REALISM AND REASON*

HILARY PUTNAM

In one way of conceiving it, realism is an empirical theory.[1] One of the facts that this theory explains is the fact that scientific theories tend to "converge" in the sense that earlier theories are, very often, limiting cases of later theories (which is why it is possible to regard theoretical terms as preserving their reference across most changes of theory). Another of the facts it explains is the more mundane fact that language-using contributes to getting our goals, achieving satisfaction, or what have you.

The realist explanation, in a nutshell, is not that language mirrors the world but that *speakers* mirror the world -- i.e., their environment -- in the sense of *constructing a symbolic representation of that environment.* In "Reference and Understanding"[2] I argued that a "correspondence" between words and sets of things (formally, a *satisfaction relation,* in the sense of Tarski) can be viewed as part of an *explanatory model* of the speakers' collective behavior.

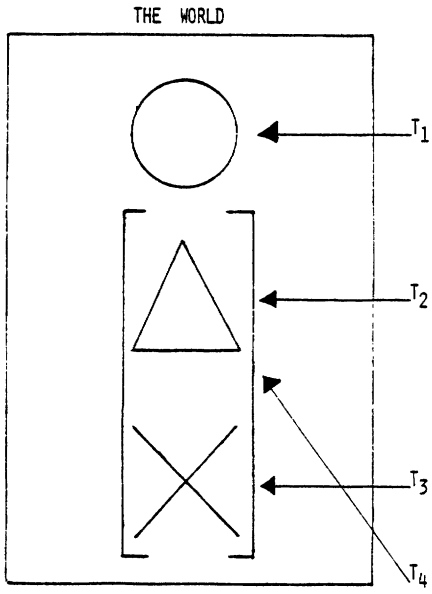
I'm not going to review this in this Address; but let me refer to realism in this sense -- acceptance of this sort of scientific picture of the relation of speakers to their environment, and of the role of language, -- as *internal* realism.

Metaphysical realism, on the other hand, is less an empirical theory than a model -- in the "colliding billiard balls" sense of 'model'. It is, or purports to be, a model of the relation of *any* correct theory to all or part of THE WORLD. I have come to the conclusion that this model is incoherent. This is what I want to share with you.

Let us set out the model in its basic form.

---

*Presidential Address delivered before the Seventy-third Annual Eastern Meeting of the American Philosophical Association in Boston, Massachusetts, December 29, 1976.



In its primitive form, there is a relation between each term in the language and a piece of THE WORLD (or a *kind* of piece, if the term is a general term).

This relation -- the relation of reference -- is given *by the truth -- conditional semantics* for the language, in the canonical versions of the theory -- i.e., *understanding* a term, say, T_1, consists in knowing what piece of THE WORLD it refers to (or in knowing a necessary and sufficient condition for it to refer to a piece of THE WORLD, in some versions). I shall not assume this account of understanding to be part of the picture in what follows, although it certainly was assumed by metaphysical realists in the past.

Minimally, however, there has to *be* a determinate relation of *reference* between terms in L and pieces (or sets of pieces) of THE WORLD, on the metaphysical realist model, whether *understanding* L is taken to consist in "knowing" that relation or not. What makes this picture different from *internal* realism (which employs a similar picture *within* a theory) is that (1) the picture is supposed to apply to *all* correct theories at once (so that it can only be stated with "Typical Ambiguity" -- i.e., it transcends complete formalization in any one theory); and (2) 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).

The most important consequence of metaphysical realism is that *truth* is supposed to be *radically non-epistemic* -- we might be "brains in a vat" and so the theory that is "ideal" from the point of view of operational utility, inner beauty and elegance, "plausibility", simplicity, "conservatism", etc., *might be false.* "Verified" (in any operational sense) does not imply "true", on the metaphysical realist picture, even in the ideal limit.

It is this feature that distinguishes metaphysical realism, as I am using the term, from the mere belief that there *is* an ideal theory (Peircean realism), or, more weakly, that an ideal theory is a regulative ideal presupposed by the notions "true" and "objective" as they have classically been understood. And it is this feature that I shall attack!

