking he is right? All he needs is this: the idea that the relative truth is idealized rational acceptability within a framework, and the idea that the objectivity of rational acceptability within a framework consists in the objectivity of what really *does* follow from the framework principles." Preston (pp. 67-68) relates this to 'Garfinkel's one-liner'; when Garfinkel *says* (i.e. believes) 'relativism isn't true for me', it does not follow (for the relativist) that his claim is true, and thus does not refute relativism.

> they were a computer program and we were the computer, but to interpret them, to criticize them, to bring them and the ideals which inform them into reflective equilibrium.[1]

In summary, then, internal realism rejects the metaphysical realist's identification of 'the world' with a collection of mind-independent objects. According to the metaphysical realist, the mind-independence of objects consists in the claim that their existence is independent of our theorizing or conceptualizing and in the claim that the nature of particular objects is determined (at least in part) by the theory-independent ontological categories into which they fall. Internal realism is anti-idealistic in the sense that it agrees that the existence of objects in the world is not (entirely) dependent upon our conceptualizing - there would be a world even if there were no words (this is one of the reasons why it counts as a type of *realism*). On the other hand, it is nominalistic in that it denies that there are self-identifying objects - *we* sort things into ontological categories *as* we conceptualize, and there are no external, mind-independent, categories which *our* categorizing must be faithful to. But, this sort of nominalism does give rise to at least an appearance of idealism. To borrow a phrase from Quine, no entity without identity. *We* give the identity conditions to objects in the sense that *we* impose ontological categories *as* we conceptualize and thus at least in *this* sense we make

---

[1]Putnam (1981c) pp. 239-240. Or, if one prefers a less subjectivist approach, a conceptual scheme is a reflection of our communal interests (which, by an large, are part of an *inherited* tradition), of which we are (by and large) not free to choose, and that such traditions themselves can be criticized (both from within and without). (Putnam (1981c), see also Putnam (1992)). Putnam's relativism boils down to what Preston (1992) calls a *non-total* (the thesis of relativism is not issued from a transcendental point) *objective* (it employs an (objective) notion of truth as idealized justification) relativism.

objects. To cut through the apparent tension, Putnam invites us to think of it this way: there is a metaphysical aspect to objects - unconceptualized reality does contribute to the nature of the world - but there is also a theoretical or conventional aspect to objects: the mind and the world jointly make up the mind and the world. It is a mistake, however, to think that one can coherently separate the conventional aspects from the non-conventional. Our very concepts of reality and the world are invariably bound up with our practice of conceptualizing.

The internal realist can, however, continue to think of truth as involving a correspondence between language and the world (and this is the second reason why it counts as a type of realism), but if we think of 'the world' as not being entirely mind-independent, then we cannot think of truth as radically non-epistemic. Truth is inherently an epistemic notion, but should not be identified with ordinary justification. To retain our intuitive notion that truth is stable and distinguishable from belief, we should identify it rather with *idealized* justification.

If we also accept that there are no absolute and eternal standards of justification - i.e. if we admit that such standards are always themselves *internal* to a particular theory or conceptual scheme - then we will admit at least the possibility that there can be more than one true and complete theory of the world. To reinforce this, if we grant that standards of justification are theory-relative, and truth is an idealization of justification, then truth itself will be theory-relative; there is no particular reason why distinct theories could not both be true - i.e. true as understood *internally* to each theory.

Finally, if objects are theory-relative, then correspondence, which take objects as

relata, will also be theory-relative. There is no unique and privileged relation between language and the world which constitutes the reference relation which defines the truth-predicate, because there is no unique and privileged way the world is. As he says: "We don't have notions of the 'existence' of things or of the 'truth' of statements that are independent of the versions we construct and of the procedures and practices that give sense to talk of 'existence' and 'truth' within those versions."[1] So, internal realism holds that objects, standards of justification, reference, and truth are all theory-relative without, Putnam claims, being committed to the traditional problems of relativism.

1.3 Putnam's Strategy

Putnam's basic strategy is to argue against metaphysical realism in its entirety by offering arguments against its specific theses. Thus, he thinks that if he can succeed in rejecting, say, MR_3, then he will have succeeded in rejecting metaphysical realism. This strategy presupposes that the four main theses depend on each other. As he says, MR_1, MR_2, and MR_3 "do not have content standing on their own; each leans on the others and on a variety of further assumptions and notions."[2]

If truth involves a correspondence relation between language and the world (MR_2), and the world consists of a fixed totality of mind-independent objects (MR_1), then at most there can be one true and complete theory of the way the world is (MR_3). For if there were more than one true theory, and they were not merely notational variations

[1]Putnam (1981c) p. 230.

[2]Putnam (1982) p. 31.

of each other, then they would not agree on the way the world is. Given the truth of each, it would follow that there is not one unique way the world is, and MR_1 would have to be given up. Thus, if MR_3 is incorrect, then MR_1 or MR_2 must also be incorrect. Secondly, if there can be more than one correspondence relation each of which equally constitutes a reference relation, then as truth is defined in terms of that reference relation, it would follow that there can be more than one true and complete theory. In other words, the incorrectness of MR_{3a} would entail the incorrectness of MR_3, which would in turn entail the incorrectness of MR_1. Finally, to say that truth is a relation between language and theory-relative objects is to deny that truth is radically non-epistemic; neither language nor theory are mind-independent entities. Thus, the rejection of any of MR_1, MR_3, or MR_{3a} would entail the rejection of MR_4.

Putnam's three main specific strategies are to argue against MR_3, MR_{3a}, and MR_4.[1] His attack on metaphysical realism can, I believe, be pared down to three specific arguments: (i) the argument from conceptual relativity aimed at MR_3, (ii) the Model-Theoretic argument aimed primarily at MR_{3a} (but also derivatively at MR_3 and MR_4), and (iii) the 'Brain in the Vat' argument aimed primarily at MR_4.

It is my contention that none of these arguments succeed, and thus Putnam is left with no particular reason to accept internal realism over metaphysical realism. In the following, I will consider each of these arguments separately, though if Putnam is right about their mutual interrelationships it is impossible to keep them entirely separated.

---

[1]He does offer arguments nominally against some of the other theses; e.g. MR_{1b} (i.e. the anti-nominalistic thesis). However, upon careful reflection it becomes apparent that they are really variants of one of the main arguments against MR_3, MR_{3a}, and MR_4.

2.0 ARGUMENTS

2.1 The Model-Theoretic Argument

2.1.1 The Argument[1]

Let T be a theory which is epistemically ideal (but is not assumed to be true in the metaphysical realist's sense) in that it embodies two features: (i) it satisfies all theoretical constraints (i.e. it is "complete, consistent, ... 'beautiful', 'simple', 'plausible', etc."[2]) and (ii) it satisfies all operational constraints (i.e. all of its sentences parallel certain experiential facts - e.g. "if 'there is a cow in front of me at such-and-such at time' belongs to T, then 'there is a cow in front of me at such-and-such a time' will certainly *seem* to be true - it will be 'exactly as if' there were a cow in front of me at that time."[3]) Because T is consistent it is guaranteed to have a model. Assuming that the model is of infinite cardinality, by the Löwenheim-Skolem Theorem T is also guaranteed to have a model of every infinite cardinality. Select one of these models M which is of the same cardinality as the world, and map its terms one-one with objects in the world. The result is a satisfaction relation SAT between T and the world which we can use to define a truth-predicate for T as TRUE(SAT).[4] In other words, T is TRUE(SAT). If we then

---

[1]The following account is an amalgamation of Putnam's four presentations in (1976b), (1980), (1981b), and (1989). The last is largely concerned with tracing the connections between its purported results and some Quinean theses.

[2]Putnam (1976b) p. 125.

[3]Putnam (1976b) p. 126.

[4]There is reason, I am told, to believe that a physicalist model of the world will be of finite cardinality. I do not think this would affect the argument, as the satisfaction relation will still consist of a one-one mapping between the terms of T and M - if the world is of finite cardinality, then select an ideal theory satisfied by a finite model.

understand truth *simpliciter* in terms of TRUE(SAT) it follows that T is guaranteed to be true, contrary to what the metaphysical realist claimed.

Obviously the move to rejecting the metaphysical realist's original claim depends upon SAT being an intended interpretation - there is no guarantee, says the metaphysical realist, that SAT *is* the correspondence relation of reference and thus no guarantee that T is true *simpliciter*. Putnam's counter is that there is no way, other than by appeal to operational and theoretical constraints - which T is presumed to satisfy - to fix one interpretation as 'intended' over others. Quite simply, we cannot rule SAT out as unintended and thus T is true under an admissible model (i.e. is true *simpliciter*). Putnam's challenge: what, other than operational and theoretical constraints, fixes an interpretation as intended?

There are, it seems, two 'natural' answers: (i) surely operational constraints will 'rule out' deviant interpretations, contrary to what the argument asserts; and (ii) *we*, through our *intentions*, fix an interpretation as intended - e.g. we intend the word 'Paris' to refer to Paris and not to the number 2.[1] Let us look at Putnam's rejection of these 'natural' answers in order.

According to the first, only one interpretation will conform to the obvious truth of various empirical sentences; for example, only an interpretation which assigns 'cat' to cats and 'mat' to mats will deliver the truth of 'a cat is on a mat' (presumed to be observationally verified). Any interpretation which assigns falsity to 'a cat is on a mat' will be rejected as unintended. However, Putnam argues that it is possible for two (or

---

[1]As per the example in §1.1.

more) alternative interpretations to preserve the truth and falsity *of all the same sentences* even across all possible worlds.[1] Consider the following sentence:

2) A cat is on a mat.

Under the 'standard' interpretation, (2) is true in all possible worlds in which:

i) there is at least one cat
ii) there is at least one mat
iii) that cat is, was, or will be on that mat

and in which 'cat' refers to cats and 'mat' refers to mats. Sentence (2) can be reinterpreted such that *in the actual world* 'cat' refers to cherries and 'mat' refers to trees *without* affecting the truth-value of (2) in *any* possible world.[2] After the reinterpretation (2) will mean:

3) A cat* is on a mat*.

where 'cat*' and 'mat*' are defined by reference to three cases:

a) Some cat is on some mat, and some cherry is on some tree.
b) Some cat is on some mat, and no cherry is on any tree.
c) Neither case (a) nor case (b) holds.

Let us now introduce the following definitions:

Definition of 'Cat*':
x is a cat* if and only if case (a) holds and x is a cherry; or case (b) holds and x is a cat; or case (c) holds and x is a cherry.

Definition of 'Mat*':
x is a mat* if and only if case (a) holds and x is a tree; or case (b) holds and x is a mat; or case (c) holds and x is a quark.

Divide all possible worlds into those in which (2) is true and those in which (2) is false.

---

[1]The following is a paraphrase of his case in Putnam (1981b) Ch. 2.

[2]Note that 'is on' retains its standard interpretation.

Take all possible worlds in which (2) is true, and further divide them up into those in which some cherry is on some tree (i.e. those for which case (a) holds); and those in which no cherry is on any tree (i.e. those for which case (b) holds). In other words, we have divided up all possible worlds into three non-overlapping classes:

i) those in which (2) is true and some cherry is on some tree
ii) those in which (2) is true and no cherry is on any tree
iii) those in which (2) is false

Consider worlds of type (i). In those worlds, cats* are cherries and mats* are trees. As the second conjunct of case (a) holds in these worlds sentence (3) comes out true in these worlds. Next, consider worlds of type (ii). In these worlds cats* are cats and mats* are mats. Then, as the first conjunct of case (b) holds in these worlds, sentence (3) likewise comes out true in these worlds. Finally, consider worlds of type (iii). In these worlds, cats* are cherries and mats* are quarks. Now, as no cherry can be on a quark, sentence (3) comes out false in these worlds. Therefore, sentence (2) and sentence (3) have exactly the same truth-values in all the same possible worlds: in all those worlds in which (2) is true, so is (3), and in all those worlds in which (2) is false, so is (3).

Let us now reinterpret 'cat' by assigning it the intension[1] we just assigned to 'cat*' and simultaneously reinterpret 'mat' by assigning it the intension we just assigned to 'mat*'. Thus, in the actual world, 'a cat is on a mat' can be reinterpreted to mean 'a cherry is on a tree' without any change in truth-value in any possible world. Putnam

[1]An intension, in Putnam's sense, is a function which determines the extension of a term in any possible world.

then asks - how can we tell that 'cat' refers to cats and not to cherries and that 'mat' refers to mats and not to trees? Appealing to operational constraints - the truth of observation sentences - will not help. In the appendix to *Reason, Truth and History* he presents a formal proof which shows both that we can extend this same result to every sentence in a language and that there will be an infinite number of such reinterpretations compatible with the assignment of truth-values in any possible world. Putnam's draws the conclusion that:

> It follows that there are always infinitely many different interpretations of the predicates of a language which assign the 'correct' truth-values to the sentences in all possible worlds, *no matter how these 'correct' truth-values are singled out.*[1]

Putnam's counter to the second 'natural' response is that the notion of intention *presupposes* the notion of reference. He distinguishes between two sorts of mental states: *pure* and *impure*. A mental state is pure " if its presence or absence depends only on what goes on 'inside' the speaker."[2] On the other hand, a mental state is impure if its presence or absence also depends what goes on 'outside' the body or mind. According to his example, *being in pain* is a pure mental state but *knowing that snow is white* is an impure one - having the mental state of knowing that snow is white depends not just on something 'in' the mind but also on the fact that snow *is* white; "the world," as he says, "has to cooperate as well."[3] Now suppose that our intending to refer to, say, water by using the word 'water' were a pure mental state. We can imagine a possible

---

[1]Putnam (1981b) p. 35.

[2]Putnam (1981b) p. 42.

[3]Putnam (1981b) p. 42.

world identical to ours - call it 'Twin Earth' - save that the substance they call 'water' is composed of XYZ instead of H_2O. The inhabitants of Twin Earth may be in exactly the same pure mental state as we when they use the word 'water', but they do not refer to the same substance as we do when we use the word (assuming, of course, that we *do* refer to H_2O when we use the term). Thus, whichever pure mental state both we and the Twin Earthers are in is insufficient to fix the reference of 'water' to water (that is, to H_2O). Thus, intentions can only fix reference if we suppose them to be *impure* mental states: "Impure mental states of intending - e.g. intending that the term 'water' refer to actual water - *presuppose* the ability to refer to (real) water"[1] and thus our intentions can be of no help in fixing an interpretation as intended.

Let us concede that pure mental states are insufficient to fix an interpretation as intended - i.e. there must be some contribution the world makes in fixing the reference of our terms. On Putnam's account of the 'meaning' of natural kind terms, 'water' refers to whatever has the 'deep structure' of certain (defeasible) paradigmatic samples.[2] On his account, we stand (or stood) in a certain relation viz-a-viz paradigmatic samples of water (e.g. drank them, bathed in them, etc.) and baptized the substance of which those samples were composed 'water'. Following this, 'water' became a *rigid designator* - any other sample composed of the same 'deep structure' as the (defeasible) paradigms can be properly referred to by 'water'.[3] The important point is that 'water' refers to water,

---

[1]Putnam (1981b) p. 43.

[2]Putnam (1984b).

[3]Cf. Putnam (1973), (1975), (1981b), and (1981d).

on this account, in virtue of a relationship R holding between our (original) use of the term and certain paradigmatic samples of water (and it is because R takes, as relata, something 'outside of the mind and body' that our intention to use 'water' to refer to water is impure).

Thus, it at least seems promising to suppose that there is some relation R holding between us (or our use of words) and items in the world which contributes towards the fixing of an interpretation as intended. In other words, it seems not unreasonable to suppose that there is a constraint C (in addition to operational and theoretical ones), satisfied whenever we (or our use of words) stand in relation R to certain items in the world, which (at least partially[1]) fixes an interpretation as intended - the most obvious being a causal one as invoked in a causal theory of reference.[2] For example, as our use of 'Paris' stands in a causal relation (of baptism, perhaps) to Paris and not to the number

---

[1]When discussing constraint C in the following, I will assume that the other constraints (i.e. operational and theoretical) are met.

[2]See Putnam (1975) and Kripke (1972) for pioneering work in the causal account of reference. Brueckner (1984), Heller (1988), and Devitt (1991) all favour such a constraint. Merrill (1980) and Lewis (1984), however, look "not to the speech and thought of those who refer, and not to their causal connections to the world, but rather to the referents themselves." (Lewis p. 227). Their idea is that there are objective 'joints' in nature - Lewis' 'elite classes' and Merrill's 'intrinsic structuring of the world' - which themselves determine which referents are eligible and which are ineligible for an interpretation. As long as an interpretation assigns referents respecting such an 'intrinsic structuring', whether or not we can know that it does so, it is admissible. The claim that there are 'elite classes' (essentially, Putnam's self-identifying objects) is thought, by Lewis at any rate, to be captured by MR_1. Thus he argues that realism (i.e. MR_{3a}) is grounded in realism (i.e. MR_1): "If I am looking in the right place for a saving constraint, then realism needs realism. That is: the realism that recognizes a nontrivial enterprise of discovering the truth about the world needs the traditional realism that recognizes objective sameness and difference, joints in the world, discriminatory classifications not of our making." (Lewis (1984) p. 228).

2, interpretation I_2 (from §1.1) can be dismissed as unintended. In terms of the model-theoretic argument, there is no guarantee that SAT satisfies constraint C, thus no guarantee that it is intended, and thus no guarantee that TRUE(SAT) is truth under an admissible model.[1]

Putnam's response to this is that constraint C is 'just more theory'. The demand that an interpretation conform to constraint C is just the demand that:

4) *a* refers to x if and only if *a* bears relation R to x^2

Given that T is assumed to be complete, it will already contain (4) (assuming that it is a *genuine* constraint on the admissibility of an interpretation). (4), interpreted according to SAT, comes out true (that is, TRUE(SAT)) and thus there is no sense in which T fails to satisfy the constraint. SAT thus satisfies all operational, theoretical, and C-constraints and hence qualifies as an intended interpretation. T can thus be guaranteed to be true

---

[1]Although a causal constraint seems promising, nothing in the following will assume that the constraint *is* causal in nature. Interestingly, Putnam the Elder subscribes to the possibility of some adequate constraint C: "...it does seem likely that unintended interpretations could be ruled out by imposing suitable requirements of simplicity upon the compositional mappings we are willing to accept." Putnam (1960) p. 79.

[2]Sentence (4) is Field's (1972) proposed analysis of reference. He takes R to be a purely physical relation devoid of any semantic terms - probably a straightforward causal relation. For our purposes, it does not matter how we specify R, although Anderson (1993) criticizes the realist for failing to supply an acceptable account of it. However, he has the order of argumentative strategy confused - Putnam's argument is that *no* substitution of R will allow the metaphysical realist to avoid the model-theoretic results. Because Putnam's strategy employs this level of generality, it can be no complaint against the realist for employing the same level. Still, Anderson has rightly suggested that if the following defense succeeds, then, to complete the project, the realist would be well advised to strive towards finding an acceptable account.

simpliciter.[1]

Interestingly, if Putnam's 'just more theory' ploy succeeds, then not even a so-called *magical* theory of reference will avoid the problem.[2] A magical theory of reference is one which supposes there to be an *intrinsic*, non-conditional, relation between a word and its referent - for example that it is a 'surd metaphysical fact' that 'Paris' refers to Paris. Even granting such 'brute' facts, an ideal theory will contain sentences expressing those facts and still be guaranteed to be true under an intended interpretation SAT.

This, in a nutshell, is Putnam's model-theoretic argument against metaphysical realism. Let me summarize its results. First of all, if it succeeds then an epistemically ideal theory can be guaranteed to be true - i.e. it makes no sense to suppose that such a theory might, in reality, be false. Thus, the truth of a theory (or sentence) cannot coherently be radically divorced from any epistemological position - truth, in other words, cannot be 'radically non-epistemic', and MR_4 must be rejected. Secondly, there is (virtually) no limit on the number of satisfaction relations which can be defined

---

[1]Putnam (1976a), (1984a), (1987a), and (1992) offer a slightly different argument against constraint C than the one discussed here. There Putnam argues that causality is an interest-relative notion. Interest, he argues, is an essentially intentional notion and therefore presupposes reference. However, it seems to me that those arguments, at base, involve a form of the 'just more theory' ploy: it is because our *notion* (or theory) of causality is interest- or theory-relative that causality itself is, or that we must be able to *refer* to causality in order for causality to ground reference.

[2]Lewis (1984) wonders why Putnam presents the model-theoretic argument as being 'bad news' only for 'moderate, naturalistic realists' (p. 232-233). If the following argument is correct, Putnam need not have given his rhetorical arguments (1981b) against magical theories of reference.

between T and pieces of the world - given that they all satisfy all operational and theoretical (or even C-) constraints, they all count as 'intended' and hence are genuinely referential. In other words, there is no single correspondence relation of reference upon which truth *simpliciter* rests, and thus MR_{3a} must be rejected. Finally, given that there are many different 'intended' interpretations, there will be many different (as many as there are 'intended' interpretations) *complete* and *true* theories (descriptions) of the way the world is, and thus MR_3 must be rejected. If the argument goes through, metaphysical realism (at least as conceived by Putnam) seems in a bad way indeed.

2.1.2 Responses

Smart (1982) attempted to side-step the entire argument by directly challenging what he took to be its central conclusion - that there is no sense in which an ideal theory can be false. He asked us to consider two 'ideal' theories: T_1, which asserts that the physical universe consists of a four-dimensional time/space manifold in which all physical entities are contained; and T_2 which asserts that our familiar four-dimensional time/space manifold is but a cross-section of a larger five-dimensional one in which its cross-sections are strictly causally isolated from one another. Because the goings-on in the purported other cross-sections are strictly inaccessible to us, T_1 and T_2 must agree on all the observable facts - that is, they can be presumed to satisfy operational constraints equally. By appeal to the theoretical constraint of simplicity, however, we have reason to reject T_2. T_1 thus satisfies all operational and theoretical constraints and thus on Putnam's account can be guaranteed of being true. But for all that, the world

might not be simple and T_2 might *in fact* be true. T_2's (imagined) truth would entail T_1's falsity. Hence there is a clear sense in which even an epistemically ideal theory might be false.

