What Is "Realism"?

Author(s): Hilary Putnam

Source: *Proceedings of the Aristotelian Society*, 1975 - 1976, New Series, Vol. 76 (1975 - 1976), pp. 177-194

Published by: Oxford University Press on behalf of The Aristotelian Society

Stable URL: https://www.jstor.org/stable/4544887

X*—WHAT IS "REALISM"?

by Hilary Putnam

While it is undoubtedly a good thing that "ism" words have gone out of fashion in philosophy, *some* "ism" words seem remarkably resistant to being banned. One such word is "realism". More and more philosophers are talking about realism these days; but very little is said about what realism *is*. This paper will not answer that very large question; but I hope to contribute a portion of an answer.

Whatever else realists say, they typically say that they believe in a Correspondence Theory of Truth.

When they argue *for* their position, realists typically argue *against* some version of Idealism—in our time, this would be Positivism or Operationalism. (This is not in itself surprising—all philosophers attempt to shift the burden of proof to their opponents. And if one's opponent has the burden of proof, to dispose of his arguments seems a sufficient defence of one's own position.) And the typical realist argument against Idealism is that it makes the success of science a *miracle*. Berkeley needed God just to account for the success of beliefs about tables and chairs (and trees in the Quad); but the appeal to God has gone out of fashion in philosophy, and, in any case, Berkeley's use of God is very odd from the point of view of most theists. And the modern positivist has to leave it without explanation (the realist charges) that "electron calculi" and "space-time calculi" and "DNA calculi" correctly predict observable phenomena if, in reality, there are no electrons, no curved space-time, and no DNA molecules. If there are such things, then a natural explanation of the success of these theories is that they are *partially true accounts* of how they behave. And a natural account of the way in which scientific theories succeed each other—say, the way in which Einstein's Relativity succeeded Newton's Universal Gravitation—is that a partially correct/partially

---

* Meeting of the Aristotelian Society at 5/7, Tavistock Place, London, W.C.1, on Monday, 23rd February 1976, at 7.30 p.m.

incorrect account of a theoretical object—say, the gravitational field, or the metric structure of space-time, or both—is replaced by a *better* account of the same object or objects. But if these objects don't really exist at all, then it is a *miracle* that a theory which speaks of gravitational action at a distance successfully predicts phenomena; it is a *miracle* that a theory which speaks of curved space-time successfully predicts phenomena; and the fact that the laws of the former theory are derivable "in the limit" from the laws of the latter theory has no explained methodological significance.

I am not claiming that the positivist (or whatever) has no *rejoinder* to make to this sort of argument. He has a number: reductionist theories of the *meaning* of theoretical terms, theories of explanation, *etc.* Right now, my interest is rather in the following fact: the realist's argument turns on the success of science, or, in an earlier day, the success of common sense material object theory. But what does the success of science have to do with the Correspondence Theory of Truth?—or *any* theory of truth, for that matter?

That science succeeds in making many true predictions, devising better ways of controlling nature, *etc.*, is an undoubted empirical fact. If realism is an *explanation* of this fact, realism must itself be an over-arching scientific *hypothesis.* And realists have often embraced this idea, and proclaimed that realism *is* an empirical hypothesis. But then it is left obscure what realism has to do with theory of *truth.* In the present paper, I shall try to bring out what the connexion between these two concerns of the realist is—what the connexion is between explaining the success of knowledge and the theory of truth.

1. *The "convergence" of scientific knowledge*

What I am calling "realism" is often called "scientific realism" by its proponents. If I avoid that term here, it is because "scientific realist", as a label, carries a certain ideological *tone*—a tone faintly reminiscent of 19th century materialism, or, to be blunt about it, village atheism. Indeed, if a "scientific realist" is one who believes, *inter alia*, that *all* knowledge worthy of the name is part of "science", then I am not a "scientific realist". But scientific knowledge is certainly an impressive part of our knowledge, and its nature and

significance have concerned all the great philosophers interested in epistemology at all. So it is not surprising that both realists and idealists should claim to be "philosophers of science", in *two* senses of "of". And if I focus on scientific knowledge in what follows, it is because the discussion has focused on it, and not out of a personal commitment to scientism.