So let T_1 be an ideal theory, by our lights. Lifting restrictions on our actual all-too-finite powers, we can imagine T_1 to have every property *except objective truth* -- which is left open -- that we like. E.g., T_1 can be imagined complete, consistent, to predict correctly all observation sentences (as far as we can tell), to meet whatever "operational constraints" there are (if these are "fuzzy", let T_1 seem to *clearly* meet them), to be "beautiful", "simple", "plausible", etc. The supposition under consideration is that T_1 might be all this *and still be* (in reality) *false.*

I assume THE WORLD has (or can be broken into) infinitely many pieces. I also assume T_1 *says* there are infinitely many things (so in *this* respect T_1 *is* "objectively right" about THE WORLD). Now T_1 is *consistent* (by hypothesis) and has (only) infinite models. So by the completeness theorem (in its model theoretic form), T_1 has a model of every infinite cardinality. Pick a model M of the same cardinality as THE WORLD.[3] Map the individuals of M one-to-one into the pieces of THE WORLD, and use the mapping to define the relations of M directly in THE WORLD. The result is a satisfaction relation SAT – a "correspondence" between the terms of L and sets of pieces of THE WORLD -- such that the theory T_1 comes out *true* -- true of THE WORLD -- provided we just interpret 'true' as TRUE(SAT)[4]. So what becomes of the claim that even the *ideal* theory T_1 might *really* be false?

Well, it might be claimed that SAT is not the *intended* correspondence between L and THE WORLD. What does 'intended' come to here?

T_1 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_1 then, "there is a

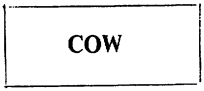
cow in front of me at such-and-such a time" will certainly *seem* to be true--it will be "exactly as if" there were a cow in front of me at that time. But SAT is a *true* interpretation of T_1. T_1 is TRUE(SAT). So "there is a cow in front of me at such-and-such a time" is "True" in this sense -- TRUE (SAT).

On the other hand, if "there is a cow in front of me at such-and-such a time" is *operationally* "false" (falsified) then "there is a cow in front of me at such-and-such a time" is FALSE(SAT). So, the interpretation of "reference" in L as SAT certainly meets all *operational* constraints on reference. But the interpretation of "reference" as SAT certainly meets all *theoretical* constraints on reference--it makes the *ideal* theory, T_1, come out *true*.

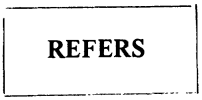
So what *further* constraints on reference are there that could single out some other interpretation as (uniquely) "intended", and SAT as an "unintended" interpretation (in the model-theoretic sense of "interpretation")? The supposition that even an "ideal" theory (from a pragmatic point of view) might *really* be false appears to collapse into *unintelligiblity*.

Notice that a "causal" theory of reference is not (would not be) of any help here: for how 'causes' can uniquely refer in as much of a puzzle as how 'cat' can, on the metaphysical realist picture.

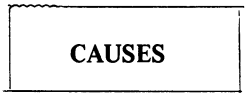
The problem, in a way, is traceable back to Ockham. Ockham introduced the idea that concepts are (mental) *particulars*. If concepts are particulars ("signs"), then any concept we may have of the *relation* between a sign and its object is *another sign*. But it is unintelligible, from my point of view, how the sort of relation the metaphysical realist envisages as holding between a sign and its object can be singled out either by holding up the sign itself, thus



--*or* by holding up yet another sign, thus



--or perhaps--



If concepts are not particulars, on the other hand, the obvious possibility is that (insofar as they are "in the head") they are *ways of using* signs. But a "use" theory, while intelligible (and, I believe, correct) as an account of what *understanding* the signs consists in, *does not single out a unique relation* between the terms of T_1 and the "real objects". If we don't think concepts are *either* particulars (signs) *nor* ways of using signs, then, I think we are going to be led back to direct (and mysterious) grasp of Forms.