Smith (1983) responded by arguing that if the other purported four-dimensional time/space manifolds are indeed strictly inaccessible to us, then there is no empirical content in the supposition that they exist - the elements of T_2 which go beyond T_1 in positing these 'other' manifolds are devoid of any real content. On the other hand, empirical content can be granted to them on the supposition that we are not *in principle* isolated from the other manifolds. But if we are not in principle isolated from them, then T_1 and T_2 cannot both satisfy all operational constraints.

Smart (in a letter to Smith quoted in Smith (1983)) argued that empirical content can be given to the disputed supposition by acknowledging that there is a *geometrical* relation between the various manifolds - i.e. content does not depend upon there being a *causal* relation. Smith argues that even if this is coherent, it is still not the case that T_1 is *false*; everything it says is still correct, although it might not be complete in the sense of containing all the truths there are.[1]

The debate continued. Melchert (1986) attempted to resurrect Smart's original intuitions by offering a new example - one from (personal) history as opposed to physics. Imagine that in the past you uttered a remark in innocence that was immediately taken as insulting. Upon reflection, you agree the best evidence (besides, that is, your own

---

[1]Smith fails to appreciate that if a theory is not complete it fails to be ideal in Putnam's sense.

'conviction' of innocence which others do not have access to) available to all (read: the best evidence possible) supports the 'insult-theory' against the 'innocence-theory'. The 'insult-theory', then can be taken as the best possible theory (read: ideal), and yet in a clear sense it might be false.

Smith (1986) quite rightly points out that all this may be true, but does nothing to disarm the model-theoretic argument. Melchert's 'best possible' theory is nothing like Putnam's 'ideal' theory. An ideal theory is, according to Smith, "an ideally well supported general theory of the world"; while Melchert offers us a fairly well supported particular theory of a human action. Melchert's 'best possible' theory, it seems to me, fails to approximate Putnam's ideal theory on another ground. An ideal theory must satisfy all operational constraints - i.e. it must conform to all the available evidence. Melchert's 'insult-theory' fails to take into account the first-person evidence of your conviction of innocence.

Davies (1987) thankfully put an end to this entire line of response. Smart's argument by counter-example (or any like it) is directed only at the *conclusion* of the model-theoretic argument, not against the argument itself. But, a counter-example can succeed only if there is some flaw in the argument itself. Thus, unless one can demonstrate the flaw, to assume that a purported counter-example is genuine begs the question. Furthermore, unless the question is begged, we cannot assume that there is a neutral interpretation from which we could judge that T_1 is false (or incomplete) and T_2 true. According to the model-theoretic argument, the truth (or falsity) of a theory can only be judged from *within* a particular interpretation. Thus, for an argument by

counter-example to succeed, it must be the case that there are constraints *other* than operational and theoretical ones which fix an interpretation as intended. Thus, an argument by counter-example can succeed only if there are *independent* grounds for supposing the model-theoretic argument fails. But if it can be shown that the model-theoretic argument fails, there is no longer any need to present the counter-example. Davies is quite correct. The only proper response to the argument is one which attacks it directly.

The model-theoretic argument, I contend, rests upon two crucial premises: that the 'just more theory' ploy succeeds, and that the notion of an ideal theory is unproblematic. Neither of these premises, I will argue, is immune from criticism.[1]

Recall that the 'just more theory' ploy depends upon converting some proposed constraint on admissible interpretation into linguistic form in such a way that a complete ideal theory would include it. In other words, it assumes an equivalence between: (a) some constraint C being satisfied by an interpretation I of an ideal theory T; and (b) T containing C-theory as a component. If C is a *genuine* constraint on interpretation, then T must contain C-theory (given the assumption that it is complete), and it is only if T contains C-theory, which also gets interpreted by SAT, that SAT can be presumed to have satisfied constraint C. However, as Lewis points out, there is a distinction between containing C-theory and satisfying C itself:

---

[1]The first set of arguments - against the 'just more theory' ploy - is taken from Gardiner (1994a).

> C is *not* to be imposed just by accepting C-theory. That is a misunderstanding of what C is. The constraint is *not* that an intended interpretation must somehow make our account of C come true. The constraint is that an intended interpretation must conform to C itself.[1]

Lewis agrees that, if constraint C is a genuine constraint on interpretation, T will include a sentence expressing it (i.e. will contain C-theory).[2] He denies, however, that an interpretation I which renders C-theory true-according-to-I will be guaranteed to be one which renders C-theory true-according-to-an-interpretation-conforming-to-C. Quite simply, we have no guarantee that SAT conforms to C *even though* C-theory is TRUE(SAT). Thus we have no guarantee that SAT is an intended interpretation and consequently no guarantee that T is true under an admissible model.

Putnam's reply is to ask what is required for an interpretation to conform to constraint C. All that can be demanded, he maintains, is that for each member of a list of the interpretation's mappings:

i) x_1 is assigned to a_1
ii) x_2 is assigned to a_2
⋮
n) x_n is assigned to a_n
⋮

be associated with a (true) member of the following set of sentences contained in the

---

[1]Lewis (1984) p. 225. Resnick (1987), Heller (1988), and Devitt (1991) follow Lewis on this point, though they extend the distinction in slightly different ways.

[2]It is not clear that he needs concede even this much. However, it makes no difference to the metaphysical realist whether it is conceded or not - the 'just more theory' ploy will fail on other grounds.

theory it interprets:[1]

i′) a_1 refers to x_1
ii′) a_2 refers to x_2
⋮
n′) a_n refers to x_n
⋮

Now, what is it for a member of (i′)-(n′)... to be true? Surely that it be true under an intended interpretation. An interpretation of, say (i′), will be intended (according to constraint C) just in case:

i″) 'a_1' refers to a_1
ii″) 'x_1' refers to x_1
iii″) 'refers' refers to reference[2]

(iii″) is the crucial clause. As long as it is contained in an interpretation, the rest of the interpretation will be acceptable. Lewis (et. al.) thus insist that only interpretations conforming to (or containing) (iii″) are admissible. Putnam's response is that unless Lewis is prepared to insist that it is a 'surd metaphysical fact' that 'refers' can only be correctly assigned relation R the same claim can be raised; namely that SAT can be guaranteed to assign relation R to 'refers' *as long as* 'relation R' is interpreted according to SAT. In other words, Putnam argues that only by accepting a magical theory of reference at *this* point; that 'refers', unlike any other term in the language, can be presumed to *intrinsically* refer to reference (that is, relation R); can the 'just more theory' ploy be halted. Once magical thinking is giving up, he says, there is no reason

---

[1]In the following, I assume that 'stands in relation R to' adequately analyzes the notion of reference. Thus, we can abbreviate '*a* stands in relation R to *x*' as '*a* refers to *x*'.

[2]That is, 'refers' stands in relation R to relation R.

to think that an interpretation which assigns 'refers' to reference (that is, relation R) is any more intended than one which assigns it to some other relation Q.

On an aside, Taylor (1991) attempts to circumvent Lewis' challenge and thus avoid the need for Putnam's response. He argues that once Putnam's argument is made clear (which neither Lewis nor Putnam do, he claims), it will be obvious that SAT conforms to constraint C. Let M be a model for an ideal theory T which is not assumed to satisfy C. M will, in fact, satisfy it just in case 'M satisfies C' is true; which is just to say, model-theoretically, 'M satisfies C' is true-in-some-intended-model. Taylor notes that the clause is a meta-linguistic statement concerning M. A meta-language for M can be constructed in terms of a meta-model M+ formed by adding constraint C as an axiom to M (which will guarantee, he claims, that M+ satisfies C), as well as adding stipulations "which form the recursive part of the theory of truth" (i.e. those stipulations which permit the derivation of statements of truth-conditions (the T sentences)). The presence of the truth-theory will be "enough to constrain [the model] to ensure that M+ now semantically explicates its embedded M".[1] This latter addition will, he claims, guarantee that M+ interprets its semantic vocabulary exactly as M does; in particular, that 'refers' as used in M+ will mean exactly what 'refers' as used in M means. So, because M+ is guaranteed (he claims) to satisfy constraint C, and M's semantic vocabulary does not differ in meaning from M+'s, M will also be guaranteed to satisfy constraint C. As such, it will qualify as intended.

However, Taylor relies upon exactly the same assumption challenged by Lewis:

---

[1]Taylor (1991) p. 160.

namely that constraint C is satisfied by a model as long as that model makes C-theory true. At best, Taylor's argument establishes the conditional claim: *if* M+ qualifies as intended, then so does M: but we have no warrant to detach the consequent. Of course, Taylor can offer a similar argument by invoking a meta-meta-model M++ which makes C-theory true and is guaranteed to agree on the interpretation of M+'s semantic vocabulary. But, Lewis' challenge is still not discharged - just because M++ makes C-theory true is no reason to suppose it satisfies constraint C. Taylor (and Putnam) will, however, argue that the realist can halt the regress only if there is "some safe conceptual haven" in which she can "formulate M+ and its Right Reference Restraint".[1] Such a 'safe conceptual haven' is to be found, presumably, only by invoking a 'magical' (hence undesirable) theory of reference. Thus, Taylor's careful reconstruction, while illuminating, does not obviate Lewis' challenge.

So, it seems that *either* the metaphysical realist must invoke a theory in which 'refers' has a special status in any particular interpretation, *or* no sense can be made of SAT failing to conform to constraint C. Clearly the first disjunct is undesirable - or is it? What is the special status which the metaphysical realist requires? According to Putnam it is that 'refers' intrinsically refer to reference (that is, to relation R) - i.e. that 'refers' admit of a magical account of reference. However, the only special status that the metaphysical realist requires is that 'refers' invariably be assigned the same relation independently of any interpretation in which it appears; i.e. that it be a precondition for the admissibility of an interpretation I that 'refers' not be interpretable in I but only

---

[1]Taylor (1991) p. 161.

)utside of it. Now, is this request unreasonable or 'magical'?

The model-theoretic argument can only be viewed as an argument against netaphysical realism if it is possible that an ideal theory admit of more than one nterpretation - if SAT is the only possible interpretation of T, then the issue of whether)r not it is intended would be irrelevant. Thus, the model-theoretic argument is only potentially) embarrassing to the metaphysical realist in so far as it suggests that there :an be (at least) two *different* interpretations. Let I_1 be one such interpretation whose nappings includes:

a_1) x is assigned to a
b_1) y is assigned to b
c_1) z is assigned to c

ınd let I_2 be another such interpretation whose mappings include:

a_2) α is assigned to a
b_2) β is assigned to b
c_2) γ is assigned to c

Notice that there is something invariant across both mappings - namely that objects stand n a particular relation - *being assigned to* - to terms. What we must assume merely in)rder for I_1 and I_2 *to be* alternative interpretations is that they each exploit the *same* elation. In other words, we must give the *same interpretation* to the relation - *is assigned to* - just in order to generate the model-theoretic argument. In still other words, is assigned to' must be assigned to the relation of assignment - and this must be done *extrinsic* to either I_1 or I_2 - otherwise Putnam's argument is a non-starter.

Why must the assignment of 'is assigned to' be carried on *outside* of either I_1 or $_2$? Suppose that it is done internally in each interpretation, then I_1 may include:

d_1) 'is assigned to' is assigned to relation P

'hile I_2 may include:

d_2) 'is assigned to' is assigned to relation Q

ut what can (d_1) and (d_2) mean? They can only mean:

d_1') 'is assigned to' stands in relation P to relation P

nd

d_2') 'is assigned to' stands in relation Q to relation Q

espectively. This, of course, forces a 're-interpretation' of, say (a_1) and (a_2), to:

a_1') x stands in relation P to a

nd:

a_2') α stands in relation Q to a

espectively. There is no longer any sense in which I_1 and I_2 are alternative *nterpretations.*

So, the very notion of an interpretation presupposes that, for any interpretation there must be some relation *not* interpretable in I - i.e. whose interpretation must be xed *externally* to I - but nonetheless which I must conform to in order *to be* an nterpretation. Once this is conceded, it is difficult to see what motivation there could e for denying that there may be a relation R, not interpretable in I, which I must onform to in order to be an *intended* interpretation. The special status which the netaphysical realist seeks for 'refers' is one which must already be granted to 'is assigned o'. Thus, nothing Putnam has said prevents the metaphysical realist from holding that, ust as the very notion of an interpretation presupposes that 'is assigned to' admit of a

ıivocal interpretation across all possible interpretations, so too does the very notion of ı *intended* interpretation presuppose that 'refers' admits of a univocal interpretation ross all intended interpretations.

Putnam may respond that it is incoherent to think that an interpretation I of a *mplete* theory T will not be complete - i.e. that there be some term of T not terpretable by I. The 'incoherence' is only apparent and can initially be softened by ıalogy. T is presumed to be a first-order system expressible in some formal language . Tarski argued that no consistent and complete formal language L_o can contain its vn truth-predicate - L_o's truth-predicate can only be formulated in its meta-language $_m$. Similarly, L_m cannot contain its own truth-predicate but must be formulated in *its* eta-language (the meta-meta-language for L_o). In the same vein I here suggest that no terpretation of a formal system can contain (i.e. interpret) its own reference-predicate. ow to say that a reference-predicate does not need to be interpreted is, I suppose, to voke a 'magical' theory of reference. The metaphysical realist, however, does not need) suppose that the reference-predicate is uninterpreted, only that *if* it is used in an terpretation I_o it cannot *also* be interpreted in I_o (but only in I_o's meta-interpretation $_1$).

It might seem that the analogy between Tarski's hierarchial truth theory and the roposed hierarchial referential theory is not complete. A language containing its own uth-predicate would generate the semantic paradoxes whereas it does not seem that milar referential paradoxes are generated if an interpretation interprets its own :ference-predicate. However, this is not so clear.

The counter-argument to the model-theoretic argument is that in an intended terpretation all assignments of objects to terms will be limited to pairs <o,t> in which stands in a genuine reference relation to t. That 'genuine reference relation' is to be ıderstood (following Field) in terms of some (perhaps physical or causal) relation R. other words, in an intended interpretation all assignments of objects to terms will be nited to pairs <o,t> in which o and t stand in relation R to each other. In still other ɔrds, an intended interpretation I is one whose assignments:

a) a is assigned x
b) b is assigned y
c) c is assigned z

n be unproblematically replaced by the list:

a′) a stands in relation R to x
b′) b stands in relation R to y
c′) c stands in relation R to z

Putnam's counter is that, for any interpretation SAT of an ideal theory T, (a)-(c) *n* be replaced by (a′)-(c′) simply because SAT itself interprets (a′)-(c′) in such a way at they come out true (that is, TRUE(SAT)). The realist rejoinder is that SAT, if it going to be an intended interpretation, must interpret (a′)-(c′) in the right way - in ırticular, it must contain the following:

d) 'stands in relation R to' is assigned relation R

ıich, in order to satisfy the realist demand, must be equivalent to:

d′) 'stands in relation R to' stands in relation R to relation R

ıtnam says that this is no problem, for SAT will make even (d′) come out true. But, ΑT is only problematic for the realist if there is some possibility that it not assign the

rrect' referents. In other words, there is only a problem if SAT assigns some relation ıer than R to 'stands in relation R to'; to be embarrassing to the realist, SAT must erpret (d′) in a 'deviant' way by containing something like:

e) 'stands in relation R to' is assigned relation Q

ere relation Q *differs from* relation R. Again, to satisfy the realist demand, (e) must equivalent to:

e′) 'stands in relation R to' stands in relation R to relation Q

On the surface it appears odd that SAT must contain both (d′) and (e′).[1] tnam will reply that the oddness is only apparent. But notice that he *cannot* resolve : oddness by resorting to a linguistic hierarchy; i.e. he cannot claim that (e′) offers *meta*-interpretation of the unquoted occurrence of 'stands in relation R to' as it ɔears in (d′). In offering such a defense, he would also then have to admit that (d′) ers a *meta*-interpretation of the unquoted occurrences of 'stands in relation R to' in :h of (a′)-(c′). (a′)-(c′) constitute an interpretation I_1 of which (d′) would ıstitute a meta-interpretation I_2 of I_1 itself. The whole thrust of the 'just more theory' ɔy is that an interpretation can interpret *its own* reference-predicate. So, Putnam must t bite the bullet and accept *both* (d′) and (e′) as parts of SAT.

The problem with this is that paradox can be generated. If SAT can offer any ignment it wants, then there is nothing to prevent it assigning 'stands in relation R to' relation Q *where* relation Q is to be analyzed as the complement of relation R - i.e.

---

[1]If it does not contain (d′) it does not meet the realist demand and is hence is intended. If it does not contain (e′) it is not potentially embarrassing to the realist.

identical to 'fails to stand in relation R to'. Understanding relation Q in this way, ') is equivalent to:

d″) 'stands in relation R to' fails to stand in relation R to relation R

ıilarly, all the other clauses would generate the same paradox; (a′) would be ıivalent to:

a″) a fails to stand in relation R to x

ł even (e′) would be equivalent to:

e″) 'stands in relation R to' fails to stand in relation R to relation Q[1]

e only way to halt such paradox would be to legislate that an interpretation is *not* ɔwed to assign the complement of relation R to 'stands in relation R to'. Such a islation, however, is tantamount to imposing an *external* constraint on admissible ɛrpretations - which is exactly what the realist wants and the 'just more theory' ploy ·ports to avoid.

Going back to the model-theoretic argument, there is no guarantee that SAT is *t* such a deviant interpretation, and thus no guarantee that it is intended. .UE(SAT), consequently, cannot be guaranteed of being truth under an admissible ·del, and the argument fails. Thus, the metaphysical realist can contend that it is a ıstraint on the admissibility of an interpretation I_o that its meta-interpretation I_m

---

[1]Nicholas Griffin points out that one of Russell's paradoxes can be generated if ations are allowed to apply to themselves (as per (d′)). Paraphrasing Russell, let T the relation which holds between a term S and a relation R whenever S does not nd in relation R to R. Then, whatever relation R might be, "S has the relation T to is equivalent to "S does not have the relation R to R". Hence, letting S name relation ınd letting R be relation T, "S has the relation T to T" is equivalent to "S does not ·e the relation T to T". (Russell (1910) Vol.1, Ch.2, §VIII, #3).

gns 'refers' to reference. I_m will correctly assign 'refers' to reference, on this account, in case *its* meta-interpretation (I_o's meta-meta-interpretation) correct assigns 'refers' eference, and so on.[1]

An independent argument can be run to the effect that there must be such levels nterpretations. Recall theory T from §1.1 which consists of the sentence: "The Eiffel ver is in Paris" and its two interpretations I_1 and I_2. It was a consequence that:

5) T is true-in-M_1
6) T is false-in-M_2

)ther words, both (5) and (6) are *true*. Heller (1988) assumes that this is enough to ıblish that there must be some notion of truth independent of truth relative to an rpretation: he says that sentences like (5) and (6) are *nonrelatively* true and hence :e is a theory-independent way that the world is; it is such, independently of any ɔry, as to make T true-in-M_1 and false-in-M_2.[2]

His argument is a bit quick. Putnam will reject it for leaving (5) and (6) true *'out* being interpreted.[3] But then Putnam must, it seems, supply us with an ıissible model under which they both come out true. Such a meta-model must ude a meta-interpretation - i.e. an interpretation I_m which interprets I_1 and I_2. I_m

---

[1]Heller (1988) and Devitt (1991) favour such a 'levels of interpretation' approach.

[2]Heller (1988) p. 116.

[3]Heller himself admits that "The question of which *un*interpreted theory is correct legitimate... [A]n uninterpreted theory has no truth value at all." (p. 116) I do not ı to suggest that Heller has flatly contradicted himself - he goes on to argue that what equired is that there be an interpretation under which (5) and (6) come out true but that we have to have any theory about such an interpretation.

t be richer than either I_1 or I_2.[1] Thus, it seems that model-theory itself (just as n-theory à la Tarski) must be committed to a hierarchy of interpretations.

Now, where does all this leave us? The model-theoretic argument depends upon ıg imperialistic - i.e. of SAT being an interpretation which renders all truths true. y under such an assumption can SAT be guaranteed to conform to any (genuine) straint on the admissibility of an interpretation. The 'levels of interpretation' ıment suggests that *no* interpretation can be like this - for any interpretation I there t be some truths (truths about, say, whether I satisfies some constraint) which I ıot make true. Lewis (et. al.) argued that even if SAT guarantees that C-theory is (that is, TRUE(SAT)) it cannot be guaranteed to conform to constraint C and thus ıot be guaranteed to be an intended interpretation. Putnam thinks that only by eal to a magical theory of reference can we make sense of this claim. If the above ıments are correct, then we can make sense of the claim by appeal to a non-magical archy of interpretations. Thus, there is strong reason for rejecting the 'just more ory' ploy; an interpretation may fail to satisfy a constraint C even though C-theory is according to it.

The other major premise of the model-theoretic argument was that the very notion ın epistemically ideal theory - a theory which satisfies all operational and theoretical straints - is unproblematic. The argument begins by *positing* such a theory and then

---

[1]It must, after all, contain the expressions 'true-in-M_1' and 'true-in-M_2' which cannot contained in either I_1 or I_2 on pain of generating the semantic paradoxes.

purports to derive various results embarrassing to the metaphysical realist. If the idea of such a theory can be discredited then Putnam's argument will be a non-starter.

Resnick (1987) maintains that "it is simply unclear what the theoretical constraints are and why there is no real question of Putnam's interpretation satisfying all of them".[1] We can view a theoretical constraint on an intended interpretation exactly as we viewed constraint C; i.e. in terms of there being a distinction between satisfying some theoretical constraint D and making D-theory (a component of T) true. On what basis can we assume that, for any given theory T, it satisfies all theoretical constraints - i.e. that T is an *ideal* theory in the required sense? We can only be assured that T is ideal, Resnick argues, if we can be assured of a correspondence between a sentence of T expressing satisfaction of some constraint D and some particular constraint on T's interpretations. Resnick's point is that there is simply no way that we can be assured of such a correspondence and consequently cannot be guaranteed that T is an ideal theory: "Putnam's argument depends upon the existence of a mapping between conditions on interpretations and those expressible with T such that one is satisfied if and only if its mate is. But there is no such mapping."[2]

The same argument can, I suppose, be run for operational constraints. There is, however, an additional problem with them. To see this, let us get clear on what Putnam

---

[1]Resnick (1987) p. 153.