To begin with, let me say that I think there *is* something to the idea of *convergence* in scientific knowledge. *What* there is is best explained, in my opinion, in an unpublished essay by Richard Boyd.[1] Boyd points out that all that follows from standard (positivist) philosophy of science is that later theories in a science, if they are to be *better* than the theories they succeed, must imply many of the *observation sentences* of the earlier theories (especially the *true* observation sentences implied by the earlier theories). It does not follow that the later theories must imply *the approximate truth of the theoretical laws of the earlier theories in certain circumstances*—which they typically do. In fact, preserving the *mechanisms* of the earlier theory as often as possible, which is what scientists try to do (or to show that they are "limiting cases" of new mechanisms)—is often the *hardest* way to get a theory which keeps the old observational predictions, where they were correct, and simultaneously incorporates the new observational data. That scientists try to do this—*e.g.*, *preserve* conservation of energy, if they can, rather than postulate violations—is a fact, and that this strategy has led to important discoveries (from the discovery of Neptune to the discovery of the positron) is also a fact.

Boyd tries to spell out Realism as an empirical hypothesis by means of two principles:

(1) Terms in a mature science typically *refer*.

(2) The laws of a theory belonging to a mature science are typically approximately true.

What he attempts to show in his essay is that scientists act as they do because they *believe* (1) and (2) and that their strategy works because (1) and (2) are *true*.

One of the most interesting things about this argument is that, if it is correct, the notions of "truth" and "reference" have a causal-explanatory rôle in epistemology. (1) and (2)

are premisses in an *explanation* of the behaviour of scientists and the success of science—and they essentially contain concepts from referential semantics. Replacing "true", in premiss (2) (of course, Boyd's argument needs many more premisses than *just* (1) and (2)) by some Operationalist "substitute"—*e.g.*, "is simple and leads to true predictions"—will not preserve the explanation.

Let us pause to see why. Suppose T_1 is the received theory in some central branch of physics (*physics* surely counts as a "mature" science if any science does), and I am a scientist trying to find a theory T_2 to replace T_1. (Perhaps I even know of areas in which T_1 leads to false predictions.) If I believe principles (1) and (2), then I know that the laws of T_1 are (probably) approximately true. So T_2 must have a certain property—the property that the laws of T_1 are "approximately true" *when we judge from the standpoint of* T_2—or T_2 will (probably) have no chance of being true. Since I want theories that are not *just* "approximately true", but theories that have a chance of being *true*, I will only consider theories, as candidates for being T_2, which have this property—theories which contain the laws of T_1 as a limiting case. But this is just the feature of the scientific method we discussed. (Boyd also discusses a great many other features of the scientific method—not just this aspect of "convergence"; but I do not need to go into these other features here.) In fine, my knowledge of the truth of (1) and (2) enables me to restrict the class of candidate-theories I have to consider, and thereby increases my chance of success.

Now, if all I know is that T_1 leads to (mainly) true predictions in some observational vocabulary (a notion I have criticized elsewhere[2]), then all I know about T_2 is that it should imply most of the "observation sentences" implied by T_1. But it does *not* follow that it must imply the truth of the *laws* of T_1 in some limit. There are many other ways of constructing T_2 so that it will imply the truth of most of the observation sentences of T_1; and making T_2 imply the "approximate truth" of the *laws* of T_1 is often the *hardest* way. Nor is there any reason why T_2 should have the property that we can assign *referents* to the terms of T_1 from the standpoint of T_2. Yet it is a fact that we can assign a referent to "gravitational field" in Newtonian theory *from the stand-*

point of Relativity theory (though not to "ether" or "phlogiston"); a referent to Mendel's "gene" from the standpoint of present-day molecular biology; and a referent to Dalton's "atom" from the standpoint of quantum mechanics. These retrospective reference assignments depend on a principle that has been called the "Principle of Charity" or the "Principle of Benefit of the Doubt";[3] but not on *unreasonable* "charity". Surely the "gene" discussed in molecular biology is the gene (or rather "factor") Mendel *intended* to talk about; it is certainly what he should have intended to talk about!—Again, if one believes that the terms of T_1 do have referents (and one's semantic theory incorporates the Principle of Benefit of the Doubt), then it will be a *constraint* on T_2, it will narrow the class of candidate-theories, that T_2 must have this property, the property that *from its standpoint* one can assign referents to the terms of T_1. And again, if I do not use the notions of truth and reference in philosophy of science, if all I use are "global" properties of the order of "simplicity" and "leads to true predictions", then I will have no analogue of this constraint, I will not be able to narrow the class of candidate-theories in this way.