Suppose we (and all other sentient beings) are and always were "brains in a vat". Then how does it come about that *our* word "vat" refers to *noumenal* vats and not to vats in the image.

If the foregoing is not to be just a new antinomy, then one has to show that there is at least one intelligible position for which it does *not* arise. And there is. It does not arise for the position Michael Dummett has been defending. Let me explain:

Dummett's idea[5] is to do the *theory of understanding* in terms of the notions of *verification* and *falsification.* This is what he calls "non-realist semantics".

What makes this different from the old phenomenalism is that there is no "basis" of hard facts (e.g., sense data) with respect to which one ultimately uses the truth conditional semantics, classical logic, and the *realist* notions of truth and falsity. The analogy is with Mathematical Intuitionism: the Intuitionist uses *his* notion of "truth" -- constructive provability -- *even when talking about constructive proof itself.* Understanding a sentence, in this semantics, is knowing what constitutes a proof (verification) of it. And this is true *even of the sentences that describe verifications.* Thus, I might take "I have a red sense datum" as a primitive sentence, or I might take "I see a cow", or, if I do the semantics from the point of view of the brain rather than the person, I might take "such and such neurons fired".

Whatever language I use, a primitive sentence -- say, "I see a cow" -- will be assertible if and only if *verified.* And we say it is verified *by saying the sentence itself,* "I see a cow". To use a term of Roderick Firth's, "I see a cow" is "self-warranting" in this kind of epistemology -- not in the sense of being *incorrigible,* not even necessarily in the sense of being fully determinate (i.e., obeying strong bivalence -- being determinately true or false). (Facts are "soft all the way down" on this picture, Dummett says.) The important

point is that the realist concepts of truth and falsity are not used in this semantics at all.

Now the puzzle about what singles out one correspondence as *the* relation of reference does not arise. The notion of "reference" is not used in the semantics. We can introduce "refers" into the language á la Tarski, but then (1) "Cow" refers to cows, will simply be a tautology -- and the *understanding* of (1) makes no reference to the metaphysical realist picture at all.

One important point. It is not good to do the non-realist semantics (I would rather call it *verificationist semantics* -- because it is not incompatible with *internal* realism), -- in terms of any level of "hard facts", even sense data. For if sense data are treated as "hard data" -- if the verificationist semantics is given in a meta-language *for* which itself we give the *truth-conditional* account of understanding -- then we can repeat the whole argument against the intelligibility of metaphysical realism (as an argument against the intelligibility of the *meta*-language) -- just think of the *past* sense-data (or the *future* ones) as the "external" part of THE WORLD. (This is a reconstruction of one aspect of Wittgenstein's private language argument.) This is why Dummett's move depends upon using the verificationist semantics all the way up (or down) -- in the meta-language, the meta-meta-language, etc.

The reason I got involved in this problem is this: in "Reference and Understanding" I argued that one could give a *model of a speaker* of the language in terms of the notion of "degree of confirmation" (which might better be called "degree of verification" when it has this understanding-theoretic role). And I contended that the realist notions of truth and reference come in not in explaining what goes on "in the heads" of speakers, but in *explaining the success* of language-using. Thus I urged that we accept a species of "verificationist" semantics. (Though not in the sense of verificationist theory of *meaning* -- for, as I have argued elsewhere[6], "meaning" is not just a function of what goes on "in our heads", but also of *reference,* and reference is determined by *social* practices and by actual physical paradigms, and not just by what goes on inside any individual speaker.) But, I claimed, one can still be a *realist,* even though one accepts this "verificationist" model. For the realist claim that there is a correspondence between words and things is not *incompatible* with a "verificationist" or "use" account of understanding. Such a correspondence, in my view, is part of an *explanatory theory* of the speakers' interaction with their environment.