[2]Resnick (1987) p. 154. We can, it seems, be assured that, for some theories, such a correspondence will not hold. As Resnick points out, according to the so-called Skolem paradox, a theory can assert that its domain is uncountable and yet have a countable domain.

means by an 'operational constraint'. In "Realism and Reason" he introduces the notion by way of example:

> [An ideal theory T] has the property of meeting all *operational* constraints. So, if 'there is a cow in front of me at such-and-such a time' belongs to [T], then 'there is a cow in front of me at such-and-such a time will certainly *seem* to be true - it will be 'exactly as if' there were a cow in front of me at that time.[1]

In "Models and Reality" he sharpens the notion:

> In my argument, I must be identifying what I call operational constraints, not with the totality of facts that could be registered by observations ... but with the totality of facts that will in actuality be registered or observed, whatever those be.[2]

And finally in *Reason, Truth and History* he supposes that, (probabilistically) associated with each (observation?) sentence S is an experiential condition E such that "an admissible interpretation is such that *most of the time* the sentence *S* is true when the experiential condition *E* is fulfilled".[3] His example is of S being "Electricity is flowing through this wire" and E being *my having the visual impression of seeing the voltameter needle being deflected.*[4]

Putnam's general idea of an operational constraint appears to contain two parts: (i) correlated with each (observation?) sentence S is a particular experiential condition E; and (ii) an interpretation I is intended just in case a sentence S is true under it only if its correlated experiential condition E obtains. So, if *my having the visual impression*

---

[1]Putnam (1976b) p. 126.

[2]Putnam (1980) p. 8.

[3]Putnam (1981b) p. 30.

[4]Putnam (1981b) p. 29.

of a cow in front of me at such-and-such a time is the 'correct' experiential condition associated with 'there is a cow in front of me at such-and-such a time', then if at that time I have a visual impression of a cow in front of me the sentence is operationally verified and, *ceteris paribus*, an interpretation of T under which that sentence would be true is intended. On the other hand, if at that time I have no visual impression of a cow in front of me, then the sentence is operationally falsified and any interpretation of T under which it came out true would be unintended.

I have no real complaint against the second component - it is (i) I find puzzling. It seems to me that if we accept the model-theoretic argument there is no longer any sense in which a purportedly ideal theory can satisfy (i). What I want to suggest is that the very notion of satisfying an operational constraint *presupposes* a determinate relation of reference.

Consider an ideal theory T which includes:

7) The Eiffel Tower is in Paris.

Because T is presumed to satisfy all operational constraints, there must be some experiential condition E such that if E obtains only those interpretations of T under which (7) comes out true will (potentially) qualify as intended. What is the experiential condition in question? The natural assumption would be that it is *having a visual impression of The Eiffel Tower being in Paris* (or something roughly similar). Is that natural assumption correct? Why would not *having a visual impression of The Calgary Tower being in Calgary* be correct? Consider the two interpretations I_1 and I_2 of T respectively:

a_1) The Eiffel Tower is assigned to 'The Eiffel Tower'
b_1) Paris is assigned to 'Paris'
c_1) the spatial relation *is in* is assigned to 'is in'

a_2) The Calgary Tower is assigned to 'The Eiffel Tower'
b_2) Calgary is assigned to 'Paris'
c_2) the spatial relation *is in* is assigned to 'is in'

Which experiential condition - *having a visual impression of The Eiffel Tower being in Paris* or *having a visual impression of The Calgary Tower being in Calgary* - must obtain in order for T to satisfy all operational constraints? Putnam will answer that it does not matter - that *relative* to I_1 the first must obtain in order for T to satisfy them and *relative* to I_2 the second must obtain. In other words, each interpretation gets to 'pick' its own experiential condition whose obtaining is sufficient for it to be intended. But then in what sense is an operational constraint a constraint on an interpretation's admissibility? A constraint for admissibility, it would seem, should be a constraint *imposed on* an interpretation, not a constraint *from within* an interpretation.

Putnam will answer that I am confused. Operational constraints, he will say, *are* imposed 'from without'. All that an operational constraint demands is that an interpretation I renders a sentence S true only if an appropriate experiential condition E obtains, and that is not a demand which comes from 'within' the interpretation at all. What *does* come from 'within' is a determination of *which* condition E must obtain in order for I to satisfy the constraint. Furthermore, an interpretation can 'pick' its own condition E and yet still fail to satisfy the constraint. Consider interpretation I_3:

a_3) The Eiffel Tower is assigned to 'The Eiffel Tower'
b_3) Calgary is assigned to 'Paris'
c_3) the spatial relation *is in* is assigned to 'is in'

I_3 selects *having the visual impression of The Eiffel Tower being in Calgary* as the experiential condition which must obtain in order for I_3 to be intended. That experiential condition fails to obtain and thus I_3 is to be rejected as unintended.

However, in order for this to work we need to presuppose that the terms describing the 'correct' experiential condition (determined by an interpretation itself) *refer* to the 'correct' objects. It might not be immediately apparent that there is a problem here. For example, in order for I_1 to qualify as intended it must be the case that:

8) I (or whoever) have a visual impression of The Eiffel Tower being in Paris.

is true, which in turn requires that 'Paris' (as used in (8)) refer to Paris - and 'Paris' *does* refer to Paris according to I_1. However, consider an interpretation I_4:

a_4) The Calgary Tower is assigned to 'The Eiffel Tower'
b_4) Calgary is assigned to 'Paris'
c_4) the spatial relation *is in* is assigned to 'is in'
d_4) Hamilton is assigned to 'Calgary'

In order for I_4 to qualify as an intended interpretation of T, it must be the case that:

9) I (or whoever) have a visual impression of The Calgary Tower being in Calgary.

is true, which in turn requires that 'Calgary' (as used in (9)) refer to Calgary; but it doesn't - according to I_4 'Calgary' refers to Hamilton.

There is a further problem. Consider an interpretation I_5:

a_5) The Eiffel Tower is assigned to 'The Eiffel Tower'
b_5) Paris is assigned to 'Paris'
c_5) the spatial relation *is in* is assigned to 'is in'
d_5) the relation *conceives of* is assigned to 'have a visual impression of'

I_5 picks *having a visual impression of The Eiffel Tower being in Paris* as the experiential

condition which must obtain in order for it to qualify as intended. However, according to I_5 itself, what is required for that condition to obtain is that I (or whoever) conceive of The Eiffel Tower being in Paris. It seems preposterous that my mere conceiving The Eiffel Tower to be in Paris is enough to operationally verify (7). Allowing an interpretation to select its own experiential condition makes satisfying operational constraints far too easy.

The moral to draw from this, I suggest, is that an interpretation cannot be allowed to interpret the description of the experiential conditions which must obtain in order for it to satisfy operational constraints. In other words, an interpretation cannot be allowed to 'pick' its own experiential condition. Which experiential condition whose obtaining is required in order for an interpretation to qualify as intended must be established (or described) *outside* of the interpretation.

Another way of saying this is that the descriptions of our phenomenal world must be interpreted invariably across any possible interpretation of an ideal theory. It is then those descriptions which an intended interpretation must be faithful to. Once this is admitted, I can no longer see any motivation in denying that *that* interpretation counts as the metaphysical realist's beloved unique relation of reference. On the other hand if we deny that the descriptions of our phenomenal world cannot be interpreted from within an interpretation, then there is no longer any reason to suppose that a given interpretation satisfies operational constraints - the very notion of an operational constraint would become incoherent. So, if the model-theoretic argument succeeds in the sense that there is no single privileged interpretation then there is no longer any

sense in which a theory can be *ideal* - i.e. one which satisfies all operational and theoretical constraints.

There is another doubt whether there could exist an ideal theory in Putnam's sense - or at least that we would ever recognize one if it came along. There are two ways we can think of an ideal theory: either as some actual, yet to be formulated, future theory (the theory we will formulate when we have done science long enough, perhaps) or as a heuristic device - an idealization based on the relationship between successor theories and their predecessors.

Take, for analogy, two types of mathematical functions. We can construct a mathematical function such that, when plotted on a graph, the values along the x-axis increase along the y-axis moving closer towards some fixed value further along the x-axis until they finally converge on that point. Alternatively, we can construct an asymptotic function where the values along the x-axis increase along the y-axis moving ever closer to some fixed further value on the x-axis without ever converging on it. In the first function, the fixed value along the x-axis is the convergence point of the increasing values along the x-axis; in the second function, the fixed value along the x-axis is the limit of the increasing values. In asymptotic functions, values never in 'reality' converge on their limits, though we sometimes find it convenient to treat them *as if* they did. If we forget that we are merely talking for 'convenience', and assume that the limit is actually the convergence point, we have made a mistake.

In terms of theory, it is commonplace to think that (past) theories are replaced

by (future) theories according to observational improvement. That is, for any two theories T_n and T_{n+1}, T_{n+1} is observationally better than T_n (it satisfies more observational constraints, as it were). Analogous to the case of the asymptotic and non-asymptotic functions, there are two distinct models by which we can view the relation between theories. On the one hand, we can view the (possible) history of (past, present, and future) theories as conforming to the model:

a) ..., T_{n-1}, T_n, T_{n+1}, ...

in which there is no observationally ideal theory (none which satisfies *all* operational constraints). On this view, the only sense of an ideal theory would seem to be analogous to that of a mathematical limit - i.e. an idealized theory to which the actual members of series of (a) approximate ever more closely. On the other hand, we can view the (possible) history of theories as conforming to the model:

b) ..., T_{n-1}, T_n

in which T_n is the last *possible* member. On this model, T_n would be observationally ideal - i.e. it would satisfy *all* operational constraints - and would suffice to play the role of Putnam's ideal theory. Which of these two models is necessitated by Putnam's argument?[1]

Suppose that model (a) is to be preferred. The ideal theory, then, is a convenient

---

[1]Putnam himself seems to lean towards the former: "'Epistemically ideal conditions', of course, are like 'frictionless planes': we cannot really attain epistemically ideal conditions, or even be absolutely certain that we have come sufficiently close to them. But frictionless planes cannot really be attained either, and yet talk of frictionless planes has a 'cash value' because we can approximate them to a very high degree of approximation." Putnam (1981b) p. 55. To be charitable, however, I will not assume that Putnam is committed to this view.

non-actualizable idealization. Even if it were the case that such a theory could, à la the model-theoretic argument, be guaranteed to be true, such a claim would have no effect on metaphysical realism. By holding that truth is radically non-epistemic, the most the metaphysical realist is committed to is that any theory which *we* possess, or *could* possess, might be false. This claim is left untouched by Putnam's model-theoretic argument *if* the ideal theory is interpreted as an idealization. The argument *starts* with the stipulation that T be an ideal theory. If there *is* no such T then the argument does not even start.

Koethe offers another argument against construing Putnam's ideal theory along (a).[1] Suppose we understand Putnam's purportedly ideal theory as one occupying some (relatively higher)[2] position in a series $\{..., T_{n-1}, T_n, T_{n+1}, ...\}$ - say T_m. Assuming that T_m is sufficient to represent Putnam's ideal theory, we can suppose that some 'intended' interpretation SAT renders T_m TRUE(SAT). The model-theoretic argument claims that because SAT is an 'intended' interpretation, TRUE(SAT) is truth under an intended interpretation, and hence qualifies for truth *simpliciter*. But, owing to the assumption that T_{m+1} is incompatible with T_m,[3] T_{m+1} must therefore be considered FALSE(SAT) and hence false *simpliciter*. It would be a consequence of the model-theoretic argument

---

[1]Koethe (1979). His argument tends to concentrate on the effects of model (a) for the model-theoretic argument whereas the one I will develop shortly tends to concentrate on the effects of model (b). Our conclusions are more or less the same, that "what the metaphysical realist ought to reply ... is that there simply is no such thing as a theory which is ideal in [Putnam's] sense." (Koethe (1979) p. 98).

[2]I.e. one that will be constructed in the (distant) future.

[3]Otherwise it would not be its successor theory.

(under assumption (a)) that a *more observationally ideal* theory be false.[1] The only way to avoid this difficulty (within assumption (a)), Koethe suggests, is to suppose that SAT fails to be an intended interpretation. and the model-theoretic argument is a non-starter. Therefore, for the model-theoretic argument to have any damaging effect on metaphysical realism, it must be understood as being committed to the (tenseless) existence of a last member of such a series - i.e. of a theory which is observationally ideal.

Thus, it would seem that the model-theoretic argument must presuppose model (b). Putnam may point out that the metaphysical realist is committed to such a view anyway. MR_3 commits the metaphysical realist to there being - an existential claim - one true and complete description/theory of the way things are (i.e. of the world). This 'one true theory' will just be an ideal theory in the sense Putnam needs to construct the model-theoretic argument.

Or will it? Taking the 'one true theory' as the ideal theory Putnam needs to construct the argument *will* yield the unintelligibility of the claim that even an epistemically ideal theory might be false, but only *trivially*:[2] Identifying the 'one true theory' with the ideal theory alluded to in the claim:

10) Even an epistemically ideal theory might be false.

would entail:

---

[1]In that event, falsity at least would be (radically) non-epistemic. Currie (1982) argues that, according to the model-theoretic argument, an ideal theory is guaranteed to be false just as much as it is guaranteed to be true.

[2]See Bailey (1983) for a similar claim.

11) Even the 'one *true* theory' could be *false*.

which certainly is unintelligible in the sense that it is self-contradictory. It would seem, then, that Putnam's ideal theory (the theory alluded to in (10)) *must* be other than the theory alluded to in MR_3 otherwise the metaphysical realist is committed to a gross inconsistency.

Putnam will of course respond that the metaphysical realist *is* committed to this gross inconsistency and that the model-theoretic argument graphically illustrates this. To be able to distinguish the theory alluded to in MR_3 from the ideal theory of the model-theoretic argument, it would have to be shown that it is logically possible for a theory which satisfies all operational and theoretical constraints to be false (as the metaphysical realist's 'one true theory' satisfies both operational and theoretical constraints *and* is true). The model-theoretic argument shows that it is *not* logically possible for a theory which satisfies all operational and theoretical constraints to be false. In other words, we can view the model-theoretic argument as pointing out a deep tension between MR_3 and the claim that even an ideal theory could be false. We can save the latter only by abandoning the former. Furthermore we cannot save the former by offering the argument that there is no ideal theory (i.e. that it is a mere idealization), for that denial would equally apply to the 'one true theory'. That is, MR_4 and (10) form an inconsistent pair.

However, *if* the theory alluded to in MR_3 is an ideal theory in the same sense as that alluded to in (10), then we certainly did not need the model-theoretic argument to point out that MR_3 and (10) are mutually inconsistent. If MR_4 does indeed entail (10),

then MR_4 is also inconsistent with MR_3. What all this shows, it seems to me, is that we need a more charitable reading of MR_4 - one that does not entail (10). To understand what the metaphysical realist means by saying that truth is radically non-epistemic, I suggest we return to the claim that truth is inherently a metaphysical notion.

By holding that truth is inherently a metaphysical notion, it does follow that for the metaphysical realist truth is inherently a non-epistemic notion. What the first claim means is that the final court of appeal, as it were, for the truth of sentences is the world itself and not any epistemic justification we have. What it does not mean, contrary to what Putnam suggests, is that there is *necessarily* a gap between the world and what we are warranted in asserting; i.e. that no theory, not even an ideal one, can be true. Typical sceptical arguments, such as those offered by Descartes, do not depend upon our *actually* being deceived, but only on the *possibility* of our being deceived. Or, in other words, they depend only upon our being unable to conclusively determine whether or not we are being deceived.

It is partly an acknowledgement of the force of these sorts of sceptical arguments which gives metaphysical realism its appeal.[1] We know that past theories have turned out to be false; we know this precisely because they failed to 'square' with the 'world'. Specifically, we deemed them to be false in virtue of their containing false observation sentences (i.e. entailing false predictions). Containing only true observation sentences,

---

[1]Of course, it is precisely the ability to undercut the skeptical arguments which give anti-realism its appeal. To those who think there is essentially something right about scepticism, this will not make anti-realism attractive. But even those who wish to resist scepticism may find the cost of anti-realism too high a price to pay.

then, is *the* generic operational constraint imposed on any adequate theory: a true theory is one which conforms to this constraint and hence entails no false predictions and a false theory is one which violates this constraint and hence entails at least one false prediction.[1] This explains why it is vital that Putnam's ideal theory satisfy all operational constraints.

Notice that Putnam's ideal theory has to embody two features: it has to contain all true observation sentences and it can contain no false observation sentence. It is this latter feature - the ideal theory cannot be falsified - which is interesting. If a theory were falsified (i.e. contained a false observation sentence), then that theory would, on that basis, fail to satisfy *all* operational constraints and hence would fail to be an ideal theory. That theory T satisfies all operational constraints; i.e. is free from falsification; is therefore essential to Putnam in constructing his model-theoretic argument - both in terms of using it to reject MR_{3a} *as well as* to reject claim (10).[2]

A theory is unfalsifiable (Koethe's unrevisable) just in case there is no possibility that falsifying evidence could come to light. For our purposes, a theory is falsifiable just in case it contains a false observation sentence and is falsified just in case it is shown (i.e. known) to contain a false observation sentence. Such knowledge requires us to be aware that some experiential condition E fails to obtain (or obtains) when some

---

[1]Quine (1990) is quite right to maintain that success of prediction, while not the aim of science, is nonetheless its test.

[2]On Koethe's account, Putnam's ideal theory must be *unrevisable* in the sense that it cannot have a successor theory improving upon its observational inadequacies simply because it is assumed to have none.

observation sentence contained in the theory asserts that it would (or would not). Conversely, a theory is unfalsifiable just in case it contains no false observation sentence. It can be known to be unfalsifiable, then, only under the assumption that we can be assured that no experiential condition E whose obtaining is asserted by some sentence of T will fail to obtain (or that some experiential condition E′ whose non-obtaining was asserted obtains). This in turn requires that we be capable of surveying all actual and possible experiential conditions *as well as* all (observation) sentences contained in T to assure ourselves that they 'match up' in the requisite way. In other words, for us to be confident that a theory could be ideal in Putnam's required sense, we would have to be possessed of capacities we do not have - quite simply we would not recognize an ideal theory even if it bit us on the nose. In still other words, the model-theoretic argument requires that there be at least one (radically) non-epistemic truth; namely that an ideal theory (tenselessly) exists.

The gap, then, between metaphysics and epistemology should not, as Putnam assumes, be thought of as between theories and the world - there may be *no* gap between an ideal theory and the world (which is precisely what MR_3 asserts) - rather it should be thought of as between the world and the state of our *knowledge*. *That* a theory T is epistemically ideal - *that* it satisfies all operational and theoretical constraints - is a metaphysical fact. *That* we can recognize T as epistemically ideal is an epistemological fact. These two facts do not necessarily converge, even *if* T *is* epistemically ideal. Seen in this light there is a perfectly good sense in which *any* theory which we might possess

might be false - namely that we cannot guarantee its truth.[1]

Let me summarize the main results. In order for MR_{3a} to be respectable, there must be some procedure which selects a single 'intended' interpretation from the (virtually) unlimited number of correspondence relations which can be defined between objects in the world and the terms of some theory T. It is not unreasonable to suppose that there be some constraint C in addition to operational and theoretical ones which fix such an interpretation as intended - causal ones seeming the most promising. An ideal theory, being complete, would have to include a sentence expressing that constraint C is a genuine constraint on admissible interpretations. However, merely guaranteeing that that sentence is true according to some interpretation is not sufficient to guarantee that that interpretation satisfies the constraint in question. What would guarantee that an interpretation satisfies the constraint is if it can be guaranteed to assign the central relation mentioned in that constraint to 'refers'. This 'guarantee' in turn requires that 'refers', as used in an interpretation, not be interpreted by that interpretation. Far from such a requirement invoking a magical theory of reference, it is a precondition of the very idea of an interpretation. In a similar vein, the very notion of an interpretation satisfying an operational constraint - a notion inherent in the very idea of an ideal theory

[1]Putnam may respond that nonetheless the model-theoretic argument succeeds against MR_{3a}. If we grant the (tenseless) existence of an ideal theory (which the metaphysical realist already concedes by granting the existence of the 'one true theory'), then, according to the model-theoretic argument, we can give it any number of distinct interpretations, all of which will come out true. This response ignores the other difficulties faced by the argument discussed earlier.

- presupposes a determinate reference (outside of the interpretation in question) of the descriptions of the experiential conditions making up the constraint. Merely in order to make sense of the claim that reference is indeterminate at one level (i.e. that at any given level there may be more than one 'intended' interpretation) we must presuppose that reference is determinate at higher levels. Putnam's model-theoretic argument, therefore, must presuppose what it aims at denying.

Secondly, for MR_4 to be respectable, there must be a clear sense in which even an epistemically ideal theory could be false. That sense is captured by admitting that for no theory could we guarantee that it is true in that we cannot guarantee that it is unfalsifiable. A theory *may be* unfalsifiable, but this is a fact which we could never be assured of. That a given theory is unfalsifiable would be a truth which transcends our epistemic capacities, and thus if there could exist an unfalsifiable theory, there would have to be at least one truth which is (radically) non-epistemic. Therefore, as the model-theoretic argument requires that an ideal theory be an actualizable possibility, it must presuppose that there is at least one non-epistemic truth. In other words, it must again presuppose what it aims at denying.

Finally, in order for MR_3 to be respectable, it must be the case that MR_{3a} is respectable. The model-theoretic argument fails to make MR_{3a} unrespectable. On a side note, in terms of MR_3 exclusively, the model-theoretic argument aims at showing that there can be a single theory which admits of distinct but equally intended *interpretations*. Even if the model-theoretic argument fails in that regard, MR_3 would still be in trouble if it could be shown that there are distinct but equally well-supported

theories. In terms of an argument against MR_3, it makes no difference whether there are distinct but equally well supported interpretations of a single theory or if there are distinct but equally well supported theories *simpliciter*. In other words, the internal workings of the model-theoretic argument against MR_3 directly (as opposed to those via MR_{3a}) are more or less indistinguishable from the internal workings of Putnam's (and Goodman's) argument from the existence of distinct but empirically equivalent theories. Those arguments will receive an extended discussion in §2.3.

2.2 Brains in Vats

Putnam's Brain-in-a-Vat argument is a curiosity. In the first place, it seems natural to read it as advancing an anti-skeptical position with respect to the external world, yet Putnam intends it to involve the metaphysical issues of realism and anti-realism. More specifically, Putnam sees it as somehow undermining metaphysical realism and somehow supporting his own preferred internal realism. In the second place, he uses it as an apparent springboard to his model-theoretic argument, which is more clearly aimed at metaphysical realism: "Why is it surprising that the Brain in a Vat hypothesis[1] turns out to be incoherent," he asks? "The reason is that we are inclined to think that *what goes on inside our heads* must determine what we mean and what our words refer to."[2] Once we give up the notion of a necessary or intrinsic connection between word and referent there is no bar, he claims, to the full-blown model-theoretic results. These two issues, the relationship between the vat argument and issues of realism and the relationship between the vat and model-theoretic arguments, are my chief concerns, but they require an extended examination of the vat argument itself.