2. *What if there were no "convergence" in scientific knowledge?*

Let me now approach these problems from the other end, from the problem of "truth". How would our notions of *truth* and *reference* be affected if we decide *there is no* convergence in knowledge?

This is already the situation according to someone like Kuhn, who is sceptical about convergence and who writes (at least in *The Structure of Scientific Revolutions*) as if the same term cannot have the same referent in different paradigms (theories belonging to or generating different paradigms correspond to different "worlds", he says), and even more so from Feyerabend's standpoint.

Let us suppose they are right, and that "electron" in Bohr's theory (the Bohr-Rutherford theory of the early 1900s) does not refer to what we *now* call electrons. Then it doesn't refer to *anything* we recognize in present theory, and, moreover, it doesn't refer to anything *from the standpoint of* present theory (speaking from that standpoint, the only thing Bohr

could have been referring to were electrons, and if he wasn't referring to electrons he wasn't referring to anything). So if we use present theory to answer the question "was Bohr referring when he used the term 'electron'?", the answer has to be "no", according to Kuhn and Feyerabend. And what other theory can *we* use but our own present theory? (Kant's predicament, one might call this, although Quine is very fond of it too.) Kuhn talks as if each theory does refer—namely, to *its own* "world" of entities—but that isn't true according to *any* (scientific) theory.

Feyerabend arrives at his position by the following reasoning (which Kuhn does not at all agree with; any similarity in their views on cross-theoretical reference does not come from a shared analysis of science): the introducer of a scientific term, or the experts who use it, accept certain laws as virtually necessary truths about the putative referent. Feyerabend treats these laws, or the theoretical description of the referent based on these laws, as, in effect, a *definition* of the referent (in effect, an *analytic* definition). So if we ever decide that nothing fits *that exact description*, then we must say that there was "no such thing". If nothing fits the exact Bohr-Rutherford description of an electron, then "electron" *in the sense in which Bohr-Rutherford used it* does not refer. Moreover, if the theoretical description of an electron is different in two theories, then the term "electron" has a different *sense* (since it is synonymous with different descriptions—Feyerabend does not say this explicitly, but if this isn't his argument he doesn't have any) in the two theories. In general, Feyerabend concludes, such a term can have neither a shared referent nor a shared sense in different theories (the "incommensurability of theories").

This line of reasoning can be blocked by arguing (as I have in various places, and as Saul Kripke has) that scientific terms are not synonymous with descriptions. Moreover, it is an essential principle of semantic methodology that when speakers specify a referent for a term they use by a *description* and, because of mistaken factual beliefs that these speakers have, that description fails to refer, we should assume that they would accept reasonable reformulations of their description (in cases where it is clear, given our knowledge, how their description should be reformulated so as to refer, and there

is no ambiguity about how to do it in the practical context). (This is, roughly, the Principle of Benefit of the Doubt alluded to above.)

To give an example: there is nothing in the world which *exactly* fits the Bohr-Rutherford description of an electron. But there are particles which *approximately* fit Bohr's description: they have the right charge, the right mass, and they are responsible for key effects which Bohr-Rutherford explained in terms of "electrons"; for example, electric current in a wire is flow of these particles. The Principle of Benefit of the Doubt dictates that we treat Bohr as referring to these particles.

Incidentally, if Bohr had not been according the Benefit of the Doubt to his earlier (Bohr-Rutherford period) self, he would not have *continued to use* the term "electron" (without even a gloss!) when he participated in the invention of (1930s) quantum mechanics.

Coming back to Kuhn, however: we can answer Kuhn by saying there *are* entities—in fact, just the entities we now call "electrons"—which behave like Bohr's "electrons" in many ways (one to each hydrogen atom; negative unit charge; appropriate mass; *etc*). And (this is, of course, just answering Kuhn exactly as we answered Feyerabend) the Principle of Benefit of the Doubt dictates that we should, in these circumstances, take Bohr to have been referring to what we call "electrons". We should just say we have a different theory of the *same* entities Bohr called "electrons" back then; his term did refer.

But we can only take this line because present theory does assert the existence of entities which fill many of the *rôles* Bohr's "electrons" were supposed to fill, even if these entities have other, very strange, properties, such as the Complementarity of Position and Momentum, that Bohr-Rutherford "electrons" were not supposed to have. But what if we accept a theory from the standpoint of which electrons are like *phlogiston*?