The point is that Dummett and I *agree* that you can't treat understanding a sentence (in general) as knowing its truth conditions; because it then becomes unintelligible what *that* knowledge *in turn* consists in. We both *agree* that the theory of understanding has to be done in a verificationist way. (Although I don't think that theory of understanding is all of theory of *meaning*,

that is of no help *here* -- theory of meaning, on my view, presupposes theory of understanding *and reference* -- and reference is what the problem is all about!) But now it looks as if in conceding that *some* sort of verificationist semantics must be given as our account of understanding (or "linguistic competence", in Chomsky's sense), I have given Dummett all he needs to demolish metaphysical realism -- a picture I was wedded to!

So *what?* At this point, *I* think that a natural response would be the following: "So metaphysical realism collapses. But internal realism -- the empirical theory of "Reference and Understanding" -- doesn't collapse (I claim). Metaphysical realism was only a *picture* anyway. If the picture is, indeed, incoherent, then the moral is surely *not* that something is wrong with realism *per se*, but simply that realism *equals* internal realism. *Internal realism is all the realism we want or need.*

Indeed, I believe that this is true. But it isn't *all* the moral. Metaphysical realism collapsed *at a particular point.* (I am going to argue that it also collapses at other points.) And the point at which it collapsed tells us something. Metaphysical realism collapses just at the point at which it claims to be distinguishable from Peircean realism -- i.e., from the claim that there is an ideal theory (I don't mean that even *that* claim isn't problematical, but it is problematical in a different way). Since Peirce himself (and the verificationists) always *said* metaphysical realism collapses into incoherence at *just* that point, and realists like myself thought they were *wrong*, there is no avoiding the unpleasant admission that "they were right and we were wrong" *on* at least one substantive issue.

I now want to talk about other points at which the metaphysical realist picture is incoherent. Consider the following simple universe: let THE WORLD be a *straight line*, thus

(If you want, there can be one-dimensional people -- with apologies to Marcuse -- on the line. How you tell the boys from the girls, I don't know.)

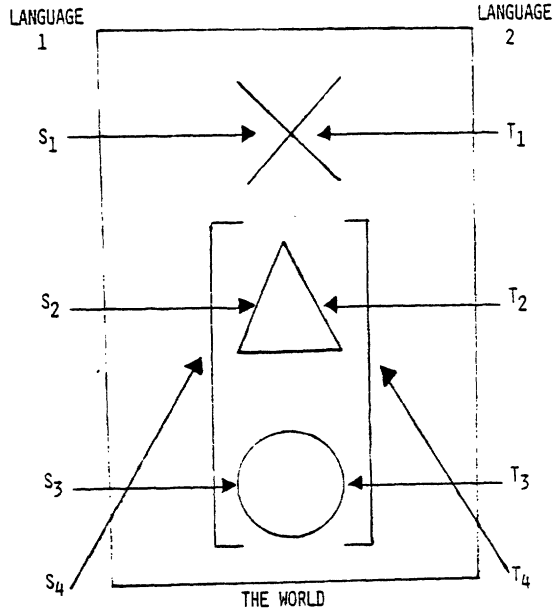
Consider the following two stories about THE WORLD:

Story 1. There are points -- i.e., the line has parts which are line *segments,* and also infinitely small parts called "points". The *same relation* -- "part of" -- holds between points and line segments which contain them, and

between line segments and bigger line segments (and between any piece of the line and the whole line).

Story 2. There are no points -- the line and its parts all have *extension.* "Of course," the teller of this story says, "I'm not saying Story 1 is *false.* You just have to understand that *points* are logical constructions out of line segments. Point talk is highly derived talk about convergent sets of line segments."

A "hard core" realist might claim that there is a "fact of the matter" as to which is true -- Story 1 or Story 2. But "sophisticated realists", as I have called them, concede that Story 1 and Story 2 are "equivalent descriptions". In effect, this concedes that line segments are a suitable set of "invariants" -- a description of THE WORLD which says what is going on in every line segment is a *complete* description. In the past, I argued that this is no problem for the realist -- it's just like the fact that the earth can be mapped by different "projections", I said (Mercator, Polar, etc.). The metaphysical realist picture now looks like this:



In particular, I believed, it can happen that what we picture as "incompatible" terms can be mapped onto the same Real Object -- though not, of course, within the same theory. Thus the Real Object that is labeled "point" in one theory might be labeled "set of convergent line segments" in another theory. And the *same* term might be mapped onto one Real Object in one theory and onto a different Real Object in another theory. It is a property of the world itself, I claimed – i.e., a property of THE WORLD itself -- that it "admits of these different mappings".