2.2.1 The Argument

At its heart the argument involves the claim that as no one can *correctly* assert the

---

[1]I.e. the hypothesis that one is a disembodied brain suspended in a vat of nutrients whose "nerve endings have been connected to a super-scientific computer which causes the person whose brain it is to have the illusion that everything is perfectly normal." (Putnam (1981b) p. 6) Hereafter referred to as the BIV hypothesis.

[2]Putnam (1981b) p. 22.

sentence "I am a brain in a vat" it cannot possibly be true. As Putnam says, even though the supposition "violates no physical law, and is perfectly consistent with everything we have experienced,"[1] we cannot correctly *say* or *think* that we are brains in a vat, and thus (necessarily) are not. Why not?

To begin with, Putnam assumes an (almost) straightforward correspondence conception of truth. "I am a brain in a vat" is true just in case it corresponds, in the right sort of way, with an actual state of affairs *independently* of our knowledge of whether that state of affairs obtains. It is true only if it is the case the asserter has a noumenal brain (forever hidden by her phenomenal appearances caused by the inputs of the super-computer) which is spatially situated in a noumenal vat (also forever hidden to her). There is, however, an additional constraint on the truth of her assertion. It must be the case that, in *her* language, 'a brain' refers to her noumenal brain and 'a vat' refers to that noumenal vat.[2]

By hypothesis there need not be any phenomenal difference in the experiences of vaters and non-vaters, and hence there need not be any syntactic difference in their respective languages. For example, each may utter the same sounds 'There is a tree in front of me' with the intent to describe their common phenomenal experience. Nonetheless, we need to distinguish the language spoken by non-vaters from its syntactic counterpart spoken by vaters. Call the former 'English' and the latter 'vat-English'. In

---

[1]Putnam (1981b) p. 7.

[2]It is important to note that while these constraints may seem independent Putnam takes the failure of the latter as conclusive evidence for the failure of the former.

this way, we can view the problem of knowing whether the BIV hypothesis is true as equivalent to the problem of knowing whether we are speaking English or vat-English.

Putnam gives two distinct formulations of the vat argument, though both are intended to be 'self-refutation' arguments.[1] In the first version, as 'brain' and 'vat' are presumed to refer to noumenal brains and noumenal vats, the assertion "I am a brain in a vat" fails to express a truth. In the second version, as 'brain' and 'vat' are presumed to refer to phenomenal brains and phenomenal vats, the assertion "I am a brain in a vat" expresses a falsehood. The differences are subtle, but can be brought out in the following reconstructions:

Version 1

a_1) I am a brain in a vat if and only if "I am a brain in a vat" is true.

b_1) "I am a brain in a vat" is true if and only if (i) I have a noumenal brain which is spatially situated in some noumenal vat, and (ii) 'a brain' refers to that noumenal brain and 'a vat' refers to that noumenal vat according to the language in which the sentence is constructed.

c_1) 'Brain' refers to noumenal brains and 'vat' refers to noumenal vats in a language L if and only if there is a causal connection between (at least) some L-tokens of 'brain' and noumenal brains and some L-tokens of 'vat' and noumenal vats.

d_1) At least some English tokens of 'brain' and 'vat' stand in a causal connection to noumenal brains and noumenal vats.

e_1) No vat-English tokens of 'brain' and 'vat' stand in causal connections to noumenal brains and noumenal vats.

f_1) I speak either English or vat-English (but not both).

---

[1]A thesis is self-refuting if "it is *the supposition that the thesis is entertained or enunciated* that implies its falsity." (Putnam (1981) pp. 7-8).

g_1) Suppose I am speaking vat-English.

h_1) Then, by (e_1), when asserting "I am a brain in a vat", my tokens of 'brain' and 'vat' fail to be causally connected to noumenal brains and noumenal vats.

i_1) Therefore, by (c_1), my tokens of 'brain' and of 'vat' fail to refer to noumenal brains and noumenal vats.[1]

j_1) Therefore, by ($b_1$ii), my assertion of "I am a brain in a vat" is not true.

k_1) Therefore, by (a_1), I am not a brain in a vat.

l_1) Suppose I am speaking English.

m_1) Then, as by definition only non-vaters speak English, I am not a brain in a vat.

n_1) Therefore, I am not a brain in a vat.

Version 2

a_2) I am a brain in a vat if and only if "I am a brain in a vat" is true.

b_2) A word-token can refer only to those objects with which it stands in a particular causal connection.

c_2) I speak either English or vat-English (but not both).

d_2) Suppose I am speaking vat-English.

e_2) My tokens of 'brain' and 'vat' can only be causally connected to purely phenomenal brains-in-the-image and vats-in-the-image.[2]

---

[1]If they refer at all, they either refer to the causal sources of my phenomenal experiences (electrical impulses or program features, Putnam tells us) or to the phenomenal experiences themselves (brains-in-the-image and vats-in-the-image). (Putnam (1981b) pp. 14-15) Version 2 presumes that a vater's use of 'brain' and 'vat' are referential in this sense.

[2]Putnam uses the prefix '-in-the-image' to indicate the objects are purely phenomenal; e.g. a tree-in-the-image would refer to an aspect of an 'hallucination' (Putnam (1981d) p. 15). Alternatively, my tokens of 'brain' and 'vat' might refer to the causal sources of my phenomenal experiences (e.g. electrical impulses). In that case, my

f_2) Therefore, by (e_2) my tokens of 'brain' and 'vat' can only refer to brains-in-the-image and vats-in-the-image

g_2) Therefore, by (f_2), "I am a brain in a vat" is true if and only if I am a brain-in-the-image in a vat-in-the-image.

h_2) But, I am not a brain-in-the-image in a vat-in-the-image.

i_2) Therefore, "I am a brain is a vat" is not true.

j_2) Therefore, by (a_2), I am not a brain in a vat.

k_2) Suppose I am speaking English.

l_2) Then, as by definition only non-vaters speak English, I am not a brain in a vat.

m_2) Therefore, I am not a brain in a vat.

2.2.2 Responses

The cornerstone of either version is Putnam's insistence that there is a causal constraint on reference - his premises (c_1) and (b_2).[1] If that constraint is rejected, his conclusion will not follow. J. Harrison (1985), for example, argues that we can and in fact do learn the correct meanings of many words even though we fail to stand in the right sort of causal relation to their referents. In his example, most of us learn the meaning of the term 'duck-billed platypus' by being causally connected not to its actual referents but only to television images. However, Putnam could easily respond that

---

token of 'brain' refers to electrical impulses of type X while my token of 'vat' refers to electrical impulses of type Y. The same argument goes through on either interpretation - merely uniformly substitute 'electrical impulse of type X' for 'brain-in-the-image' and 'electrical impulse of type Y' for 'vat-in-the-image' throughout.

[1]See Putnam (1975), (1981b) pp. 1-21, and (1984b) for his arguments in support of the causal constraint.

while such a person is not *directly* causally connected to duck-billed platypus, the television image is, and thus such a person is nonetheless *indirectly* causally connected with its correct referent. Such indirect connections, it seems reasonable to suppose, suffice for success of reference. On the other hand, Lewis (1984) and Fales (1988) argue that a purely causal account cannot exhaust an adequate referential theory. They argue for a hybrid causal-descriptivist account, where a vater's use of 'brain' and 'vat' will succeed in their reference on descriptivist grounds. For my part, I am willing to grant Putnam his causal account. The problem with the argument lies not in its premises, but in its form.[1]

There is a *non sequitur* in version 1. One can accept that the truth of the *assertion* "I am a brain in a vat" requires a particular causal connection between the asserter's use of 'brain' and 'vat' and noumenal brains and vats without also accepting that the truth of the *proposition* requires such a connection. In other words, there is at least a *prima facie* distinction between (i) 'P' is true and (ii) 'P' is correctly assertible. Recognizing the distinction forces a reinterpretation of (b_1) as:

b_1') "I am a brain in a vat" is correctly assertible if and only if ...

That reinterpretation forces a similar reinterpretation of (j_1) as:

j_1') Therefore, by (b_1'ii), "I am a brain in a vat" is not correctly assertible.

which, in order to yield Putnam's desired (k_1) requires the following suppressed premise:

---

[1]As I will later argue, the causal constraint poses serious problems for the model-theoretic argument.

$*_1$) "I am a brain in a vat" is true only if "I am a brain in a vat" is correctly assertible.

However, to accept ($*_1$) is already to abandon a realist conception of truth. Thus, as the vat argument must presuppose a non-realist conception of truth, it can in no way be seen as an argument *against* realism.

Putnam's argument plays on the fact that there are no possible circumstances in which one can correctly say that one is a brain in a vat. It is in recognition of that fact that he wishes to conclude that no one can *be* a brain in a vat. But, there are no possible circumstances in which one can correctly say that they are not speaking, but from this it does not follow that everyone is constantly talking.

Putnam might counter that whereas I cannot correctly say "I am not speaking", nonetheless the state-of-affairs of my being silent can at least be connected with someone else's correct assertion of "Mark Gardiner is not speaking". It is, he might say, in virtue of the correct assertibility of "Mark Gardiner is not speaking" (by someone other than myself) that my being silent is a possible state-of-affairs. The BIV hypothesis, as he says, is one in which "all sentient beings are brains in a vat".[1] Thus, the state-of-affairs of my being a brain in a vat cannot be connected with the correct assertibility of "Mark Gardiner is a brain in a vat", for no actual utterer can correctly assert it (for if such an utterer *is* a brain in a vat, as per the supposition, their tokens of 'brain' and 'vat' fail to be causally connected with noumenal brains and vats). Thus, Putnam may conclude, the state-of-affairs of my (or anyone's) being a brain in a vat cannot be

---

[1]Putnam (1981b) p. 8.

connected with any correct assertion, and thus cannot be a genuinely possible state-of-affairs.

It is surely possible that there only exist a single language user (perhaps not at all times, but at some time - following a near total nuclear extinction, perhaps) and that that language user be sometimes silent. Suppose I am that sole survivor. In such a case "I am not speaking" would not be correctly assertible, nor could it be connected with a correct assertion of "Mark Gardiner is not speaking", and hence on Putnamian grounds my being a single surviving presently silent language user could not be a genuinely possible state-of-affairs. Either there must necessarily exist other language users (in which case solipsism has been refuted) or I am constantly speaking.

Putnam might avoid this undesirable consequence by grounding the state-of-affairs of my silent isolation in the truth of a counterfactual: *if* there were any other language users, they *could* correctly assert "Mark Gardiner is not speaking". This response is not open to Putnam on two grounds. In the first place, he wants to limit the extent of true sentences to those that are correctly assertible. It is a constraint on the assertibility of a sentence that there be an asserter. By hypothesis, there are no other asserters than myself, and my silence precludes me from asserting the counterfactual (and it would be incorrect if I *did* assert it). Thus, Putnam must either reject the truth of the counterfactual or else declare it to be a sentence whose truth transcends correct assertibility. If the former, then Putnam cannot ground the truth of the proposition "I am not speaking" in the correctness of the assertion "If there were any other language users, they could correctly assert 'Mark Gardiner is not speaking'". If the latter, then

truth is not co-extensive with correct assertibility, and there would be no special reason to think that the non-assertibility of "I am a brain in a vat" entails its falsity.

In the second place, if resort to such counterfactuals were warranted, it would also ground the possibility of my being a brain in a vat: *if* there were any language-using non-vaters, they *could* correctly assert "Mark Gardiner is a brain in a vat".[1] Thus, it seems that there is a strong counter-example to Putnam's move from the non-correct-assertibility of "I am a brain in a vat" to its falsity and thus to the necessity being a non-vater. Version 1 is simply invalid.

In regards to version 2, several commentators accuse Putnam of ambiguity - of vacillating between English and vat-English. The sentence:

12) I am a brain in a vat.

is true-in-vat-English just in case I am a brain-in-the-image in a vat-in-the-image and false-in-vat-English just in case I am not a brain-in-the-image in a vat-in-the-image. On the other hand, it is true-in-English just in case I am a noumenal brain in a noumenal vat and false-in-English just in case I am not a noumenal brain in a noumenal vat. Thus, (12)'s truth-in-vat-English-conditions differ significantly from its truth-in-English-conditions.

Putnam wishes to draw the conclusion:

13) I am not a brain in a vat.

but are we meant to interpret (13) according to English or vat-English? It seems obvious

[1]Stephens and Russow (1985) and Brueckner (1986) note that the vat argument fails as long as there is at least one language using non-vater.

that only the former will yield him the metaphysical results he is looking for, namely:

14) I am not a noumenal brain in a noumenal vat.[1]

According to Version 2, the truth of (14) follows from the truth of:

15) I am not a brain-in-the-image in a vat-in-the-image.

but there is no reason to suppose that (15) entails (14). It is true that (15) entails "I am not a brain in a vat", but only if that sentence is interpreted as being in vat-English (whose English meta-interpretation is simply (15)), but not the case that it entails the English meta-interpretation "I am not a noumenal brain in a noumenal vat".[2] To be consistent, we would need to replace (i_2) with:

i_2') Therefore, "I am not a brain in a vat" is not true-in-vat-English.

or:

i_2'') Therefore, "I am a brain-in-the-image in a vat-in-the-image" is not true.

which simply will not, in conjunction with (a_2), yield Putnam's desired (j_2). In other words, once the argument is disambiguated, it is seen to involve a *non sequitur.*

---

[1]Attempting to simultaneously discuss sentences of English and vat-English presents special difficulties. For the most part, we are forced to use English as meta-language for both. To alleviate some of the difficulty, I will use the words 'brain' and 'vat' when the interpretation is ambiguous, 'brain-in-the-image' and 'vat-in-the-image' when giving an English meta-interpretation of the vat-English usage, and 'noumenal brain' and 'noumenal vat' when giving an English meta-interpretation of the English usage.

[2]Tymoczko (1989) offers a 'reconstruction' of Putnam's argument which clearly commits the ambiguity, as argued in Gardiner (1994b). On the other hand, the vast bulk of the literature surrounding the vat argument centres on just this point (although in many different ways). J. Harrison (1985), Malachowski (1986), and Collier (1990) deny that the falsity of the assertion "I am a brain in a vat" entails that I am not a brain in a vat. Accusations of ambiguity between English and vat-English come from McIntyre (1984 - perhaps the clearest statement), Feldman (1984), Smith (1984), Stephens and Russow (1985), Brueckner (1986), and Iseminger (1988).

This response to Version 2 is quite similar to the response to Version 1. "I am a brain in a vat" cannot be used by a vater to assert that they are a brain in a vat - at best it can only be used to assert the false English meta-interpretation "I am a brain-in-the-image in a vat-in-the-image". Similarly, it cannot be used by a non-vater to make a correct assertion. So, there is no one for whom the sentence can be used to make a correct assertion, and it is for this reason that Putnam concludes that, necessarily, I am not a brain in a vat. Putnam then is relying on the principle:

> CAT[1]) If no one can correctly assert that they are a brain in a vat, then no one is a brain in a vat.

or

> CAT′) If "I am a brain in a vat" is not correctly assertible, then "I am a brain in a vat" is false.

If this assessment is correct, then the real problem with the vat argument is the assumption of the co-extensiveness of truth and correct assertibility. That assumption has been seriously challenged. All in all, Putnam has not made a sufficient case for the necessary falsity of the BIV hypothesis.

2.2.3 Interrelationships

A realist conception of truth, which Putnam and Dummett take to underlie metaphysical realism itself, allows for the possibility of a gap between the conditions under which a sentence is (ideally) justified and the conditions under which it is true.[2]

---

[1]I.e. the Co-extensiveness of correct Assertion and Truth.

[2]Recall that, for internal realism, truth just is idealized justification.

As Putnam likes to put it, metaphysical realism is committed to the view that even an epistemically ideal theory might, in reality, be false.[1] The potential gap between justification conditions (given empirically) and truth conditions (given ontologically) would be secured if we were guaranteed an access to the former but not to the latter. The BIV hypothesis, if it expresses a genuine possibility, would secure such a gap - in such a setting we would have access only to the phenomenal (empirical) world and not the noumenal (ontological) world.[2] As such, we would have no guarantee that the one corresponded to the other.

So, the BIV hypothesis can be understood as an example of how it could be possible that there be a gap between justification and truth conditions. Thus, if the vat argument is successful, then at least one of the supports for the gap collapses, and consequently one of the supports for metaphysical realism is lost. However, it is fairly clear that Putnam intends the vat argument to be more general in its scope. He intends it to illustrate how any description involving such a potential gap will (necessarily) fail.

Assume Putnam's causal constraint on reference - our linguistic tokens can only refer to those objects to which they stand in a particular causal relation. Suppose that there is a gap between our experience and the world - the one does not correspond to the other in the sense that the content of our experience is not caused by the nature of

---

[1]Putnam (1976b) p. 125 and (1980) p. 13.

[2]As Putnam says, metaphysical realism is characterized (in part) as the view which holds that "THE WORLD is supposed to be *independent* of any particular representation we have of it - indeed, it is held that we might be *unable* to represent THE WORLD correctly at all (e.g. we might all be 'brains in a vat', the metaphysical realist tells us)." (Putnam (1976b) p. 125).

the world (e.g. it is not noumenal vats which cause my experience of phenomenal vats, but something else). Thus, our linguistic tokens cannot stand in an appropriate causal connection with objects in the noumenal world, and thus cannot refer to them. As our tokens cannot refer to objects in the noumenal world, none of our sentences can express truths about that world. If none of our sentences can express truths about that world, then there are no truths about that world.[1] If there are no truths about that world, then that world does not exist. If that world does not exist, then there can be no gap between that (non-existent) world and the world of our experience. It would not matter, then, what description or explanation is offered for the possibility of such a gap - it is simply impossible.

However, as mentioned many times, the specific argument against the BIV hypothesis, and the above general argument, rests on the move from the absence of any correctly assertible sentence about such a world to the absence of truths about such a world. In other words, again the following principle is invoked:

CAT″) 'P' is true if and only if 'P' is correctly assertible.

In other words, it disallows, by *assumption*, the possibility of a true but non-assertible sentence. That assumption is tantamount to a rejection of a realist conception of truth. Thus, as an argument *against* metaphysical realism, it clearly begs the question. Secondly, if Putnam intends it to be an argument *in support* of his preferred internal

---

[1]This is, of course, Putnam's problematic premise.

realism, it similarly begs the question.[1] Principle (CAT″) then is needed both to ensure the validity of the original vat argument and to allow it to deliver an indictment against metaphysical realism. We have seen a strong counter-example to it from the sentence "I am not speaking", and thus have good reason to reject it.

However, it may be the case that Putnam intends the causal account of reference itself to support the internalist conception of truth. If it is the case that the extent of true sentences is limited to the referential constraints on the terms they contain, and if the referent of a term is limited to what it stands in particular causal relations to, and if the relata of those causal relations are limited by the objects of our phenomenal experience, it would seem to follow that the truth of sentences cannot transcend our phenomenal experience. In other words, it would seem to be the case that truth cannot outstrip provability, or correctness of assertion. If this is the case, however, the vat argument is impotent as either an argument against realism or as an argument in support of internalism, for it would rest upon independent considerations which would themselves present difficulties for metaphysical realism.

I was content above to accept the causal constraint on reference. If Putnam proposes that constraint to exhaust the truth about reference, and if that constraint entailed principle (CAT″), then I would be inclined to follow Lewis (1984) and Fales

---

[1]A number of commentators agree on this point. Stephens and Russow (1985) argue that the vat argument depends upon the assumption that reality is exhausted by what can correctly be asserted. That assumption, coupled with the claim that there is no privileged language from which to correctly assert "I am a brain in a vat" yields Putnam's desired result. That assumption, they note, presupposes a non-realist conception of truth. See also Smith (1984) and Dell'Utri (1990) (criticized by Casati and Dokic (1991)).

(1988) in opting for a hybrid causal-descriptivist account of reference. Such an option is not necessarily *ad hoc*. Consider the sentence:

16) I am not an undetectable gremlin.

By definition, an undetectable gremlin would be a creature beyond any possible experience we might have. In other words, whatever the causal source of the token 'undetectable gremlin' may be, it cannot be undetectable gremlins. Thus, by Putnam's causal constraint, 'undetectable gremlin' fails to refer to undetectable gremlins, and thus (16) cannot be correctly assertible. If, in addition, the causal constraint entails (CAT″), it would follow that (16) is not true - it would (necessarily) not be true that I am not an undetectable gremlin![1]

All things considered, it is my contention that the vat argument fails to eliminate the potential gap between truth and justification conditions. As such it fails both to be an argument against metaphysical realism and an argument in support of internal realism.

Putnam also intends the vat argument to be related to the model-theoretic argument. The BIV hypothesis suggests that it is possible for there to be two distinct, but syntactically indistinguishable, equally well supported theories. Let T_1 be an ideal theory in Putnam's sense constructed by a non-vater. Let T_2 be T_1's vat-counterpart. As

---

[1]Putnam might maintain that (16) lacks a truth-value. Besides involving the manifest (classical as well as intuitionistic) inconsistency of the existence of such sentences discussed in the Dummett section, similar reasoning would establish that "I am a brain in a vat" lacks a truth-value and would thus preclude Putnam from drawing his desired conclusion that one cannot, necessarily, be a brain in a vat.

T_1 is assumed to be complete, it will contain the sentence "I am not a brain in a vat". T_2 must then contain its counterpart. Forgetting the vat argument for a moment, it would seem to be the case that T_2's component sentence "I am not a brain in a vat" is false, thus invalidating T_2. But, T_1 and T_2 are equally well supported - i.e. if T_1 is an epistemically ideal theory, so is T_2. Thus, it seems there is a clear sense in which an ideal theory, in Putnam's sense, can be false contrary to what the model-theoretic argument is designed to show.

Lepore and Loewer (1988) suggest that this problem illustrates Putnam's intended relation between the vat argument and the model-theoretic one. If it is not possible to be a brain in a vat, then T_2 is not a possible theory: the 'falsity' of an impossible theory does not show that an ideal theory can be false. Thus, the vat argument would undercut potential responses to the model-theoretic argument.