Then we will have to say electrons don't really exist. What if this keeps happening? What if *all* the theoretical entities postulated by one generation (molecules, genes, *etc.*, as well as electrons) invariably "don't exist" from the standpoint of later science?—this is, of course, a form of the old sceptical

"argument from error"—how do you know you aren't in error *now*? But it is the form in which the argument from error is a *serious* worry for many people today, and not just a "philosophical doubt".

One reason this is a serious worry is that eventually the following meta-induction becomes overwhelmingly compelling: *just as no term used in the science of more than 50* (or whatever) *years ago referred, so it will turn out that no term used now* (except maybe observation terms, if there are such) *refers.*

It must obviously be a desideratum for the Theory of Reference that this meta-induction be blocked; that is one justification for the Principle of Benefit of the Doubt. But Benefit of the Doubt can be *unreasonable*; we don't carry it so far as to say that *phlogiston* referred. If there is no convergence, if later scientific theories cease having earlier theories as "limiting cases", if Boyd's principles (1) and (2) are clearly false from the point of view of future science, then Benefit of the Doubt will always turn out to be unreasonable—there will not be a reasonable *modification* of the theoretical descriptions of various entities given by earlier theories which makes those descriptions refer to entities with somewhat the same rôles which do exist from the standpoint of the later theory. Reference will collapse.

But what happens to the notion of *truth* in theoretical science if none of the descriptive terms refer? Perhaps all theoretical sentences are "false"; or some convention of narrowest scope or widest scope, *etc.*, for assigning truth-values when predicates don't refer takes over. In any case, the notion of "truth-value" becomes uninteresting for sentences containing theoretical terms. So truth will collapse too.

Now, dear reader, I want to argue that the foregoing *isn't* quite what would happen. But this will turn on rather subtle logical considerations.

3. *Mathematical Intuitionism—an application to empirical knowledge*

On the assumption that the reader has not studied Mathematical Intuitionism (the school of mathematical philosophy developed by Brouwer, Heyting, *etc.*), let me mention a few facts that I will use in what follows.

A key idea of the Intuitionists is to use the logical connectives in a "non-classical" sense. (Of course, Intuitionists do this because they regard the "classical" sense as inapplicable to reasoning about infinite or potentially infinite domains.) They explain this sense—that is, they explain *their* meanings for the logical connectives—in terms of constructive *provability* rather than (classical) *truth.*

Thus:

(1) Asserting p is asserting p *is provable.*

("$p \cdot \ulcorner p$ is not provable$\urcorner$" is a *contradiction* for the Intuitionists.)

(2) "$\neg p$" ($\neg$ is the intuitionist symbol for negation) means *it is provable that a proof of* p *would imply the provability of* $1 = 0$ (or any other patent absurdity). In other words, $\neg p$ asserts the *absurdity of* p*'s provability* (and not the classical "falsity" of p).

(3) "$p \cdot q$ means p *is provable and* q *is provable.*

(4) "$p \vee q$" *means there is a proof of* p *or a proof of* q *and one can tell which.*

(5) "$p \supset q$" means *there is a method which applied to any proof of* p *yields a proof of* q *(and a proof that the method does this).*

These meanings are clearly different from the classical ones. For example, $p \vee \neg p$ (which asserts the decidability of every proposition) is not a theorem of Intuitionist propositional calculus.

Now, let us *reinterpret* the *classical* connectives as follows:

(1) $\sim$ is identical with $\neg$.

(2) $\cdot$ (classical) is identified with $\cdot$ (Intuitionist).

(3) $p \vee q$ (classical) is identified with $\neg(\neg p \cdot \neg q)$.

(4) $p \supset q$ (classical) is identified with $\neg(\neg\neg p \cdot \neg q)$.

Then, with this interpretation, the theorems of *classical* propositional calculus become theorems of Intuitionist propositional calculus! In other words, this is a *translation* of classical propositional calculus *into* Intuitionist propositional

calculus—not, of course, in the sense of giving the classical *meanings* of the connectives in terms of Intuitionist notions, but in the sense of giving the classical theorems. (It is not the only such "translation", by the way.) The meanings are still not classical, if the classical connectives are reinterpreted in this way, because these meanings are explained in terms of *provability* and not *truth and falsity.*

To illustrate: classically $p \vee \sim p$ asserts that every proposition is true or false. Under the above "conjunction-negation translation" into Intuitionist logic, $p \vee \sim p$ asserts $\neg(\neg p \cdot \neg\neg p)$, which says that it is absurd that the negation of a proposition and its double negation are both absurd—NOTHING ABOUT BEING TRUE OR FALSE!