The problem – as Nelson Goodman has been emphasizing for many, many years -- is that this story 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 upon passing from one correct theory to another correct theory -- e.g., 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 concede that Story 1 and Story 2 are Equivalent Descriptions, then the property *being an object* (as opposed to a class or set of things) will be theory relative. Consider now a third story, Story 3: *There are only line segments with rational end points* (i.e., since there aren't "points", in this story, except as logical constructions, (1) every line segment has rational length; (2) the piece of the line[7] between any two line segments is a line segment, and so has rational length; (3) every line segment is divisible into n equal pieces, for every integer n; (4) there is at least one line segment; and (5) the union of two line segments is a line segment.) Irrational line segments are treated as logical constructions – sets of "points" are themselves Cauchy convergent sets of *rational* line segments.

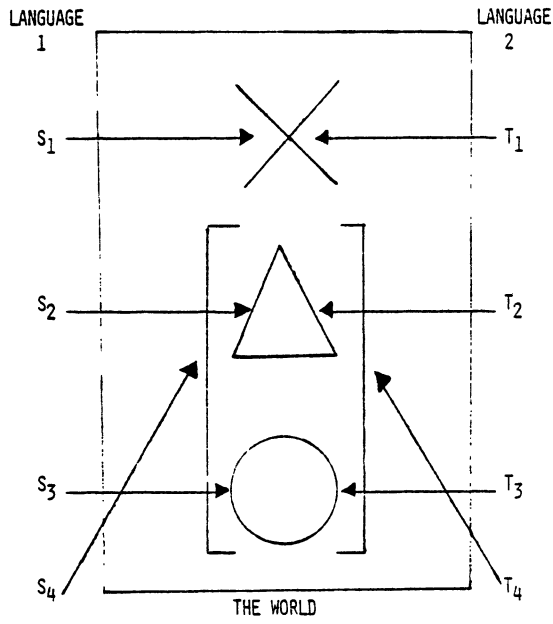
A "hard core" realist might again object, this time because this story makes an irrational line segment of a different logical type than a rational line segment. But the defender of this story can reply: "Isn't it common in mathematics that objects are identified with sets of other objects which are *pre-analytically* of the same logical type? Thus, negative integers and positive integers whole numbers and rationals, rationals and reals, reals and imaginaries are pre-analytically all 'numbers', but in *formalizing* mathematics we are used to treating negative numbers (or more generally, 'signed numbers') as *ordered pairs* of 'natural numbers', *rational* numbers as *ordered pairs* of 'signed numbers', irrationals as *sets* of rationals, etc. So what is wrong with treating irrational line segments as sets of sets of rational line segments? After all, the rational line segments are a basis for the topology; if you know what is going on in every rational line segment, you have a complete description of all events, etc."

If we accept *Story 3* as yet another Equivalent Description of THE WORLD, however, then even the *cardinality* of the world becomes theory relative! For there are only denumerably many objects in Story 3, and non-denumerably many in Stories 1 and 2! (We might try to avoid this by treating *sets* as "objects", too -- but, as I've shown elsewhere, "set" talk can be translated away into *possibility* talk.)

All this isn't an artifact of my simple example: actual physical theory is rife with similar examples. One can construe space-time points as objects, for example, or as properties. One can construe fields as objects, or do everything with particles acting at a distance (using retarded potentials). 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 Kantian "noumenal" world, a *mere* "thing in itself", If one cannot say *how* THE WORLD is theory-independently, then talk of all these theories as descriptions of THE WORLD is empty.

Another point at which the metaphysical realist picture runs into trouble.