The causal account of reference, however, suggests a more subtle relationship between the two. Suppose that the BIV hypothesis is not, as the vat argument aims to show, incoherent. In other words, suppose that T_2 is not an impossible theory, and that it *will* contain "I am not a brain in a vat" as long as T_1 does. "I am not a brain in a vat", as contained in T_2, must, however, be interpreted according to vat-English. Its English meta-interpretation, "I am not a brain-in-the-image in a vat-in-the-image", being true, does not invalidate T_2. In other words, as long as T_1 is ideally supported, so is T_2 *as long as* it is interpreted according to its own language. In still other words, as long as T_2 is allowed to interpret itself its truth is not in jeopardy. The model-theoretic argument is designed to show this: as long as interpretation is internal to a theory, there is no way

to reject an interpretation of an ideal theory as unintended and hence no way to reject such a theory as false. So, in a sense, we can take the vat argument as a graphic illustration of the model-theoretic argument. I say 'in a sense' because there is a significant difference between the two arguments.

The model-theoretic argument is designed to show that the notion of a privileged 'intended' interpretation is suspect - an ideal theory will admit of many different interpretations, all of which will count equally as 'intended'. In the vat argument, however, Putnam wishes to establish the conclusion "I am not a brain in a vat", but only as long as it is understood interpreted according to English (i.e. "I am not a noumenal brain in a noumenal vat"). The 'same' conclusion interpreted according to vat-English (i.e. "I am not a brain-in-the-image in a vat-in-the-image") does not deliver the metaphysical punch he is looking for. In other words, only English, not vat-English, supplies the 'right' or 'intended' interpretation for the conclusion. The vat argument, then, assumes what the model-theoretic argument rejects - the notion of a privileged 'intended' interpretation.

What happens to the vat argument if that notion is abandoned? Model-theoretically, there is a sense in which sentence (12) (the BIV hypothesis) can be *guaranteed* to come out true. Suppose that we are vaters constructing a theory T which includes (only) (12). Once we have given the syntax, we start to supply it an interpretation. Suppose we supply it interpretation I_1:

a_1) 6 is assigned to 'I'
b_1) mathematical equality ('=') is assigned to 'am'
c_1) 2 is assigned to 'a brain'
d_1) multiplication ('x') is assigned to 'in'

e_1) 3 is assigned to 'a vat'

Under I_1, (12) comes out true. So, at least relative to I_1, I am a brain in a vat, contrary to Putnam's conclusion. So the vat argument depends upon there being constraints on the admissibility of an interpretation - constraints which go beyond operational and theoretical ones (for I_1 satisfies *those*). I_1 just supplies the *wrong* interpretation of (12) to understand Putnam's argument.

What, then, is the *right* interpretation? It will *at least* have to involve something like the following:

a_2) the asserter is assigned to 'I'
b_2) identity is assigned to 'am'
c_2) a (noumenal) brain is assigned to 'a brain'
d_2) the spatial relation of containment is assigned to 'in'
e_2) a (noumenal) vat is assigned to 'a vat'

Putnam's opponent will grant I_2 (the interpretation containing a_2-e_2) as the intended interpretation for (12), and will hold that (12) is true just in case the asserter is in fact identical to some (noumenal) brain actually contained in some (noumenal) vat[1]; i.e. she will insist that (12) is true just in case it is true-under-I_2. But, according to Putnam's causal constraint on reference, just as truth-under-I_1 is insufficient to guarantee the truth of (12) in any interesting sense - i.e. in the sense which allows for the prospect of radical deception - so too is truth-under-I_2 insufficient to guarantee the truth of (12) in the required sense. The difference is that whereas I_1 is an inappropriate interpretation (we simply do not use 'a vat' to refer to the number 3), I_2 is not a possible one, at least not for a vater. (It *may* be possible for a non-vater, but then (12) would come out false-

[1]I.e. no mention is made of assertibility conditions.

under-I_2).

I_2 would not be a possible interpretation for a vater owing to proposed causal constraints. I_2 satisfies those constraints only if it conforms to the following:

a_2') the asserter's use of 'I' is causally related to herself
b_2') the asserter's use of 'am' is causally related to identity
c_2') the asserter's use of 'a brain' is causally related to a (noumenal) brain
d_2') the asserter's use of 'in' is causally related to the spatial relation of containment
e_2') the asserter's use of 'a vat' is causally related to a (noumenal) vat

Clearly, according to Putnam's argument, if I_2 is an interpretation proposed by a vater, then it would fail to conform to at least (e_2'), and thus truth-under-I_2 would not be sufficient to establish the truth of (12) in the required sense. Thus, Putnam, in giving the vat argument, is clearly committed to there being causal constraints on reference (that is, on admissible interpretations).

That being the case, it is difficult to reconcile the vat argument with the model-theoretic one. The latter argument depends upon it being the case that causal constraints are insufficient to fix an interpretation as intended - this is the 'just more theory' ploy.

To see the tension, suppose we are constructing an ideal theory T in Putnam's sense. Suppose further that we want it to include (12). Putnam will immediately say that T cannot include (12), for then it will fail to satisfy all operational constraints - it is simply not the case that it 'seems' that we are brains in a vat. Or is it? Relative to I_1, it certainly *does* 'seem' that we are brains in a vat (that is, it certainly does seem that 6=2x3). Thus, at least relative to an interpretation I_3 which includes I_1 as a proper subset, T satisfies operational constraints. It is further assumed to satisfy all theoretical

constraints. Thus, T is true-under-I_3, and thus true under an intended interpretation, and thus, contrary to Putnam's conclusion, a vater *can* correctly assert that they are a brain in a vat.

Suppose Putnam were to say that T itself, not its interpretations, has to satisfy operational constraints in order to count as ideal. I am not sure what this would mean - an uninterpreted theory is meaningless - it can neither satisfy nor fail to satisfy operational constraints. Putnam *cannot* say that a theory itself satisfies operational constraints just in case *all* of its interpretations do so. A complete theory would have to include at least some statements of identity, say b=b. If we interpret '=' as non-identity (which surely is a *possible* interpretation), then at least one interpretation of an ideal theory will fail to satisfy all operational constraints, and thus, under the proposal, no theory, however ideal, could satisfy them all.[1] To say that a theory satisfies an operational constraint is only to say that some interpretation of it does so. But then T interpreted according to I_3 is true under an intended interpretation, and thus (12) is true *simpliciter.*

But of course we want to say big deal - this does not show us that we are, in fact, brains in a vat. But this is only to say that *even if* I_3 satisfies all operational and theoretical constraints, it is not guaranteed to be intended. Quite right, it must also satisfy the proposed causal constraints listed as (a_2')-(e_2'). But according to the model-theoretic argument, I_3 itself interprets (a_2')-(e_2') in such a way that they come out true-

[1]Currie (1982) argues that by similar reasoning it can be shown that an epistemically ideal theory can be guaranteed to be false.

under-I_3. According to Putnam, this is all that we can ask of an interpretation in satisfying some constraint. Thus, it seems that if we accept the model-theoretic argument, we must deny Putnam's claim that we cannot correctly assert that we are brains in a vat.

But still we want to say big deal. All that the model-theoretic argument shows is that we can correctly assert that we are brains in a vat if what we mean is that 6=2x3; it does not show that we can correctly assert that we are brains in a vat if what we mean is that we are brains in a vat. In other words, we *intend* "I am a brain in a vat" to mean that I am a brain in a vat, and this according to the vat argument is what we cannot correctly assert. But, the model-theoretic argument is designed to show that there is not a single *intended* meaning of any sentence, including "I am a brain in a vat". Thus, if we accept the model-theoretic argument, then there can be an *intended* interpretation under which "I am a brain in a vat" comes out true, and thus there is nothing more we need to reject Putnam's claim we cannot "really, actually, possibly *be* brains in a vat".[1]

Is there any way out of this mess? The best way, it seems to me, is to take seriously the notion that causal constraints can serve both to select certain interpretations as admissible and reject others as inadmissible; but only if such constraints are *imposed* upon an interpretation and cannot themselves be interpreted by them. This is enough, it seems to me, to reject the model-theoretic argument. There is no guarantee that SAT conforms to such a constraint, and thus no guarantee that SAT is intended.

---

[1]Putnam (1981) p. 15.

Thus if we take the vat argument seriously, then there are causal constraints on reference. Those constraints themselves cannot be internal to but must be imposed upon acceptable interpretations. This allows us to avoid the results of the model-theoretic argument, but seems to leave us with the apparently undesirable results of the vat argument. However, as argued, there is serious doubt about the validity of the argument[1] and about its purported relationship to the issues of realism. Thus, in any event, metaphysical realism seems to emerge unscathed.

[1]Though not perhaps with the acceptability of the causal constraint on reference - at least in the sense of giving *some* of the truth about reference.

2.3 Arguments from Equivalence

2.3.1 Incompatible Empirical Equivalence

2.3.1.1 The Argument

Consider a possible world consisting of a line:

---

and two theories concerning that world: T_1, which asserts that the line can be divided up into line segments and infinitely small 'points', and T_2 which asserts that the line is composed only of line segments with extension. In other words, T_1 contains "There are points" while T_2 contains (or implies) "There are no points".[1] There is a definite sense, then, in which T_1 and T_2 are incompatible. According to MR_{1b}, the world sorts itself into ontological categories. Either it is such as to sort its objects into line segments *and* points or such as to sort its objects into line segments only; i.e. either it is such as to include points or such as to exclude points. If it includes points, then T_1 is true while T_2 is false. If it excludes points, then T_2 is true and T_1 is false.

But of course we do not *know* whether it includes or excludes points - all we know is that it contains the line. Our reasons for accepting T_1 are no better or no worse than our reasons for accepting T_2; there is no non-pragmatic reason for accepting one theory over the other. The conclusion one might draw from this is that there is no 'fact of the matter' whether the world contains points or not: T_1 and T_2 are equally good or equally true. But if T_1 and T_2 are equally true, then it cannot be the case that there is a unique true and complete theory of the way the line-world is. In other words, accepting that

---

[1]Putnam (1976b) pp. 130-135.

both T_1 and T_2 might be true is tantamount to rejecting MR_3 concerning the line-world.

There are initially three possible attitudes regarding the possibility of incompatible theories T_1 and T_2 having equal claim to truth: (i) *in fact*, either T_1 is true and T_2 is false or T_2 is true and T_1 is false, although perhaps we can never determine which; (ii) T_1 and T_2 each describe different worlds, W_1 and W_2, such that T_1 accurately describes W_1 and T_2 accurately describes W_2[1]; (iii) there is a single world, W, which can be accurately described in incompatible ways according to both T_1 and T_2. Both (i) and (ii) are consistent with MR_3, while (iii) is not.

Goodman argues that, as we "flinch at recognition of conflicting truths; for since all statements follow from a contradiction, acceptance of a statement and its negate erases the difference between truth and falsity,"[2] we will be loath to accept that such conflicting but empirically equivalent theories can be true of the same world. His proposal is to accept attitude (ii); given two empirically equivalent but incompatible theories, T_1 and T_2, we should 'postulate' two distinct worlds, W_1 and W_2, such that T_1 accurately describes W_1 and T_2 accurately describes W_2. However, according to other aspects of Goodman's irrealism, W_1 and W_2 are not independent of T_1 and T_2. Recall his idealism: W_1 and W_2 are 'created' by our creating T_1 and T_2. In other words, his irrealism allows one to retain MR_3 at the price of abandoning MR_{1a}. It is possible, I

---

[1]E.g. W_1 includes points and is accurately described by T_1 and W_2 excludes points and is accurately described by T_2.

[2]Goodman (1984) p. 30. Interestingly, Goodman assumes that worlds can only be described using non-paraconsistent logics; i.e. it is as if he allows consistency to impose an *external* constraint on the admissibility of a model.

suppose, for a metaphysical realist to accept this sort of ontological pluralism: there are distinct worlds, each populated with mind-independent entities, such that, for each world, there is exactly one true and complete description or theory of it. The problem with this is obvious - *we*, at least, seem to simultaneously inhabit both worlds; for it is *we* that are formulating theories about them. The thesis of ontological pluralism, at least within a metaphysical realist framework, is quite counter-intuitive.

Attitude (iii) retains ontological monism in the face of incompatible but equally true theories, but only at the price of abandoning MR_3. Let us take a closer look at this argument. It derives, I think, from three sources:

> a) There exist, or may exist, pairs of theories, T_1 and T_2, which are incompatible but empirically equivalent.
> b) The truth (or degree of truth) of a theory is determined by the evidence for the theory.
> c) There is exactly one true and complete description of the way the world is (i.e. MR_3).

The logical positivists, invoking the verification principle, argued that as meaning was exhausted by empirical consequences, two theories alike in empirical consequences were also alike in meaning; in other words, they held that any two theories which were empirically equivalent - i.e. had equivalent evidential bases and predictive powers - were by that fact theoretically or cognitively equivalent. Once the verification principle had been given up, however, most philosophers accepted that empirical equivalence was insufficient to establish cognitive equivalence.[1] This paves the way for claim (a), but is

---

[1]C.f. Quine (1975), Putnam (1983a) and Glymour (1970).

not sufficient to establish it. Take any two theories T_1 and T_2 which are empirically equivalent yet cognitively inequivalent. If T_1 and T_2 are mutually compatible, then they can be conjoined into a consistent super-theory T_3 (in other words, all compatible true theories can be conjoined into the unique true theory alluded to in (c), if such a theory is possible). Thus, empirically equivalent but cognitively inequivalent theories pose no special problems for realism as long as they are compatible. The anti-realist argument, therefore, depends upon the existence of not merely empirically equivalent while cognitively inequivalent theories, it requires that they be mutually incompatible. Thus, (a) asserts that there exist pairs of theories T_1 and T_2 such that (i) they are empirically equivalent and (ii) they are cognitively inequivalent in that they are mutually incompatible.[1]

There are, in addition, two distinct theses related to (a). The stronger thesis maintains that, for *any* theory T_1, there exists (or can be constructed) another theory T_2 which is empirically equivalent but incompatible with T_1. Quine, for example, argues

---

[1]One should not think, however, that compatibility of two theories together with their empirical equivalence will be sufficient for their cognitive equivalence. A good deal of the literature on this problem focuses on what constraint, in addition to empirical equivalence, is needed for cognitive equivalence; most of it agrees that some sort of mapping function from the sentences of each theory to the other is required - e.g. what Putnam (1983a) calls 'mutual relative interpretatibility', what Quine (1975) calls a 'reconstrual of the predicates that transforms the one theory into a logical equivalent of the other', or what Sklar (1982) calls 'appropriate structural mapping at the theoretical level'. For the purposes of this anti-realist argument, it is unimportant to precisely state what this additional constraint is - any two theories which fail to be compatible will fail to satisfy such a mapping. That is, it is the *incompatibility*, not the cognitive inequivalence *per se*, which will generate the anti-realist argument. It is possible, I suppose, for two theories to be compatible and empirically equivalent yet fail to satisfy such an additional constraint - but if they are compatible, they can be conjoined into a super-theory in the manner I suggested, thus yielding no special problems for the realist.

that as all (scientific) theories are underdetermined by data, all (scientific) theories must contain theoretical sentences which transcend their evidential sentences. Thus, for any theory T_1 which contains the set of evidential sentences Γ and some theoretical sentence P, another theory T_2 can be constructed which agrees with T_2 on Γ but disagrees with it on P. T_1 and T_2 will be empirically equivalent, as they agree on the same set of evidential statements, while being mutually incompatible. In other words, for Quine, the thesis of the underdetermination of theory by data is equivalent to (a)[1], and as the underdetermination thesis holds, he maintains, for all (scientific) theories, he asserts the stronger thesis regarding (a). The weaker thesis, on the other hand, asserts only that, for *some* theories, there exist (or can be constructed) incompatible yet empirically equivalent theories. As the weaker thesis is all that is required for the anti-realist argument it is all we need assume.

In terms of the second source, it should initially be recognized that evidence can be related to theory in at least two ways. First of all, evidence can take the form of describing phenomena that particular theories seek to explain or account for - *that* there are such phenomena is taken as evidence for the truth or acceptability of a theory. For example, the presence of petrified dinosaur bones is taken as evidence for the acceptability of paleontological theory. Such evidence forms what we can call the *evidential base* for a theory. Secondly, evidence can take the form of predicting future or hitherto unobserved phenomena. Success of prediction for a theory is taken as

---

[1]C.f. Quine (1975).

evidence *for* it, just as failure of prediction is taken as evidence *against* it.[1] Such evidence forms what we can call, borrowing Marxist terminology[2], the *predictive superstructure* of a theory. Generally, the evidential base of a theory contains what Quine calls *occasion sentences* while the predictive superstructure contains what he has called *observation categoricals*.[3] The line between these types of evidence is not hard and fast, nor need it be for our purposes. For all intents and purposes, they can be conflated. Let us therefore say that a theory entails or contains the observation sentences in both its evidential base and its predictive superstructure.

The relation between theory and evidence alluded to in (b) needed for the anti-realist argument is that the truth or falsity of a theory is a function solely of the truth or falsity of the observation sentences in both its evidential base and predictive superstructure.[4] Moreover, the notion of 'degree of truth' or 'approximation of truth' of a theory can also be derived from this basic one: the degree of the truth of a theory will be, on this account, a function of the ratio between the true observation sentences

---

[1]It is interesting that Quine now agrees that prediction does not serve as the aim of a scientific theory but rather serves as its testing procedure: "...prediction is not the main purpose of the science game. It is what decides the game, like runs and outs in baseball." (Quine (1990) p. 20). However, he cannot mean that success of prediction *alone* is what 'decides the game'. Surely failure to adequately account for observed phenomena is taken as evidence *against* a theory.

[2]Suggested by Nicholas Griffin.

[3]For a succinct description of these sentence types, see Quine (1990) § I.

[4]In essence, what (b) asserts is that the only constraints on the truth of theories are operational ones. Theoretical constraints ultimately serve only pragmatic purposes and cannot contribute to the determination of the truth or falsity of theories.

entailed by a theory and its false ones.[1] Two theories with exactly the same ratio of true empirical statements to false ones must be deemed to approximate truth to the same degree - we can have no non-pragmatic reason to accept one such theory over the other.

Take, then, two theories T_1 and T_2 which are empirically equivalent yet mutually incompatible. Suppose all of the sentences in T_1's evidential base and predictive superstructure are true. It follows, by (b), that T_1 itself is true. But, because T_1 and T_2 are empirically equivalent, T_1's evidential base and predictive superstructure is identical to T_2's. By (b), then, it follows that T_2 itself is true. But, T_1 and T_2 are incompatible - they cannot be conjoined into a consistent super-theory. Thus, there are two distinct true theories of the way the world is, contrary to what (c) asserts. (c), therefore, - that is, MR_3 - is unacceptable.

2.3.1.2 Responses

There is a tension between (b) and a metaphysical realist account of truth given in terms of MR_2 - a tension which is responsible, it seems to me, for an certain ambiguity in the term 'true' as applied to theories. According to MR_2, a theory is true just in case its terms refer and its sentences correctly describe 'reality'. On the other hand, (b) maintains that a theory is true just in case its observational evidential sentences are true. This claim entails that two theories alike in the truthfulness of their evidential sentences

[1]One could modify this basic account with the Popperian notion of verisimilitude, where the presence of even a single false observation sentence falsifies the theory (i.e. reduces its degree of verisimilitude to 0). The point is that, according to (b), the degree of a theory's confirmation is tantamount to its degree of truth.

are alike in their truthfulness as theories. A metaphysical realist, however, would insist that if two theories are incompatible, then even if they are empirically equivalent, *at most* one of them is true.[1] A metaphysical realist would simply reject (b) whenever it was in tension with MR_2. In other words, the argument draws upon two distinct senses of 'truth of theory'. On the one hand, (b) maintains that the truth of a theory is solely a function of the truth of its evidential sentences. On the other hand, the metaphysical realist maintains that the truth of a theory is solely a function of a correspondence relation between that theory and the way the world is. Thus, (b), as a premise in the anti-realist argument, begs the question.

The tension between these two conceptions of truth of theory tends to be ignored, it seems to me, because on the surface it appears quite harmless. For the metaphysical realist, truth has always been a predominately metaphysical notion; she wants to maintain a distinction between the *truth* of a theory and the *acceptability* of a theory. However, even the metaphysical realist will admit that the two are closely related, and for the most part it does no harm to ignore the theoretical distinction. The metaphysical realist will deem a theory to be true (and hence will deem its terms refer and that its sentences accurately describe reality) whenever it satisfies certain acceptability-conditions. Those acceptability-conditions are given by the truth of its observational evidential statements. However, for the metaphysical realist, satisfaction of those acceptability-conditions are only *guides* to truth - they do not *constitute* truth. That is, the metaphysical realist will always be open to the possibility that a theory satisfying

---

[1]I.e. a metaphysical realist would, it seems, opt for attitude (i) listed in §2.3.1.1.

acceptability-conditions will nonetheless violate truth-conditions, although we may never be in a position to know this.

However, where there is no serious question about our being led astray, we can talk about the truth of a theory when only acceptability-conditions have been satisfied. Where there is serious question of being so led astray, we must, as metaphysical realists, insist on the distinction. Serious question emerges precisely when we have two incompatible theories equally satisfying acceptability-conditions. If this is correct, then the proper realist response to the anti-realist argument is to reject (b). Moreover, this realist response is not merely an *ad hoc* one to avoid the problem - insisting on that distinction is part and parcel of realism. Unless, then, the anti-realist can give an independent argument for (b), the argument fails in rejecting MR_3.

Regarding (a), Quine remarks:

> Consider all the observation sentences of the language: all the occasion sentences that are suited for use in reporting observable events in the external world. Apply dates and positions to them in all combinations, without regard to whether observers were at the place and time. Some of these place-timed sentences will be true and the others false, by virtue simply of the observable though unobserved past and future events in the world. Now my point about physical theory is that physical theory is underdetermined even by all these truths. Theory can still vary though all possible observations be fixed. Physical theories can be at odds with each other and yet compatible with all possible data even in the broadest sense. In a word, they can be logically incompatible and empirically equivalent. This is a point on which I expect wide agreement...[1]

Quine sees the 'fact' of the underdetermination of theory by data as providing sufficient evidence for accepting (a). His expected 'wide agreement' did not, however, come about.