One can extend all this to the quantifiers—I omit details.

> *One thing this shows is that, contrary to what a number of philosophers—including, surprisingly, Quine —have asserted, such inference rules as* $p \cdot q / \therefore p$; $p \cdot q / \therefore q$; $p / p \vee q$; $q / \therefore p \vee q$; $\sim p, \sim q / \therefore \sim(p \vee q)$ *do not fix the "meanings" of the logical connectives. Someone could accept all of these rules (and all classical tautologies, as well) and still be using the logical connectives in the non-classical sense just described—a sense which is not truth-functional.*

Suppose, now, we apply *this* interpretation of the logical connectives (the interpretation given by the "conjunction-negation translation" above) to empirical science (this idea was suggested to me by reading Dummett on Truth, although he should not be held responsible for it) in the following way: replace *constructive provability* (in the sense of Intuitionist mathematics) by *provability from* (some suitable consistent reconstruction of) *the postulates of the empirical science accepted at the time* (or, if one wishes to be a realist about "observation statements", those together with the set of true observation statements. If the empirical science accepted at the time is itself *inconsistent* with the set of true observation statements—because it implies a false prediction—then some appropriate subset would have to be specified, but I shall not consider here how this might be done). If B_1 is the empirical science accepted at *one* time and B_2 is the empirical

science accepted at a *different* time, then, according to this "quasi-Intuitionist" interpretation, the very *logical connectives* would refer to "provability in B_1" when used in B_1 and to "provability in B_2" when used in B_2. The *logical connectives* would change meaning in a systematic way as empirical knowledge changed.

> *A technical complication is that "provability in* B_1*" cannot be understood as formal (syntactic) provability, if we are to satisfy the axioms of Heyting's (Intuitionist) propositional calculus when we reinterpret the Intuitionist propositional calculus into which we are doing our conjunction-negation translation. Rather, we must take this notion in the sense of "informal" provability. But this same problem—the need for such an "informal" notion (satisfying the axioms of* S4*, with* $\Box$ *taken as denoting informal provability) arises in all Intuitionist mathematics.*

4. *Truth*

Suppose we formalize empirical science or some part of empirical science—that is, we formulate it in a formalized language L, with suitable logical rules and axioms, and with empirical postulates appropriate to the body of theory we are formalizing. Following standard present day logical practice, the predicate "true" (as applied to sentences of L) would not itself be a predicate of L, but would belong to a stronger "meta-language", *ML*. (Saul Kripke is currently exploring a method of avoiding this separation of object language and meta-language, but this would not affect the present discussion.) This predicate might be defined (using the logical resources of *ML* but no descriptive vocabulary except that of L) by methods due to Tarski; or it might be taken as a primitive (undefined) notion of *ML*. In either case, we would wish all sentences of the famous form:

(T) "Snow is white" is true if and only if snow is white

—all sentences asserting the equivalence of a sentence of L (pretend "snow is white" is a sentence of L) and the sentence of *ML* which says of that sentence that it is true—to be

theorems of *ML*. (Tarski called this "Criterium *W*" in his *Wahrheitsbegriff*—and this somehow got translated into English as "*Convention T*". I shall refer to the requirement that all sentences of the form (*T*) be theorems of *ML* as *Criterion T*.)

What happens to "true" if we reinterpret the logical connectives in the "quasi-Intuitionist" manner just described? *It is possible to define it exactly à la Tarski.* Only "truth" becomes *provability* (or, to be more precise, the double negation of provability. I shall ignore this last subtlety.) In short: the *formal* property of Truth: the Criterion of Adequacy (Criterion *T*)—only *fixes* the extension of "true" *if the logical connectives are classical.*

> *This means that we can extend the remark we made in section 3, (the first indented remark): even if the "natives" we are studying accept the Criterion* T *in addition to accepting all classical tautologies, it doesn't follow just from that that their "true" is the classical "true".*

"Truth" (defined in the standard recursive way, following Tarski) becomes *provability* if the logical connectives are suitably reinterpreted. What does "reference" become?