This has to do with what Quine calls "ontological relativity". Suppose we confine attention, for the moment, to *complete* theories. If T is a complete theory, we can define an equivalence relation on its terms -- *provable coextensiveness* -- with the property that if two terms belong to different equivalence classes, then in *no* model of the theory do they refer to the same referent, whereas if they belong to the same equivalence class, then they have the same referent in *every* model of the theory. So, for our purposes, we may count terms as the same if they lie in the same equivalence class – i.e., if they are "coextensive taking the theory at face value". With this preliminary identification made, we notice that if our picture is correct -- I repeat the picture



-- then there is a *unique* reference-preserving "translation" connecting the Languages.

But it is notorious that there are often *inequivalent* relative interpretations of one theory in another. Story 1 can be interpreted in Story 2 (in the case of our example) in *many* different ways. "Points" can be sets of line segments whose lengths are negative powers of 2, for example, or sets of line segments whose lengths are negative powers of 3.

If the picture as I drew it were correct, there would have to be a "fact of the matter" as to *which* translation *really* preserves reference in every such case!

Just as we complicated the picture by allowing the same term to be mapped onto different Real Objects when it occurs in different theories to meet the previous objection, so we could complicate the picture *again* to meet the second objection: we could say that the language has *more than one* correct way of being mapped onto THE WORLD (it must, since it has more than one way of being correctly mapped onto a language which is *itself* correctly mapped onto the world). But now *all* grasp of the picture seems to vanish: if what is a *unique* set of things *within a correct theory* may not be a unique set of things "in reality", then the very heart of the picture is torn out.

Why all this doesn't refute internal realism.

Suppose we try to stump the *internal* realist with the question, "How do you know that 'cow' refers to cows?" "After all", we point out, "there are other interpretations of your whole language -- non-denumerably many interpretations (in the sense of satisfaction-relations), which would render true an *ideal* theory (in your language). Indeed, suppose God gave us the *set of all true sentences* in your language (pretend we have infinite memories, for this purpose). Call this set the *perfect* theory. Then there would *still* be infinitely many admissable interpretations of the *perfect* theory -- interpretations which, as we saw, satisfied *all* the operational and theoretical constraints. Even the *sentence* "'Cow' refers to cows" is true in all of these interpretations. So how do you know that it is true in the sense of being true in a *unique* "intended" interpretation? How do you know that 'cow' refers to cows in the sense of referring to *one* determinate set of things, as opposed to referring to a determinate set of things *in each admissible interpretation?"* (This is, of course, just arguing against the internal realist exactly as we argued against the metaphysical realist.)

The internal realist should reply that "'Cow' refers to cows" follows immediately from the definition of 'refers'. In fact, "'cow' refers to cows" would be true even if internal realism were false: although we can revise "'Cow' refers to cows" by scrapping the theory itself (or at least scrapping or challenging the notion of a *cow*) -- and this is how the fact that "'Cow' refers to cows" is not *absolutely* unrevisable manifests itself -- *relative to the theory,* "'Cow' refers to cows" is a logical truth.

The critic will now reply that his question hasn't been answered. "'Cow' refers to cows" is indeed analytic relative to the theory -- but his question challenged *the way the theory is understood.* "'Cow' refers to cows" is true in all admissible interpretations of the theory -- but that isn't at issue.

The internal realist should now reply that (1) "the way the theory is understood" can't be discussed *within* the theory; and (2) the question whether the theory has a *unique* intended interpretation has *no* absolute sense. Viewed from within Story 1 (or a meta-language which contains the object language of Story 1), "point" has a "unique intended interpretation". Viewed from within Story 2 (or a meta-language which contains the object language of Story 2), the term "point" *as used in Story 1* has a plurality of admissible interpretations. The critic's "how do you know?" question assumes a theory-independent fact of the matter as to what a term in a given theory corresponds to -- i.e., assumes the picture of metaphysical realism; and this is a picture the *internal* realist need not (and better not) accept.