---

[1]Quine (1970) p. 179.

Tennant, for example, declares that there simply is no evidence that such pairs of theories exist.[1] Laudan and Leplin (1991) and Boyd (1973) argue (within a Quinean framework) that as the testing of theories does not solely involve a consideration of their empirical consequences, but presupposes a wide range of background conditions and auxiliary hypotheses, it is (almost) impossible to find theories which agree on both the empirical consequences and on the auxiliary hypotheses: theories which appear empirically equivalent assuming one set of auxiliary hypotheses will appear inequivalent assuming a different set.[2]

However, this argument is unconvincing. It may be the case that theories which appear on first consideration to be empirically equivalent can be made inequivalent by altering the background conditions, but there is no guarantee that, on second consideration, rival empirically equivalent theories will not be found. That is, while T_1 and T_2 may appear empirically equivalent at t_1, they can be made inequivalent by modifying the auxiliary hypotheses of, say, T_1.[3] Thus, at t_2, T_1 and T_2 are inequivalent. However, if Quine is right, then at t_2 we should expect to find a rival theory T_3 which is empirically equivalent to T_1. Quine's opponent[4] can respond by modifying the auxiliary

---

[1]Tennant (1987) pp. 29-30.

[2]See Kukla (1993) for criticism of Laudan and Leplin's argument, as well as Laudan and Leplin's (1993) reply.

[3]If the auxiliary hypotheses are considered as background information, and not part of the theory *per se*, modifying them will not produce a new theory.

[4]Quine's opponent is anyone opposed to establishing (a) on the basis of the underdetermination thesis. It is important to note that realists need not be opposed to Quine (for they may reject (b) rather than (a)), but rejecting (a) would be a way to halt the anti-realist argument. In this context, then, I refer to Quine's opponent as the

conditions at t_2 in order to make T_1 and T_3 inequivalent. However, again if Quine is right, we should expect to find a rival theory T_4 at t_3 which is empirically equivalent to T_1. As an argument against the possibility of incompatible empirically equivalent theories, the realist must be guaranteed to make the last move in this game, and there is no reason to suppose that the realist holds the trump card. The realist, by resorting to this argument, merely postpones the undesirable effects of rival theories. However, at least this line reveals the difficulty in showing the empirical equivalence of alternative theories - i.e. the anti-realist argument mistakenly assumes that empirical equivalence is easy to show. As such, doubt is cast on its initial premise.

Devitt (1991) argues that there is simply no good reason to suppose that there will be alternative empirically equivalent theories for any given (scientific) theory. He notes that Quine's use of 'possible evidence' must be construed in so liberal a way that there is no longer any reason to accept the thesis - we simply cannot rule out the possibility that, for any pair of seemingly empirically equivalent theories, novel future experiments or instruments will deliver different results for the two theories.[1]

Closely related to the problem of satisfactorily interpreting 'possible evidence' is the currently suspect distinction between observational and theoretical sentences. For the argument to succeed, two theories must agree on their observational sentences but disagree on their theoretical ones. If no such distinction can be adequately drawn, then

---

realist.

[1]Problems with interpreting 'possible evidence' is a standard complaint against the anti-realist argument. See also Lukes (1978).

it is at least questionable that any two theories satisfy being incompatible yet empirically equivalent. Quine's (1990) position that the distinction is not categorical but rather one of degree does not quite remove the problem. Laudan and Leplin (1991) accept Quine's account, but note that it entails that what counts as observational and what as theoretical will then be relative to particular points in our scientific history and our technological development. As the notion of 'evidence' is given in terms of 'evidential statements', all of which are held to be observational, what counts as evidence will then also be so relative. That is, there seems to be no adequate content to the notion of 'all possible evidence', nor the derived notion of 'empirical equivalence'.[1]

Alternatively, Clenndinen (1989) offers two arguments against (a). First of all, he distinguishes the evidential components of a theory from the predictive ones.[2] Empirical equivalence, he argues, involves only the latter. However, for any two theories, we can only determine whether or not their evidential components are equivalent, but, he argues, this is not sufficient to establish equivalence of the predictive components. Predictive power may transcend a theory's empirical adequacy. Secondly, he thinks that there may be theoretical constraints which can serve, in non-pragmatic ways, to decide between competing theories. He offers an argument to the effect that 'simplicity' may be a 'truth-tracking' criteria and not merely a pragmatic one. He argues

---

[1]Laudan and Leplin (1991) p. 454.

[2]In my reconstruction of the main anti-realist argument, I was also careful to distinguish them - in the literature the distinction is typically not made. Anti-realists focus almost exclusively on the truth or falsity of the predictive superstructure of theories.

that, as what we are warranted to accept is what we have empirical evidence for, if we have the same empirical evidence for T_1 and T_2, where T_2 is more complex than T_1, then the additional elements of T_2 are not supported by the evidence and as such we have no reason to accept it over T_1. However, Clendinnen's argument assumes, rather naively, that when we have two empirically equivalent theories, one will be a sub-theory of the other (i.e. they are identical up to a point, then one of them contains more elements). However, I think he has offered a telling argument for choosing between two empirically equivalent theories which are of this sort - we will come back to this when we look at pairs of theories of which one is a 'gratuitous extension' of the other.

Finally, consider what Newton-Smith (1978) calls the 'realist's dilemma': faced with two incompatible but empirically equivalent theories, one must either maintain that one of such a pair of theories is true and the other false, but that we cannot know which (which he calls the 'ignorance' response) or one must maintain that the world is simply indeterminate with respect to the rival theories - that if we cannot (in principle) know which way the world is, then there is nothing to know about it (which he calls the 'arrogance' response).[1] Newton-Smith opts for the 'arrogance' response, with its anti-realist consequences. Bergstrom (1984), however, argues that the 'ignorance' response is necessitated, paradoxically, by the anti-realist argument. Given that the 'ignorance' theory of theory choice is empirically equivalent to the 'arrogance' theory, then accepting

---

[1]The dilemma appears to be widely appreciated. For example, Glymour (1970) p. 285 asserts that "...the admission that there are empirically equivalent theories which are not synonymous seems to entail either that the true theory is sometimes unknowable or that, more simply, even all possible evidence can sometimes have more than one correct explanation."

the arrogance theory is tantamount to rejecting it. The 'arrogance' theory, "according to itself, is neither true nor false. So, if it is true, it is not true. Hence, it is not true."[1] In other words, Bergstrom argues that we can accept all of the ingredients leading up to the supposed anti-realist conclusion without accepting the conclusion.[2]

So, given all this, where does that leave (a)? I am inclined to think that reasons to accept it are far weaker than the anti-realist supposes.[3] It is hard to see why one should accept that for any theory (at least any empirical theory) T_1, there exists (or can be constructed) a rival incompatible theory T_2 which agrees on all *possible* evidence. Proponents of (a) typically either rely on an expected 'wide agreement' (which, as we have seen, has not come about) or else offer examples of such theories; but, there is no universal agreement on the success of the various examples.[4] I do not wish to enter into the debate of whether the purported examples are genuine or not - for my purposes it is enough to note that (a) is far from established. Moreover, there is reason to doubt

---

[1]Bergstrom (1984) p. 357.

[2]Newton-Smith gives a strange response to this. To accept the 'ignorance' response is, he says, to accept that "there are *facts* concerning which we can have no evidence", and that this is too high a price to pay. ((1978) p. 88). But, the view that truth may transcend our ability to determine it is part and parcel of metaphysical realism. The price to pay for realism, Newton-Smith seems to be asserting, is realism!

[3]Notice, however, that the counterarguments are not aimed at the underdetermination thesis *per se*, but only that the underdetermination thesis *entails* the existence of incompatible yet empirically equivalent theories.

[4]For example, Newton-Smith (1978) offers two examples which are dismissed by Bergstrom (1984) as violating conditions for either empirical equivalence or incompatibility. Goodmanesque (1978) and Putnamian (1976c) type examples such as point vs. line-segment universes and field vs. particle universes are criticized by J. Wright (1989) for violating Tarski's Convention (T).

that any such example *could* succeed - at least not without controversy. No matter how well entrenched or intuitive two theories are, *if* they are incompatible, then, as I suggested above, the proper realist attitude would be that at least one of them is false. That is, it seems to me that our conviction that at least one of a pair of incompatible theories is false is more reasonable than is our conviction that any given theory is true.

However, a general argument in support of (a) seems initially promising. Let Γ and Δ be the set of evidential sentences and the set of theoretical sentence of T_1 respectively. Add to T_1 any non-observation sentence P which is not derivable from either Γ or Δ, thus forming the theory T_2. T_1 and T_2 are empirically equivalent in virtue of having the same set of evidential statements, but are cognitively inequivalent in that they are not, say, relatively mutually interpretable; i.e. they agree on all the evidential statements but disagree on the theoretical statements. I can see no reason why this cannot be done for any given theory.[1]

---

[1]Kukla (1993) proposes that introduction of grue-like predicates will also supply an 'algorithm' for generating empirically equivalent alternatives: "Take any theory T with observable consequences O, and construct from it the theory T′ which says that T is true of the universe under the initial condition that the universe is being observed; but when nobody's looking, the universe follows the laws of T*, where T* is any theory which in incompatible with T. Clearly, one can find such a T′ for any T, and just as clearly, T′ is empirically equivalent to T." (pp. 4-5). Besides falling into the 'gratuitous extension' problem, Laudan and Leplin (1993) rightly question whether T′ can be a genuine *theory*, let alone an empirically equivalent one to T. For example, suppose T says that Mars will (be observed to) be in some location at some time, based on where it is (observed to be) at some previous time. T also (implicitly) projects where it will be (unobserved, say, but not unobservably) in the intervening times. T′ would locate it at different places at those times (otherwise it would not be incompatible with T), and hence when it *is* observed to be where it is supposed to be (according to both T and T′), it would have had to instantaneously cross space (presumably violating T's laws governing movement). In other words, T′ would have to posit "utterly mysterious physical events - instantaneous shifts - devoid of any theoretical mechanism, the additional provision of

However, this general method is not sufficient to establish (a). As I argued previously, cognitive inequivalence *per se* is not sufficient - incompatibility is needed. Unless T_1 denies P - which it cannot, given the constraints - T_1 and T_2 are not incompatible. They can be conjoined into the consistent super-theory $T_1 \cup T_2$ (which, eliminating redundancies, just is T_2). What is needed is a method for generating incompatible theories.

Again add to T_1 any non-observation sentence P committed to the existence of some x which is derivable from neither Γ nor Δ to form T_2. Add to T_1 the negation of P to form T_3. Thus, T_2 and T_3 will be empirically equivalent, cognitively inequivalent, *and* mutually incompatible. However, this method does not quite do the trick either - it only gives us a method for constructing *pairs* of incompatible empirically equivalent theories based on any given theory - it *does not* yield a theory which is incompatible but empirically equivalent to the theory we started with, namely T_1 (i.e. T_1 is not incompatible with either of T_2 or T_3).

Moreover, a general argument can perhaps be given *against* constructing the requisite rival theory by any such method of adding sentences to some base theory. In any such pair of theories, one will be a sub-theory of the other. As the larger theory will agree with the smaller theory in all of its evidential statements, any additional statement not contained in the smaller theory will be what Bergstrom (1984) calls a 'gratuitous

---

which would undoubtedly create further observationally attestable divergences between T and T′... Kukla's analysis ignores the fact that physical events characteristically initiate causal chains that initiated them... What Kukla portrays as an algorithm for producing equivalent scientific theories looks increasingly like the Evil Genius of the sceptics, requiring total suspension of science." (p. 11).

extension', and as such Clendinnen's argument from simplicity mentioned above may very well apply.

All in all, I cannot see any general method for ensuring that there will be the required incompatible empirically equivalent theories needed to generate the anti-realist conclusion. Therefore, as the reasons for accepting (a) seem weak, that (b) is question-begging and, what amounts to the same thing, that there seems to be good reason not to accept the conclusion of the anti-realist argument *even if* one accepts the premises, the metaphysical realist need not be overly worried by this argument.

2.3.1.3 Interrelationships

In the context of Putnam's model-theoretic argument, commitment to the existence of incompatible empirically equivalent theories may provide an argument *for* metaphysical realism.

The model-theoretic argument aims at showing that MR_3 is unintelligible. Thus, it would seem that Putnam, in denying that claim, is committed to either asserting that there is (or can be) two or more true and complete theories of the world or else that there is (or can be) no true and complete theory of the world. Given his internal realism with its attendant epistemic notion of truth as well as the model-theoretic argument he offers in its support, it would seem that Putnam must opt for the former. That is, any theory which was ideal from an epistemic point of view must be considered a true

theory.[1] Let T be a theory which is epistemically ideal. Map the individuals of M one-to-one with objects in the world. This mapping will define a satisfaction relation SAT between the terms of T and the world, and will similarly define a truth-predicate TRUE(SAT). Call the theory interpreted according to SAT T_1. Remap the individuals of M onto the world so as to define a new satisfaction relation TAT. This mapping will similarly define a truth-predicate TRUE(TAT). Call the theory interpreted according to TAT T_2. Now, all of the sentences of T_1, given that it is epistemically ideal, will come out true (that is, TRUE(SAT)) and all of the sentences of T_2, given that it is epistemically ideal, will also come out true (that is, TRUE(TAT)). Putnam's argument is that there is no way to decide which of SAT or TAT is the unique and intended relation of reference - each has equal claim and as such there is no way of deciding which of TRUE(SAT) or TRUE(TAT) defines the unique and intended truth-predicate. Hence, what we have are two theories which are mutually incompatible yet each has equal claim to truth. We simply cannot say, claims Putnam, that there is exactly one true and complete theory or description of the world - for each such theory there will be (at least) another which can also justifiably claim to be true and complete. Thus, it would seem that Putnam is committed to asserting that there are (or can be) two or more true and complete theories of the world.

What I want to argue is that there is a strange tension between the argument from incompatible empirically equivalent theories and the model-theoretic argument, although

---

[1]If there is (or can be) no such theory then, as argued in §2.1.2, the model-theoretic argument fails.

on the surface it would seem that they function in tandem.

To begin with, it is correct, I suppose, to claim that T_1 and T_2 are cognitively inequivalent. However, as I argued above, cognitive inequivalency is not sufficient for the anti-realist argument - the rival theories must also be *mutually incompatible*. It does not seem to me that a case has been made for T_1 and T_2 being incompatible (they *may* be, but I cannot see that they are *guaranteed* to be). Let me engage in some speculation. Suppose that T_1 and T_2 are not incompatible - that is, there is no sentence in T_1 which is FALSE(TAT) nor any sentence in T_2 which is FALSE(SAT). They can then be conjoined to form the super-theory T_3 which will be interpreted according to the reference relation RAT which consists of the disjunction of SAT and TAT. Similarly, its truth-predicate TRUE(RAT) will be defined as the disjunction of TRUE(SAT) and TRUE(TAT).[1] If we then take TRUE(RAT) as *the* truth-predicate, then there is a theory, namely T_3, which can justifiably claim to be the heralded metaphysical realist's 'one true theory', contrary to what the anti-realist argument intends to show. Thus, if

---

[1]Let the domain be {a,b,c}. Let SAT and TAT respectively be the interpretative functions for T_1 and T_2. Let SAT(α)=a, SAT(β)=b, SAT(γ)=c, TAT(α)=b, TAT(β)=c, and TAT(γ)=a. If T_1 and T_2 are consistent (as per the hypothesis), then it will never be the case that, say, SAT(ϕ)=TRUE(SAT) and TAT(ϕ)=FALSE(TAT). Thus, we can conjoin T_1 and T_2 and construct an interpretation RAT where RAT(α)=a $\vee$b, RAT(β)=b $\vee$c, and RAT(γ)=c $\vee$a. Similarly, the truth-predicate TRUE(RAT) can be viewed as TRUE(SAT) $\vee$TRUE(TAT). I admit that it is unclear whether we can guarantee, for any pair of rival theories, that if SAT(ϕ)=TRUE(SAT) then TAT(ϕ)=TRUE(TAT), but failure to do so would only be tantamount to showing that we cannot guarantee that the two theories are mutually consistent. Putnam needs the stronger claim that the two theories *cannot* be consistently conjoined, and for that he needs the strong claim that the two theories can be *guaranteed* to be mutually incompatible. What I intend to show is that Putnam *cannot* (on his own account) guarantee, for any pair of rival theories, that they are mutually incompatible.

this is correct, then the argument's conclusion depends upon there being no consistent super-theory T_3, which in turn demands that T_1 and T_2 be mutually incompatible. Is the anti-realist able to guarantee that they are? If not, then the model-theoretic argument is disarmed.

There may be reason to suppose that he cannot show them to be so incompatible. Let P_1 be any sentence of T_1 interpreted according to SAT and let P_2 be any sentence of T_2 interpreted according to TAT. What is required to show that T_1 and T_2 are incompatible is that, for some pair of sentences $<P_1, P_2>$, P_1 is the negation of P_2 (and, of course, vice versa). The problem is this - *from what interpretation could* P_1 *be judged the negation of* P_2? They cannot be so judged from within SAT, for P_2 interpreted according to SAT is not a sentence of T_2; nor can they be so judged from within TAT, for P_1 interpreted according to TAT is not a sentence of T_1. Nor is there any interpretation external to either T_1 or T_2 from which to make such judgements - this is precisely what Putnam's internal realism amounts to.[1] In other words, it seems that Putnam's own internal realism precludes him from being able to say that T_1 and T_2 are incompatible, and hence from drawing the anti-realist conclusion of the argument from incompatible empirical equivalent theories.

Thus, if Putnam cannot make a case for T_1 and T_2 being incompatible theories,

---

[1]The argument can be expressed as an indirect proof. Suppose there is an interpretation I from which T_1 and T_2 can be judged mutually incompatible (i.e. I interprets $T_1 \cup T_2$). It follows that I is inconsistent, and thus cannot serve as an interpretation for either T_1 or T_2 (for, being ideal theories, their intended interpretation are assumed to satisfy all theoretical constraints). Thus, there can be no intended interpretation which will allow one to say that T_1 and T_2 are mutually incompatible.

then the model-theoretic argument loses its teeth. However, one must be careful with how they treat this result. If T_1 and T_2 are mutually compatible, then they can be conjoined into the consistent super-theory T_3, thus vindicating MR_3. However, if they can be shown to be mutually incompatible, then MR_3 is in jeopardy. The foregoing argument was to the effect that the internal realist cannot show that T_1 and T_2 are mutually inconsistent, and thus cannot draw the conclusion rejecting MR_3. However, being unable to assert that T_1 and T_2 are mutually incompatible does not allow one to assert that they are mutually *compatible*. The metaphysical realist, unlike the internal realist, is not precluded from appealing to a 'neutral' interpretation to determine the mutual consistency of T_1 and T_2, and thus has the resources from which the mutual incompatibility of T_1 and T_2 can be demonstrated. So, at most the foregoing argument shows that the internal realist is unable to appeal to the argument from incompatible but empirically equivalent theories to reject metaphysical realism; it cannot also be used as a general vindication of metaphysical realism (i.e. MR_3).

However, suppose Putnam can make a case for T_1 and T_2 being incompatible - then an argument can be made that the model-theoretic argument in tandem with the argument from incompatible empirically equivalent theories actually *supports* realism.[1] Let us tentatively assume that Putnam is able to make the following claims: (i) T_1 is an epistemically ideal theory, as such, (i′) it is true and complete and (i″) it contains as its evidential sentences the set of all and only true observation sentences Γ; (ii) T_2 is an epistemically ideal theory, as such, (ii′) it is true and complete and (ii″) it contains as

[1]The core of the argument was suggested by Nicholas Griffin.

its evidential sentences the set of all and only true observation sentences Γ; (iii) (as a result of (i″) and (ii″)) T_1 and T_2 are empirically equivalent; and (iv) T_1 and T_2 are mutually incompatible.

Now, given that T_1 and T_2 are mutually incompatible, there must be some sentence P which one asserts and the other denies (i.e. which is contained in one and its negation is contained in the other). Neither P nor $\neg$P can be members of Γ nor be derivable in any way from Γ.[1] Thus, in order for T_1 and T_2 to be mutually incompatible, each must contain a sentence which transcends the data. Furthermore, as each is epistemically ideal (i.e. true, in Putnam's sense), the evidence-transcending sentences contained in each must be true - i.e. both theories must contain *true* statements which are in principle undecidable by reference to the data. But this is possible only if one is a metaphysical realist - the claim that truth may transcend the evidence is precisely one of the things which makes one a metaphysical realist. T_1 and T_2 can only be incompatible empirically equivalent ideal theories if we assume a metaphysical realistic conception of truth! There are, however, two initially plausible anti-realist responses.

It is conceded that there must be an individuating sentence P which one theory contains and the other does not - otherwise T_1 and T_2 are not different theories. However, from this, it does not necessarily follow that P is *true*, even when viewed from within the theory which contains it. If we follow Dummett and assume that truth is co-extensive with decidability (where decidability involves the determination of the truth-

---

[1]For, suppose T_1 contains P and P is derivable from Γ; then T_2, which contains $\neg$P, would also contain P (as it contains Γ) and hence be inconsistent (i.e. not epistemically ideal, contrary to the assumption); similarly for T_2 containing P.

values of only observational sentences and their consequences), then we are forced to conclude that any theoretical sentence not derivable from Γ of either T_1 or T_2 must lack a truth-value - in particular, P lacks a truth-value. In other words, the anti-realist of a Dummettian persuasion can hold that T_1 and T_2 are indeed individuated on the basis of only one theory containing P, but that this does not force the realist consequence (i.e. that there can exist true sentences which transcends the data) as we need not be committed to the truth (or falsity) of P.

However, while this response may seem initially promising, it violates the initial conditions of the argument. If the individuating sentence P lacks a truth-value, in what sense are T_1 and T_2 mutually incompatible? If they are not incompatible, then T_1 and T_2 can be conjoined into the super-theory T_3. T_3 will, I agree, contain both P and $\neg$P, but as, according to this response, they both lack a truth-value, T_3 will not be inconsistent. If we are not bothered by one (true) theory containing sentences lacking truth-values, we can be in no way bothered by another (true) theory containing pairs of merely syntactic contradictories.