The critic now replies as follows: "reference" (strictly speaking, 'satisfies') is defined so that (1) 'Cow' refers to cows. -- just says (in the case of where 'cow' is a primitive expression of L) that the ordered pair <'cow', {cows}> belongs to a certain *list* of ordered pairs. If anything, this *presupposes* that 'cow' refers (in some *other* sense of 'refers'); it doesn't *explicate* it.[8]

Answer: the *use* of 'cow' does presuppose that 'cow' is *understood. And if my account of* understanding was a *truth conditional* (or *reference conditional)* account, then the objection would be good. But I gave a verificationist account of understanding (in terms of degree of confirmation); thus *my use of the term 'cow' in the language has already been explained, and I am free to use it--even to use it in explaining what 'cow' refers* to.

What I am saying is that, in a certain "contextual" sense, it is an *a priori* truth that 'cow' refers to a determinate class of things (or a more-or-less determinate class of things -- I neglect ordinary vagueness). Adopting "cow talk" is adopting a "version", in Nelson Goodman's phrase, from within which it is *a priori* that the word 'cow' refers (and, indeed, that it refers to cows).

One of the puzzling things about the metaphysical realist picture is that it makes it unintelligible how there can *be a priori* truths, even contextual ones, even as a (possibly unreachable) limit. An a *priori* truth would have to be the product of a kind of direct "intuition" of the things themselves. Even verbal truth is hard to understand. Consider "all bachelors are unmarried". It can be "verbal" that this is in some sense "short for" "all unmarried men are unmarried". And this, in turn, is an instance of "All AB are A". But why is *this* true?

Suppose there *were* unrevisability -- absolute unrevisability. And suppose we held "All AB are A" (and even "All unmarried men are unmarried") absolutely immune from revision. Why would this make it *true*?

Suppose, unimaginably, there are some AB that are not A. (After all, there are lots of things in modern science we can't imagine) Then, on the metaphysical realist picture, our refusal to give up assenting to "All AB are A" doesn't make *it* true -- it just makes *us* stubborn.9

Once we abandon the metaphysical realist picture, the situation becomes quite different. Suppose we include a sentence S in the ideal theory T_1 just because it is a feature we *want* the ideal theory to have that it contain S. (Suppose we even hold S "immune" from revision, as a behavioristic fact about us.) Assuming S doesn't make T_1 inconsistent, T_1 *still* has a model. And since the

model isn't fixed *independently* of the theory, T_1 will be *true* -- true in *the* model (from the point of view of *meta*-T_1; true in all *admissible* models, from the point of view of a theory in which the terms of T_1 do not determinately refer to begin with. So S will be true! "S" is "analytic" -- but it is an "analyticity" that resembles Kant's account of the *synthetic a priori* more than it resembles his account of the analytic. For the "analytic" sentence is, so to speak, part of "the form of the representation" and not "the content of the representation". It can't be false of the world (as opposed to THE WORLD), because the world is not describable independently of our description.

Even if T_1 were *inconsistent,* if we were consistently inconsistent (assigned "truth" and "falsity" to sentences in a stable way), this would not block this argument: for *stable* inconsistency can be viewed as *reinterpretation of the logical connectives.* When we give up the metaphysical realist picture we see for the first time how a truth can be "about the world" ("All AB are A" *is* "about the world" -- it is about all *classes* A,B) and "without content".

In the foregoing, I used the idea of an absolutely "unrevisable" truth as an idealization. Of course, I agree with Quine that this is an unattainable "limit". Any statement can be "revised". But what is often overlooked, although Quine stresses it again and again, is that the revisability of the laws of Euclid's geometry, or the laws of classical logic, does not make them mere "empirical" statements. This is why I have called them *contextually a priori.*[10] Quine put the point very well when he said that "the lore of our fathers" is black with fact and white with convention, and added that there are no *completely* white threads and no quite black ones. One might describe this as a soft (and de-mythologized) Kantianism. A trouble with the meta-physical realist picture is that one cannot see how there can be white at all -- even greyish white.