Secondly, the anti-realist may argue that, to avoid the realist consequence, the incompatibility must be located in Γ. Recall Putnam's internal realism: the sentences of Γ in both T_1 and T_2 are only syntactically identical. To be precise, we must semantically distinguish the evidential sentences of T_1 (interpreted according to SAT - call the set Γ_1) from the evidential sentences of T_2 (interpreted according to TAT - call the set Γ_2). We can then say that T_1 and T_2 are incompatible in the sense that T_1 contains some sentence $P_1 \in \Gamma_1$ which T_2 does not (similarly, T_2 will contain some sentence

$P_2 \epsilon \Gamma_2$ which T_1 does not). This will allow Putnam to maintain that T_1 and T_2 are mutually incompatible, but at the price of precluding him from saying that they are empirically equivalent - far from agreeing on *all* evidential sentences, they agree on *none*! So again this response violates the initial conditions for the argument.

All in all, the argument from incompatible empirically equivalent theories is not threatening to the metaphysical realist. In the first place, the anti-realist has not made a case for equating empirical equivalence with an equivalence of truthfulness (i.e. it does not non-question-beggingly follow that two theories empirically equivalent have the same degree of truthfulness). Secondly, there is serious question of whether there do (or could) exist such pairs of theories. The model-theoretic argument may be raised to illustrate how to construct such theories, but then either such theories must contain sentences whose truth transcends evidence (vindicating MR_4) or they cannot, contrary to appearances, be presumed incompatible (violating the initial conditions of the argument).

2.3.2 Conceptual Relativity

A closer reading of Putnam reveals that he is offering a slightly different argument to Goodman's argument from incompatible empirically equivalent theories. Such empirically equivalent theories are only *apparently* incompatible; they seem incompatible "when taken at face value", or have "what at least seem to be quite different ontologies",

but whose incompatibility can be resolved.[1]

Recall that two theories are empirically equivalent just in case they are verified or falsified by all the same experiences. Formally, two theories are empirically equivalent just in case all the sentences in their evidential base and predictive superstructure coincide. Two such theories are also incompatible (at least when taken at face value) when one contains a theoretical sentence denied by the other. However, Putnam maintains that such *prima facie* incompatibilities can be eliminated if the two theories are 'mutually relatively interpretable', where:

> The definition of T_1 *as relatively interpretable in* T_2 is: there exist possible definitions (i.e., formally possible definitions, whether these correspond to the meanings of the terms or not) of the terms of T_1 in the language of T_2 with the property that, if we 'translate' the sentences of T_1 into the language of T_2 by means of those definitions, then all the theorems of T_1 become theorems of T_2.[2]

Take, for example, a 'world' containing three items, x_1, x_2, and x_3, and two theories of that world, T_1 (Carnap's theory) which asserts that the world is populated with exactly three objects: x_1, x_2, and x_3; and T_2 (the Polish Logician's theory, which admits objects as mereological sums of other objects) which asserts that the world is populated with seven objects: x_1, x_2, x_3, x_1+x_2, x_1+x_3, x_2+x_3, and $x_1+x_2+x_3$.[3] Furthermore, suppose x_1 to be completely red, x_2 to be completely black, and x_3 to be completely green. T_1 and T_2 appear, at face value, to be incompatible in that the sentence:

17) There is at least one object which is partly red and partly black.

---

[1]Putnam (1982) p. 39.

[2]Putnam (1983a) p. 38.

[3]Putnam gives this example in at least two places: Putnam (1987a) and (1987c).

will come out false according to T_1 and true according to T_2; T_1 will instead contain the sentence:

18) There is no object which is partly red and partly black.

But, Putnam maintains, the incompatibility is only apparent; its source is to be found in T_1's and T_2's divergent interpretation of the neutral term 'object'. T_1 interprets 'object' parsimoniously in such a way that only x_1, x_2, and x_3 satisfy it, whereas T_2 interprets it liberally in such a way as to allow mereological sums to satisfy it.

Let T_1 introduce a new term, 'smobject', which is satisfied by that which is formed by conjoining two 'objects' (interpreted normally-to-T_1). The new definition will allow a translation of T_2's sentence "There is at least one object which is partly red and partly black." *into* T_1 as

19) There is at least one smobject which is partly red and partly black.

Thus, both T_2's sentence *and* T_1's translation of that sentence will come out true. In other words, T_2 contains (17), which comes true according to its own interpretation, whereas T_1 translates (17)-according-to-T_2 as (19), which also comes out true according to its own interpretation.

On the other hand, let T_2 introduce a new term, 'lobject' which is satisfied only by the atomistic components of 'objects' (interpreted normally-to-T_2). The new definition will allow a translation of (18)-according-to-T_1 *into* T_2 as:

20) There is no lobject which is partly red and partly black.

As long as each theory is allowed to interpret its own sentences, and there exists a truth- or theorem-preserving translation of the sentences of each theory into the other, there

is no real incompatibility between T_1 and T_2.

2.3.2.1 The Argument

Why is this a problem for metaphysical realism? According to MR_3, there is exactly one true and complete theory of the way the world is. Part of what this entails is that, for any two theories which are incompatible in virtue of assigning different truth-values to a particular sentence P, at least one of them must be false. However, two theories may be so incompatible on the surface yet each be mutually relatively interpretable. If so, there is no longer any reason to insist that, for any such pair of theories, at least one of them must be false. Thus, it is possible that there be distinct but equally true theories concerning some subject matter, and MR_3 is in jeopardy.

On the other hand, if, say T_1 and T_2 count as equally true, then it would seem that ontological categories are determined theoretically and not, as is assumed by the metaphysical realist's commitment to MR_{1b} (i.e. the anti-nominalistic thesis), by the mind-independent world itself. Putnam calls this the phenomenon of 'conceptual relativity' - in this case the very concept of an object is relative to each of T_1 and T_2.[1] What is lost, according to Putnam, is any theory-neutral understanding of what an object is. It is in this sense that, according to internal realism, objects are theory-relative, and MR_{1b} is in jeopardy.

---

[1]Notice that Putnam seems to be assuming that concepts are defined extensionally rather than intensionally. It is because 'object' interpreted according to T_1 has a different extension than 'object' interpreted according to T_2 that they appear *prima facie* incompatible.

The typical response from the metaphysical realist, says Putnam, is to invoke a 'cookie cutter' metaphor:

> Now, the classic metaphysical realist way of dealing with such problems is well known. It is to say that there is a single world (think of this as a piece of dough) which we can slice into pieces in different ways.[1]

That is, it is *we*, by our practice of theorizing, who decide the shape of the cookie cutter, but it is the world itself (the noumenal dough) which is being cut up. Moreover, it is due to the nature of the dough itself which allows it to be cut into different shapes. Putnam's response is that the metaphor founders on the question "What are the 'parts' of this dough?",[2] for no cookie-cutter-neutral answer can be given. The noumenal dough - the metaphysical realist's world - becomes an Kantian thing-in-itself of which nothing can be said:

> But talk of 'theory-independent objects' is hard to keep. The problem is that such talk may retain 'the world' but at the price of giving up any intelligible notion of *how* the world is. Any sentence that changes truth value on passing from one correct theory to another correct theory - an equivalent description - will express only a *theory-relative* property of the world. And the more such sentences there are, the more properties of the world will turn out to be theory relative. For example, if we conceded that [T_1 and T_2] are equivalent descriptions, then the property *being an object* (as opposed to a class or set of things) will be theory-relative.... The fact is, so many properties of 'the world' - starting with just the categorical ones, such as cardinality, particulars or universals, etc. - turn out to be 'theory relative' that 'the world' ends up as a *mere* 'thing in itself'. If one cannot say *how* 'the world' is, theory-independently, then talk of theories as descriptions of 'the world' is empty.[3]

---

[1]Putnam (1987c) p. 97.

[2]Putnam (1987c) p. 97 See also (1987a) pp. 32-37.

[3]Putnam (1983a) pp. 44-45. See also Putnam (1976b) pp. 130-133, (1982) pp. 40-41, (1987b) pp. 26-28, (1988) pp. 107-116, and (1992) chapter 6.

2.3.2.2 Responses

The initial obvious response to the phenomenon of conceptual relativity is to insist that T_1 and T_2 are not, in fact, distinct theories. They are merely notational variants of each other - different ways of describing the same thing. They amount to no more difference than a theory stated in either French or English. Against this Putnam points out that the translations used to eliminate the *prima facie* incompatibility between T_1 and T_2 are merely truth-preserving; they are not also meaning-preserving:

> Relative interpretability is a purely *formal* relation; it in no way involves the *meanings* of the terms... Thus mutual relative interpretability, however interesting as a formal relation, guarantees no sort of sameness of meaning or even subject matter between theories; it only testifies to the existence of similar formal structures in both theories. Conceivably two theories about wholly disparate subject matters - say, an axiomatic system of genetics and an axiomatic system of number theory - could turn out to be mutually relatively interpretable, but they would hardly be equivalent in cognitive meaning.[1]

If Putnam is correct in this, and I am willing to concede it, then the fact that two *prima facie* distinct theories concerning some single subject matter can be truth-preservingly translated into each other is no reason to regard them as mere notational variants.

However, the argument from conceptual relativity involves an ambiguity in the notion of truth reminiscent of that discussed in §2.3.1.2. According to the metaphysical realist, a theory is true just in case its terms refer and its sentences accurately describe the world. Combining this with MR_{1b}, for any two theories with differing ontologies, as

---

[1]Putnam (1983a) pp. 38-39. Of course Putnam is a bit quick in this remark. MR_3 is only in jeopardy is it is possible for there to be two *distinct* but equally true theories *concerning some single subject matter.* That there may be one true theory concerning astronomy and another true theory concerning paleontology is in no way embarrassing to the metaphysical realist. T_1 and T_2 are, however, concerning with the same subject matter - namely the population of the 'world' and the colour composition of its members.

at most one of those ontologies will correspond to the way the world divides itself up, at least one must be false. Of course, it is also part of metaphysical realism that we may never be in a position to determine which ontology is accurate, but that is only to reinforce the basic view that truth is primarily a metaphysical as opposed to epistemological notion. In other words, when faced with a pair of such theories, whether they be relatively mutually interpretable or not, the proper metaphysical realist attitude would be to insist on the falsity of at least one. Thus, MR_2 and MR_{1b} simply precludes one from accepting the possibility that two distinct theories concerning some single subject matter can equally be true; i.e. they preclude one from seeing the phenomenon of conceptual relativity as an argument *against* MR_3. The phenomenon of conceptual relativity can only provide an argument against MR_3 if a rejection of MR_{1b} is assumed. But, unless independent evidence is offered against MR_{1b}, the argument begs the question against the metaphysical realist.

On the other hand, the phenomenon of conceptual relativity can be viewed as providing an argument against MR_{1b}, but only by assuming a rejection of MR_3. One can only conclude that objects are theory-relative, say to T_1 and T_2, if one accepts that each of T_1 and T_2 are equally true; if a *false* theory asserts that there are Xs, it does not follow that there *are* Xs. So, as an argument against MR_{1b}, it assumes a rejection of MR_3, and thus begs the question against the realist. Putnam's argument is circular: in order for the phenomenon of conceptual relativity to provide evidence against MR_{1b}, it must assume a rejection of MR_3, and in order for it to provide evidence against MR_3, it must assume a rejection of MR_{1b}. Unless independent arguments can be given against either

of MR_{1b} or MR_3, the phenomenon of conceptual relativity cannot be seen as providing an argument against metaphysical realism.

Secondly, MR_3 asserts that there exists a unique true *and complete* theory of the way the world is. As long as T_1 and T_2 are not, in fact, incompatible (in virtue of the mutual relative interpretability), then, as was suggested in §2.3.1.1, they can be conjoined into a consistent super-theory T_3. According to T_3, *objects* come in two sorts: lobjects and smobjects. Smobjects consist of a conjunction of lobjects, and lobjects consist of the atomic components of smobjects; no lobject is partly red and partly black, but at least one smobject is partly red and partly black.

The upshot of internal realism is that the world can be correctly described in many different ways. The metaphysical realist can accept this, as long as the different ways are not strictly incompatible. But, if the various descriptions are not incompatible, there is nothing which precludes one putting together all of those descriptions into one big description. Each of the sub-descriptions only provide partial descriptions; only the super-description lays claim to being complete.

CONCLUSION

Realism, I submit, is not in serious danger from the semantic arguments of Michael Dummett or Hilary Putnam.

Regarding Dummett, suppose it is possible to construct a theory of meaning for a language. Suppose further that it could conclusively be shown that such a theory must take the form of a molecular axiomatic system in the manner proposed by Dummett. Suppose finally that such a theory must ultimately be grounded in a manifestable capacity to use sentences of the language in various communally accepted ways. Even granting all of these suppositions, it simply does not follow that we could never have acquired the concept of recognition-transcendent truth or could ever manifest our knowledge of the recognition-transcendent truth-conditions of any sentence. There is no reason, then, to doubt that such a concept is a legitimate and coherent one or to doubt that sentences have recognition-transcendent truth-conditions.

In discussing Dummett's argument, it was necessary to distinguish between recognition-transcendent and unrecognizable truth-conditions. Realism in its semantic guise, as I characterized it, is the doctrine that sentences have recognition-transcendent truth-conditions, *not* the doctrine that any have unrecognizable truth-conditions. The concept of recognition-transcendence is, I suggest, a non-epistemic one in the sense that, according to the realist, a sentence may have recognition-transcendent truth-conditions regardless of the state of human knowledge. The concept of unrecognizability, however, cannot be a non-epistemic one in that whether a sentence has unrecognizable truth-conditions certainly does depend upon the state of human knowledge.

As we saw, Dummett's negative programme assumes that acceptance of a concept of truth as recognition-transcendent commits one to the existence of sentences with unrecognizable truth-conditions. But, once the two concepts are carefully separated and understood, the assumption is seen to be unwarranted. Thus, even if it were the case that one could not have acquired the concept of an unrecognizable truth-condition, or that one could not manifest knowledge of unrecognizable truth-conditions, it simply would not follow that (semantic) realism was in any serious difficulty. Indeed, a concept of truth as recognition-transcendent seems to be necessitated by the failure of a purely epistemic concept to accommodate all of the compositional facts about how we use our language. Thus, not only does (semantic) realism resist Dummett's attack, it would appear to be vindicated.

Regarding Putnam, the main claim weaving through all of his arguments is that it is incoherent to suppose that a theory (or description, or point-of-view) can be understood (or discussed, or conceived) from its *outside*. SAT serves as an 'intended' interpretation of an ideal theory, he claims, because there is no way to judge that it fails to satisfy certain required constraints - according to SAT itself, it *does* satisfy all required constraints. One cannot be a brain in a vat, he says, for the BIV hypothesis cannot be judged true from *within* the vat world. Ontology must be theory-relative, he maintains, for one cannot stand outside of theory to determine which of several alternative ontologies are correct. However, once his arguments are clearly understood, it becomes clear that little actual support is given for the main claim. One simply cannot take the model-theoretic, vat, and empirical equivalence arguments as making it any more

plausible for, on the one hand, they can all be answered, and on the other, their acceptability rest upon it.

Where, specifically, did the anti-realist go wrong? The anti-realist wants, crudely, to reduce truth to verifiability. She tries to ensure this by arguing that truth-conditions must be co-extensive with verification-conditions. According to Dummett, only knowledge of verification-conditions can consistently be manifested. According to Putnam, we cannot have a concept of the truth of a theory apart from it satisfying theory-relative criteria for it. The realist, as we have seen, can accept that truth-conditions and verification-conditions are co-extensive without conceding that truth reduces to verifiability. Even if truth-conditions are in fact co-extensive with verification-conditions, it does not follow that the limits of truth are the limits of what we can verify, or, in metaphysical lingo, that the limits of 'reality' are the limits of our experience.

The main argument advanced was to the effect that the co-extensiveness of truth and verification is no argument for the claim that truth reduces to (or is dependent upon) verification: that a truth-condition is recognition-transcendent is one 'fact', and that a truth-condition is recognizable is another 'fact', even if those two 'facts' always coincide.

However, suppose that there does exist a sentence whose truth-condition actually transcends the possibility of our verification (for the realist must at least, I argued, remain open to such possibilities). What would follow from this? We could never *know* that such a truth-condition obtained when it did, but that is only to say that an

unverifiable event cannot be verified to have taken place - but *logic*, not *anti-realism* tells us this. The anti-realist's mistake is to accept a certain chain of reasoning: we can have no evidence that any such truth-condition obtained, so we could have no evidence that there *are* any such truth-conditions, so we are warranted to conclude that there are *no* such truth-conditions. This chain of reasoning is, of course, fallacious. It is fallacious on an superficial level,[1] but also an a deeper level: it is fallacious because it assumes what it purports to establish - that the limits of what we can establish is the case are the limits of what *is* the case.

Nonetheless, there are some valuable lessons to be learned from the semantic anti-realisms of Dummett and Putnam. In the Introduction I distinguished two elements in any given realism towards some subject matter: an *existential* element and what I called a *meta-theoretic* element. For illustration, consider a person, S, who holds that 'pain' is a private mental state which humans and other sentient beings might be in, and that sentences expressing attributions of pain are true (or false) independently of our capacity to determine which. For instance, S would say that "Mark's cat Cinder is in pain" is true just in case Cinder is in the (private) mental state of being in pain - a state which, of course, no one could verify as obtaining. If S is like the rest of us, she is not likely to just guess whether Cinder is in such a state when contemplating *when* to make such an assertion - she will look for behavioral clues. The point is that S takes such

---

[1]Compare to the 'lottery' fallacy: just because I cannot have any reason to claim, for any given person who purchased a lottery ticket, that they will win, it does not follow that I cannot have any reason to claim that *someone* will win.

behaviour *only* as (reliable) *evidence* of the truth of the utterance, not as *constituting* the truth of the utterance. Contrast S with P, who denies that 'pain' is a private mental state at all. According to P, 'pain' is to be *identified* with overt types of behaviour. P thus denies the *existential* claim made by S, and thus, relative to this denial, P is to be considered an anti-realist. P need not, however, deny the *meta-theoretical* element of S's realism - P might agree that Cinder can be in pain even when no one is in a position to determine whether Cinder displays the overt behaviour. P, then, need not be considered an anti-realist in the second sense.[2]

The point is that, even if I have succeeding in demonstrating that the realist's meta-theoretic claims have not been repudiated, I have not thereby ended the realism/anti-realism debate, as least not as it has traditionally (that is, metaphysically) been understood. Nothing whatsoever in my counter-arguments to Dummett and Putnam can be taken as evidence that there are private mental states, or objective values, or macroscopic material objects, or any other existential claim this or that realist would want to make. What I wanted to show was that the traditionally metaphysical debates cannot be settled by (at least these) semantic arguments, *not* that repudiating semantic anti-realism vindicates any specific existential claim - this is all part of my denial that semantic considerations *alone* can have metaphysical consequences. But, if semantic considerations (alone) cannot settle such debates, what can?

What we are left with is what we have always been left with in philosophy -

---

[2]Dummett spends much time separating his anti-realism from such reductionistic anti-realisms. See Dummett (1963b), (1969), (1978), and (1991b) Ch. 15.

rational argumentation. The proponent of mental realism must marshall the strongest arguments she can as to why postulating such entities is the most reasonable belief one can have, and why behaviourism fails to be satisfactory: she must give us sufficient *reason - argument* - to accept her view over that of the behaviourist. But, argumentation itself is a human practice. It is a practice couched in our more general linguistic practices, and it invokes particular conceptions of rationality.

Dummett has convinced me of two things. First of all, belief building - argumentation - is a practice inherently embedded in linguistic practice. In order to understand the former, we need to understand the latter. But, we cannot take our linguistic practices at face value - on the one hand, it may not be entirely clear what are practices actually are (or at least how we should interpret them) and on the other hand, actual practices are susceptible to critical scrutiny. Secondly, claims to truth must be sensitive to *evidence*, and evidence properly so-called must be capable of being *recognized* as such. Similarly, Putnam has persuasively pointed out that what *we count* as evidence (at any given time) is a function of what our beliefs concerning *rationality* itself are. Specific conceptions of rationality are, by and large, historically constituted artifacts. In our philosophizing we must be aware that what we count as acceptable, or explanatory, or true is fluid and susceptible to critical scrutiny.

These are the valuable insights of Dummettian and Putnamian anti-realism. When the realist, recognizing that truth is distinct from the provable here and now or ever, concludes that evidence is irrelevant or that cannons of rationality are merely subjective artifacts, wanders in idle metaphysical speculation away from our familiar

world of experience, the anti-realist is quite right to object and to reel them back in. But, anti-realists are capable of their excesses as well.

REFERENCES

Anderson, David Leech (1993)
"What is the Model-Theoretic Argument?", The Journal of Philosophy, 90, 311-322.

Appel, Kenneth and Haken, Wolfgang (1977)
"The Solution to the Four-Colour Problem", Scientific American, 237, 108-121.

Appiah, Anthony (1986)
For Truth in Semantics, Oxford: Basil Blackwell.

Bailey, George (1983)
"Putnam and Metaphysical Realism", International Studies in Philosophy, 15, 11-14.

Bergstrom, Lars (1984)
"Underdetermination and Realism", Erkenntnis, 21, 349-365.

Boolos, George S. and Jeffrey, Richard C. (1989)
Computability and Logic, Cambridge: Cambridge University Press.

Boyd, Richard N. (1973)
"Realism, Underdetermination, and a Causal Theory of Evidence", Nous, 7, 1-12.

Brandom, Robert (1976)
"Truth and Assertibility", The Journal of Philosophy, 73, 137-149.

Brouwer, L.E.J. (1908)
"The Unreliability of the Logical Principles", Collected Works Volume 1, Heyting (ed.), Amsterdam: North-Holland Publishing Company, 1975, 107-111.

_____ (1923)
"On the Significance of the Principle of Excluded Middle in Mathematics, Especially in Function Theory", Frege to Gödel: A Sourcebook in Mathematical Logic 1879-1931, van Heijenoort (ed.), van Heijenoort and Bauer-Mengelberg (trans.), Cambridge: Harvard University Press, 1967, 334-345.

_____ (1952)
"Historical Background, Principle and Methods of Intuitionism", Collected Papers Volume I: Philosophy and the Foundations of Mathematics, Heyting (ed.), Amsterdam: North-Holland, 1975, 508-515.

_____ (1975)
"Essentially Negative Properties", Collected Papers Volume I: Philosophy and the Foundations of Mathematics, Heyting (ed.), Amsterdam: North-Holland, 1975, 478-479.

Brown, Curtis (1988)
"Internal Realism: Transcendental Idealism?", Midwest Studies in Philosophy, 12, 145-155.