Let me close with a last philosophical metaphor. Kant's image was of knowledge as a "representation" -- a kind of play. The author is me. But the author also appears as a character in the play (like a Pirandello play). The author in the play is not the "real" author -- it is the "empirical me". The "real" author is the "transcendental me".

I would modify Kant's image in two ways. The authors (in the plural -- my image of knowledge is social) don't write just *one* story: they write many versions. And the authors *in* the stories are the *real* authors. This would be "crazy" if these stories were *fictions.* A fictitious character can't also be a real author. But these are true stories.

Hilary Putnam
Harvard University

Footnotes

1. This is spelled out in my "What is 'Realism' ", in *Proceedings of the Aristotelian Society*, 1976, pp. 177-194.

2. Delivered in Jerusalem, May 1976, and forthcoming as Part II of my *Meaning and the Moral Sciences* (to be published by Routledge and Kegan Paul).

3. If THE WORLD is finite, let the theory be compatible with there being only N individuals (where N is the cardinality of THE WORLD), and pick a model with N individuals instead of using the stronger model-theoretic theorem appealed to in the text.

4. Here, if SAT is a relation of the same logical type as "satisfies", TRUE (SAT) is supposed to be defined in terms of SAT exactly as "true" is defined in terms of "satisfies" (by Tarski). Thus "TRUE(SAT)" is the truth-property "determined" by the relation SAT.

5. This is most completely spelled out in his (unpublished) William James Lectures. A partial account appears in his contribution to the conference on "Language, Intentionality, and Translation Talk" reprinted in *Synthese*, Vol. 27, Nos. 3/4, July/August 1974.

6. Cf. my "The Meaning of 'Meaning' " in my *Mind, Language, and Reality (Philosophical Papers*, Vol. 2), Cambridge University Press, 1976.

7. The mathematical reader will note that in Story 3 there is no distinction between *open* and *closed* line segments -- because there are no such things as *points!*

8. This objection to Tarskian definitions of reference is due to Hartry Field (cf. his "Tarski's Theory of Truth", *The Journal of Philosophy*, Vol. 69, No. 13 (1972), pp. 347-375) and is discussed in my John Locke Lectures (to appear in the book cited in note 2).

9. The reader may be tempted to reply that even a metaphysical realist is entitled to the notion of a *verbal convention*. And why can't it be a verbal *convention* that no state of affairs is to be referred to as the *conjuction* of the states of affairs described by sentences *p*, *q* unless it entails both *p* and *q* separately? This would mean that "*p.q* entails *p*" is true by convention"; and similarly, "All AB are A" could be "true by convention".

The difficulty is that such a "convention" only makes it analytic that the conjunction *p.q* entails *p if the state of affairs in question exists.* But how can the *existence* of a state of affairs with the properties of entailing *p* and entailing *q* and being entailed by every state of affairs which entails both *p* and *q* be *itself* a matter of *convention,* on the meta-physical realist picture?

To establish that this is not trifling, let me remark that there are logics (studies by David Finkelstein in connection with certain "far out" physical theories -- *not* standard quantum theory) in which (1) there are propositions *incompatible* with any given proposition; but (2) there is no such thing as *the* negation of a given proposition -- i.e., no logically weakest proposition incompatible with a given proposition. (These logics correspond algebraically to lattices which are not orthocomplemented.) If "the logic of the world" is one of *these* logics (as Finkelstein believes), then *the existence of a complement* to a given state of affairs is *false as a matter of fact* -- and *no* linguistic convention *could* render it true!

It seems to me that a consistent metaphysical realist must *either* view logic as empirical, not just in the sense of thinking that logic is revisable (which I believe), but in the sense of having *no* conventional component at all (so that even our confidence that statements aren't *both* true and false becomes ultimately just *inductive* confidence), *or* he must believe that logic is *a priori* in a sense of *a priori* which is not explainable by the notion of convention at all.

10. In "It Ain't Necessarily So", reprinted in my *Mind, Language and Reality.* cited in note 6.