Brueckner, Anthony L. (1984)
"Putnam's Model-Theoretic Argument Against Metaphysical Realism", Analysis 44, 134-140.

_____ (1986)
"Brains in a Vat", The Journal of Philosophy, 83, 148-167.

Casati, Roberto and Dokic, Jérôme (1991)
"Brains in a Vat, Language and Metalanguage", Analysis, 51, 91-93.

Clendinnen, F. John (1989)
"Realism and the Underdetermination of Theory", Synthese, 81, 63-90.

Collier, John D. (1990)
"Could I Conceive of Being a Brain in a Vat?", Australasian Journal of Philosophy, 68, 413-419.

Cooper, Neil (1978)
"The Law of Excluded Middle", Mind, 87, 161-180.

Craig, Edward (1982)
"Meaning, Use and Privacy", Mind, 91, 541-564.

Crosthwaite, Jan (1983)
"On The Theoretical Representation of Linguistic Ability", Pacific Philosophical Quarterly, 64, 151-164.

Currie, Gregory (1982)
"A Note on Realism", Philosophy of Science, 49, 263-267.

_____ (1993)
"On The Road to Antirealism", Inquiry, 36, 465-483.

Currie, Gregory and Essenberger, Peter (1983)
"Knowledge and Meaning", Noûs, 17, 267-280.

Daniels, Charles B. (1990)
"A Note on Negation", Erkenntnis, 32, 478-479.

Davidson, Donald (1967)
"Truth and Meaning", Inquiries into Truth & Interpretation, Oxford: Clarendon Press, 1981, 17-36.

_____ (1981)
"A Coherence Theory of Truth and Knowledge", Truth and Interpretation, LePore (ed.), Oxford: Basil Blackwell, 1986, 307-319.

Davies, David (1987)
"How Not to Outsmart the Anti-Realist", Analysis, 47, 1-8.

Davies, Martin (1981)
Meaning, Quantification, Necessity: Themes in Philosophical Logic, London: Routledge & Kegan Paul.

Dell'Utri, Massimo (1990)
"Choosing Conceptions of Realism: The Case of the Brains in a Vat", Mind, 99, 79-90.

Devitt, Michael (1991)
Realism and Truth; 2nd Edition, Oxford: Basil Blackwell.

Dummett, Michael (1959a)
"Truth", Truth and Other Enigmas, London: Duckworth, 1978, 1-24.

_____ (1959b)
"Wittgenstein's Philosophy of Mathematics", Truth and Other Enigmas, London: Duckworth, 1978, 166-185.

_____ (1963a)
"The Philosphical Significance of Gödel's Theorem", Truth and Other Enigmas, London: Duckworth, 1978, 186-201.

_____ (1963b)
"Realism", Truth and Other Enigmas, London: Duckworth, 1978, 145-165.

_____ (1969)
"The Reality of the Past", Truth and Other Enigmas, London: Duckworth, 1978, 358-374.

_____ (1970)

"Wang's Paradox", Truth and Other Enigmas, London: Duckworth, 1978, 248-268.

_____ (1973a)
Frege: Philosophy of Language, London: Duckworth.

_____ (1973b)
"The Justification of Deduction", Truth and Other Enigmas, London: Duckworth, 1978, 290-318.

_____ (1973c)
"The Philosophical Basis of Intuitionistic Logic", Truth and Other Enigmas, London: Duckworth, 1978, 214-247.

_____ (1975)
"What is a Theory of Meaning? (I)", Mind and Language, Guttenplan (ed.), Oxford: Clarendon Press, 1975, 97-138.

_____ (1976a)
"Is Logic Empirical?", Truth and Other Enigmas, London: Duckworth, 1978, 269-289.

_____ (1976b)
"What is a Theory of Meaning? (II)", Truth and Meaning: Essays in Semantics, Evans and McDowell (eds.), Oxford: Clarendon Press, 1976, 67-137.

_____ (1977)
Elements of Intuitionism, Oxford: Clarendon Press.

_____ (1978)
"Preface", Truth and Other Enigmas, London: Duckworth, 1978, ix-li.

_____ (1982)
"Realism", Synthese, 52, 55-112.

_____ (1987)
"Replies to Essays", Michael Dummett: Contributions to Philosophy, Taylor (ed.), Dordrecht: Martinus Nijhoff Publishers, 221-330.

_____ (1990)
"The Source of the Concept of Truth", Meaning and Method: Essays in Honor of Hilary Putnam, Boolos (ed.), Cambridge: Cambridge University Press, 1-15.

_____ (1991a)
Private communication to Nicholas Griffin.

_____ (1991b)
The Logical Basis of Metaphysics, Cambridge: Harvard.

Edgington, Dorothy (1981)
"Meaning, Bivalence and Realism", Proceedings of the Aristotelian Society, 81, 153-173.

_____ (1985)
"Verificationism and Manifestations of Meaning", Aristotelian Society, Supp. 59, 33-52.

Fales, Evan (1988)
"How to be a Metaphysical Realist", Midwest Studies in Philosophy, 12, 253-274.

Feldman, Susan (1984)
"Refutation of Dogmatism: Putnam's Brains in Vats", Southern Journal of Philosophy, 22, 323-330.

Field, Hartry (1972)
"Tarski's Theory of Truth", The Journal of Philosophy, 69, 347-375.

_____ (1982)
"Realism and Relativism", The Journal of Philosophy, 79, 553-567.

Frege, Gottlob (1892)
"On Sense and Meaning", Translations from the Philosophical Writings of Gottlob Frege, Geach and Black (eds.), Black (trans.), Oxford: Basil Blackwell, 1952, 56-78.

_____ (1918)
"Thoughts (Der Gedanke)", Logical Investigations, Gearch (ed. and trans.), New Haven: Yale University Press, 1977, 1-30.

Gardiner, Mark Q. (1994a)
"Just More Theory?", accepted for publication in Australasian Journal of Philosophy.

_____ (1994b)
"Tymozcko on Putnam's Brain", accepted for publication in Erkenntnis.

George, Alexander (1987)
"Reply to Weir on Dummett and Intuitionism", Mind, 96, 404-406.

Glymour, Clark (1970)

"Theoretical Realism and Theoretical Equivalence", Boston Studies in the Philosophy of Science, VIII, 275-288.

Goodman, Nelson (1978)
Ways of Worldmaking, Indianapolis: Hackett.

_____ (1980)
"On Starmaking", Synthese, 45, 211-215.

_____ (1984)
Of Mind and Other Matters, Cambridge: Harvard University Press.

Griffin, Nicholas (1993)
Private communication from Dr. Griffin: "The High Cost of Anti-Realism".

Haack, Susan (1982)
"Dummett's Justification of Deduction", Mind, 91, 216-239.

Hacking, Ian (1983)
Representing and Intervening, Cambridge: Cambridge University Press.

Harman, Gilbert (1982)
"Metaphysical Realism and Moral Relativism: Reflections on Hilary Putnam's *Reason, Truth and History*", The Journal of Philosophy, 79, 568-575.

Harrison, Bernard (1983)
"Meaning, Truth and Negation", Proceedings of the Aristotelian Society, Supp. 57, 179-204.

Harrison, J. (1985)
"Professor Putnam on Brains in Vats", Erkenntnis, 23, 55-57.

Heller, Mark (1988)
"Putnam, Reference, and Realism", Midwest Studies in Philosophy, 12, 113-127.

Heyting, Arend (1956)
Intuitionism: An Introduction, Amsterdam: North-Holland.

Horwich, Paul (1982)
"Three Forms of Realism", Synthese, 51, 181-201.

Hossack, Keith G. (1990)
"A Problem about the Meaning of Intuitionist Negation", Mind, 99, 207-219.

Iseminger, Gary (1988)
"Putnam's Miraculous Argument", Analysis, 48, 190-195.

Johnson, Jeffrey L. (1991)
"Making Noises in Counterpoint or Chorus: Putnam's Rejection of Relativism", Erkenntnis, 34, 323-345.

Kirkham, Richard L. (1989)
"What Dummett Says About Truth and Linguistic Competence", Mind, 98, 207-224.

Koethe, John (1979)
"Putnam's Argument Against Realism", The Philosophical Review, 88, 92-99.

Kripke, Saul (1972)
Naming and Necessity, Cambridge: Harvard University Press.

Kukla, André (1993)
"Laudan, Leplin, Empirical Equivalence and Underdetermination", Analysis, 53, 1-7.

Landini, Gregory (1987)
"Putnam's Model-Theoretic Argument, Natural Realism, and the Standard Conception of Theories", Philosophical Papers, 16, 209-233.

Laudan, Larry and Leplin, Jarrett (1991)
"Empirical Equivalence and Underdetermination", The Journal of Philosophy, 88, 449-472.

_____ (1993)
"Determination Underdeterred", Analysis, 53, 8-16.

Lepore, Ernest and Loewer, Barry (1988)
"A Putnam's Progress", Midwest Studies in Philosophy, 12, 459-473.

Lewis, David (1973a)
Counterfactuals, Cambridge: Harvard University Press.

_____ (1973b)
"Counterfactuals and Comparative Possibility", Ifs: Conditionals, Belief, Decision, Chance, and Time, Harper, Stalnaker, and Pearce (eds.), Dordrecht: D. Reidel Publishing Company, 1981, 57-85.

_____ (1984)

"Putnam's Paradox", Australasian Journal of Philosophy, 62, 221-236.

Loar, Brian (1987)
"Truth Beyond all Verification", Michael Dummett: Contributions to Philosophy, Taylor (ed.), Dordrecht: Martinus Nijhoff Publishers, 81-116.

Lukes, Steven (1978)
"The Underdetermination of Theory by Data", Aristotelian Society Supp. 52, 93-107.

Luntley, Michael (1988)
Language, Logic and Experience: The Case for Anti-Realism, LaSalle: Open Court.

Lycan, William (1984)
Logical Form in Natural Language, Cambridge: MIT Press.

Malachowski, Alan (1986)
"Metaphysical Realist Semantics: Some Moral Desiderata", Philosophia, 16, 167-174.

McDowell, John (1976)
"Truth Conditions, Bivalence, and Verificationism", in Truth and Meaning: Essays in Semantics, Gareth and McDowell (eds.), Oxford: Clarendon Press, 1976, 42-66.

_____ (1987)
"In Defense of Modesty", Michael Dummett: Contributions to Philosophy, Taylor (ed.), Dordrecht: Martinus Nijhoff Publishers, 59-80.

McGinn, Colin (1976)
"Truth and Use", Reference, Truth and Reality: Essays on the Philosophy of Language, Platts (ed.), London: Routledge & Kegan Paul.

_____ (1979)
"An A Priori Argument for Realism", The Journal of Philosophy, 76, 113-133.

_____ (1981)
"Reply to Tennant", Analysis, 41, 120-122.

_____ (1982a)
"Realist Semantics and Content-Ascription", Synthese, 52, 113-134.

_____ (1982b)
"Two Notions of Realism?", Philosophical Topics, 13, 123-134.

McIntyre, Jane (1984)
"Putnam's Brains", Analysis, 44, 59-61.

Melchert, Norman (1986)
"Metaphysical Realism and History", Analysis, 46, 36-38.

Merrill, G.H. (1980)
"The Model-Theoretic Argument Against Realism", Philosophy of Science, 47, 69-81.

Meyer, Robert (1980)
"Syntactic Treatment of Negation", Analysis, 40, 74-78.

Mitchell, Samuel William (1992)
"Dummett's Intuitionism is Not Strict Finitism", Synthese, 90, 437-458.

Moriconi, Enrico and Napoli, Ernesto (1988)
"Dummett's Transcendence", Philosophia, 18, 371-383.

Moser, Paul (1990)
"A Dilemma for Internal Realism", Philosophical Studies, 59, 101-106.

Newton-Smith, W. (1978)
"The Underdetermination of Theory by Data", Aristotelian Society, Supp. 52, 71-91.

Page, James (1991)
"Dummett's Mathematical Antirealism", Philosophical Studies, 63, 327-342.

Peacocke, Christopher (1980)
"Causal Modalities and Realism", Reference, Truth and Reality, Platts (ed.), London: Routledge & Kegan Paul, 1980, 41-68.

Prawitz, Dag (1980)
"Intuitionistic Logic: A Philosophical Challenge", Logic and Philosophy, von Wright (ed.), The Hague: Martinus Nijhoff, 1980, 1-10.

_____ (1987)
"Dummett on a Theory of Meaning and its Impact on Logic", Michael Dummett: Contributions to Philosophy, Taylor (ed.), Dordrecht: Nijhoff, 117-165.

Preston, John (1991)
"On Some Objections to Relativism", Ratio, 5, 57-73.

Price, Huw (1983)
"Sense, Assertion, Dummett and Denial", Mind, 92, 161-173.

_____ (1990)
"Why 'Not'?", Mind, 99, 221-238.

Putnam, Hilary (1960)
"Do True Assertions Correspond to Reality", Mind, Language and Reality: Philosophical Papers Volume 2, Cambridge: Cambridge University Press, 1975, 70-84.

_____ (1973)
"Explanation and Reference", Mind, Language and Reality: Philosophical Papers Volume 2, Cambridge: Cambridge University Press, 1975, 196-214.

_____ (1974)
"Language and Reality", Mind, Language and Reality: Philosophical Papers Volume 2, Cambridge: Cambridge University Press, 1975, 272-290.

_____ (1975)
"The Meaning of 'Meaning'", Mind, Language and Reality: Philosophical Papers Volume 2, Cambridge: Cambridge University Press, 1975, 215-271.

_____ (1976a)
"Meaning and Knowledge: The John Locke Lectures", Meaning and the Moral Sciences, London: Routledge & Kegan Paul, 1978, 7-80.

_____ (1976b)
"Realism and Reason", Meaning and the Moral Sciences, London: Routledge & Kegan Paul, 1978, 123-140.

_____ (1976c)
"Reference and Truth", Realism and Reason: Philosophical Papers Volume 3, Cambridge: Cambridge University Press, 1983, 69-86.

_____ (1978)
"Reference and Understanding", Meaning and the Moral Sciences, London: Routledge & Kegan Paul, 1978, 97-117.

_____ (1980)
"Models and Reality", Realism and Reason: Philosophical Papers Volume 3, Cambridge: Cambridge University Press, 1983, 1-25.

_____ (1981a)

"Philosophers and Human Understanding", Realism and Reason: Philosophical Papers Volume 3, Cambridge: Cambridge University Press, 1983, 184-204.

_____ (1981b)
Reason, Truth and History, Cambridge: Cambridge University Press.

_____ (1981c)
"Why Reason Can't Be Naturalized", Realism and Reason: Philosophical Papers Volume 3, Cambridge: Cambridge University Press, 1983, 229-247.

_____ (1981d)
"Why There Isn't a Ready-Made World", Realism and Reason: Philosophical Papers Volume 3, Cambridge: Cambridge University Press, 1983, 205-228.

_____ (1982)
"A Defense of Internal Realism", Realism with a Human Face, Cambridge: Harvard University Press, 1990, 30-42.

_____ (1983a)
"Equivalence", in Realism and Reason: Philosophical Papers Volume 3, Cambridge: Cambridge University Press, 1983, 26-43.

_____ (1983b)
"Introduction", Realism and Reason: Philosophical Papers Volume 3, Cambridge: Cambridge University Press, 1983, vii-xviii.

_____ (1984a)
"Is the Causal Structure of the Physical Itself Something Physical?", Realism with a Human Face, Cambridge: Harvard University Press, 1990, 80-95.

_____ (1984b)
"Is Water Necessarily H_2O?", Realism with a Human Face, Cambridge: Harvard University Press, 1990, 54-79.

_____ (1986)
"Why is a Philosopher?", Realism with a Human Face, Cambridge: Harvard University Press, 1990, 105-119.

_____ (1987a)
The Many Faces of Realism, LaSalle: Open Court.

_____ (1987b)
"Realism with a Human Face", Realism with a Human Face, Cambridge: Harvard University Press, 1990, 3-29.

_____ (1987c)
"Truth and Convention", Realism with a Human Face, Cambridge: Harvard University Press, 1990, 96-104.

_____ (1988)
Representation and Reality, Cambridge: The M.I.T. Press.

_____ (1989)
"Model-Theory and the 'Factuality' of Semantics", Reflections on Chomsky, George (ed.), Oxford: Basil Blackwell, 213-232.

_____ (1990)
"Preface", Realism with a Human Face, Cambridge: Harvard University Press, 1990, vii-xi.

_____ (1991)
"Replies and Comments", Erkenntnis, 34, 401-424.

_____ (1992)
Renewing Philosophy, Cambridge: Harvard University Press.

Quine, W.V.O (1950)
Methods of Logic, New York: Holt, Rinehart, and Winston.

_____ (1957)
"Speaking of Objects", Ontological Relativity and Other Essays, New York: Columbia University Press, 1969, 1-25.

_____ (1960)
Word and Object, Cambridge: MIT Press.

_____ (1966)
"Existence and Quantification", Ontological Relativity and Other Essays, New York: Columbia University Press, 1969, 91-113.

_____ (1968)
"Ontological Relativity", Ontological Relativity and Other Essays, New York: Columbia University Press, 26-68.

_____ (1970)
"On the Reasons for Indeterminacy of Translation", The Journal of Philosophy, 67, 178-183.

_____ (1975)

"On Empirically Equivalent Systems of the World", Erkenntnis, 9, 313-328.

_____ (1990)
The Pursuit of Truth, Cambridge: Harvard University Press.

Rasmussen, Stig Alstrup and Ravnkilde, Jens (1982)
"Realism and Logic", Synthese, 52, 379-437.

Resnick, Michael D. (1987)
"You Can't Trust an Ideal Theory to Tell the Truth", Philosophical Studies, 52, 151-160.

Rosenberg, Jay F. (1983)
Thinking Clearly about Death, Englewood Cliffs: Prentice Hall.

Russell, Bertrand (1910)
Principia Mathematica, Cambridge: University Press.

Scruton, Roger (1976)
"Truth-Conditions and Criteria", Proceedings of the Aristotelian Society, Supp. 50, 193-216.

Sintonen, Matti (1982)
"Realism and Understanding", Synthese, 52, 347-378.

Sklar, Lawrence (1982)
"Saving the Noumena", Philosophical Topics, 13, 89-110.

Slater, B.H. (1988)
"Excluding the Middle", Crítica, 20, 55-71.

Smart, J.C.C. (1982)
"Metaphysical Realism", Analysis, 42, 1-3.

Smith, Peter (1983)
"Smart's Argument for Realism", Analysis, 43, 74-78.

_____ (1984)
"Could We be Brains in a Vat?" Canadian Journal of Philosophy, 14, 115-123.

_____ (1986)
"Metaphysical Realism and Historical Interpretation", Analysis, 46, 157-158.

Solomon, Miriam (1990)

"On Putnam's Argument for the Inconsistency of Relativism", The Southern Journal of Philosophy, 28, 213-220.

Stalnaker, Robert (1968)
"A Theory of Conditionals", Ifs: Conditionals, Belief, Decision, Chance, and Time, Harper, Stalnaker, and Pearce (eds.), Dordrecht: D. Reidel Publishing Company, 1981, 41-55.

_____ (1980)
"A Defense of Conditional Excluded Middle", Ifs: Conditionals, Belief, Decision, Chance, and Time, Harper, Stalnaker, and Pearce (eds.), Dordrecht: D. Reidel Publishing Company, 1981, 87-104.

Stephens, James and Russow, Lilly-Marlene (1985)
"Brains in Vats and the Internalist Perspective", Australasian Journal of Philosophy, 63, 205-212.

Tarski, Alfred (1931)
"The Concept of Truth in Formalized Languages", Logic, Semantics, Metamathematics, Oxford: Clarendon Press, 1956, 152-278.

_____ (1944)
"The Semantic Conception of Truth and the Foundations of Semantics", Readings in Philosophical Analysis (1949), Feigl and Sellars (eds.), Atascadero: Ridgeview, 52-84.

Taylor, Barry (1991)
"'Just More Theory': A Maneouvre in Putnam's Model-Theoretic Argument for Antirealism", Australasian Journal of Philosophy, 69, 152-166.

Tennant, Neil (1981)
"Is This a Proof I See Before Me?", Analysis, 41, 115-119.

_____ (1984)
"Were Those Disproofs I Saw Before Me?", Analysis, 46, 68-72.

_____ (1985)
"Weir and Those 'Disproofs' I Saw Before Me", Analysis, 45, 208-212.

_____ (1987)
Anti-Realism and Logic: Truth as Eternal, Oxford: Clarendon Press.

Throop, William (1989)
"Relativism and Error: Putnam's Lessons for the Relativist", Philosophy and

Phenomenological Research, 49, 675-686.

Throop, William and Doran, Katheryn (1991)
"Putnam's Realism and Relativity: An Uneasy Balance", Erkenntnis, 34, 357-369.

Tymoczko, Thomas (1989)
"In Defense of Putnam's Brains", Philosophical Studies, 57, 281-297.

Vision, Gerald (1988)
Modern Anti-Realism and Manufactured Truth, London: Routledge & Kegan Paul.

Weir, Alan (1983)
"Truth Conditions and Truth Values", Analysis, 43, 176-180.

_____ (1985)
"Rejoinder to Tennant", Analysis, 46, 68-72.

_____ (1986)
"Dummett and Meaning and Classical Logic", Mind, 95, 465-477.

Weiss, Bernhard (1992)
"Can an Anti-Realist be Revisionary about Deductive Inference?" Analysis, 52, 216-224.

Williamson, Timothy (1988)
"Bivalence and Subjunctive Conditionals", Synthese, 75, 405-421.

Wright, Crispin (1987)
Realism, Meaning and Truth, Oxford: Basil Blackwell.

_____ (1986)
"How Can a Theory of Meaning be a Philosophical Project?", Mind and Language, 1, 31-43.

_____ (1988)
"Realism, Antirealism, Irrealism, Quasi-Realism", Midwest Studies in Philosophy, 12, 25-49.

_____ (1992)
Truth and Objectivity, Cambridge: Harvard University Press.

Wright, John (1989)
"Realism and Equivalence", Erkenntnis, 31, 109-128.

Young, James (1987)
"Global Anti-Realism", Philosophy and Phenomenological Research, 47, 641-647.

_____ (1992)
"The Metaphysics of Anti-Realism", Metaphilosophy, 23, 68-76.