On the Tarski definition of *truth* and *reference*,

(a) "Electron" refers.

—is equivalent to

(b) There are electrons.

But if "there are" is interpreted Intuitionistically (b) asserts only

(c) There is a description *D* such that ⌜*D* is an electron⌝ is provable in B_1.

—and *this* could be true (for suitable B_1) even if there are no electrons! In short, the effect of reinterpreting the logical connectives Intuitionalistically is that "existence" becomes *intra-theoretic*. Actually, the effect is even more complicated than (c) if, in addition to understanding the connectives "quasi-intuitionistically" (*i.e.*, in the Intuitionist manner, but with "provability" relativized to B_1), we use the conjunction-

negation translation to interpret the "classical" connectives, as suggested here. But this complication does not change the point just made: if the quantifiers, like the other logical connectives, are interpreted in terms of the notion of *provability,* then existence becomes intra-theoretic.

5. *Correspondence Theory of Truth*

Now, what I want to suggest (the reader has probably been wondering what all this is leading up to!) is that the effect of abandoning realism—that is, abandoning the belief in any describable world of unobservable things, and accepting in its place the belief that all the "unobservable things" (and, possibly, the observable things as well) spoken of in any generation's scientific theories, including our own, are *mere* theoretical conveniences, destined to be replaced and supplanted by quite different and unrelated theoretical constructions in the future—would *not* be a total scrapping of the predicates *true* and *refers* in their *formal* aspects. We could, as the above discussion indicates, *keep* formal semantics (including "Tarski-type" truth-definitions); even keep classical logic; and yet *shift* our notion of "truth" over to something approximating "warranted assertibility". And I believe that this shift is what would in fact happen. (Of course, the formal details are only a rational reconstruction, and not the only possible one at that.)

Of course, there isn't any question of *proving* such a claim. It is a speculation about human cognitive nature, couched in the form of a prediction about an hypothetical situation. But what makes it plausible is that just such a subsitution—a subsitution of "truth within the theory" or "warranted assertibility" for the realist notion of truth—has *always* accompanied scepticism about the realist notion from Protagoras to Michael Dummett.

If this is right, then what is the answer to our original question: what is the relation between realist explanations of the scientific method, its success, its convergence, and the realist view of truth?

We remarked at the outset that realists claim to believe in something called a Correspondence Theory of Truth. But what is that?

If I am right, it isn't a different *definition* of truth. There

is only one way anyone knows how to *define* "true" and that is Tarski's way. (Actually, as we mentioned earlier, Saul Kripke has a *new* way—but the difference from Tarski is inessential in this context, although it is important for the treatment of the antinomies.) But is Tarski's way "realist"?

Well, it depends. If the logical connectives are understood realistically ("classically", as people say), then a Tarski-type truth-definition is "realist" to at least this degree: satisfaction (of which truth is a special case) is a relation beween words and things—more precisely, between formulas and finite sequences of things. ("Satisfies" is the technical term Tarski uses for what I have been calling *reference*. For example, instead of saying " 'Electron' refers to electrons", he would say "The sequence of length one consisting of just x satisfies the formula 'Electron (y)' if and only if x is an electron". "Satisfies" has the technical advantage of applying to n-place formulas. For example, one can say that the sequence *Abraham;Isaac* satisfies the formula 'x is the father of y'; but it is not customary to use "refers" in connexion with dyadic, *etc.*, formulas, *e.g.*, to say that "father of" *refers* to *Abraham;Isaac*.) This certainly conforms to an essential part of the idea of a Correspondence Theory.

Still, one tends to feel dissatisfied with the Tarski theory as a reconstruction of the "Correspondence Theory of Truth" *even if* the logical connectives are understood classically. I think that there are a number of sources of this dissatisfaction, which I have expressed myself in some of my writings, but it seems to me that Hartry Field[4] put his finger on the main one: the fact that primitive reference (*i.e.*, *satisfaction* in the case of primitive predicates of the language) is "explained" by a *list* is the big cause of distress.

But the list has a very special *structure*. Look at the following clauses from the definition of primitive reference:

(1) "Electron" refers to electrons.

(2) "Gene" refers to genes.

(3) "DNA molecule" refers to DNA molecules.

These are similar to the famous

(4) "Snow is white" is true if and only if snow is white.

—and the similarity is not coincidental: "true" is the O-adic case of satisfaction (a formula is true if it has no free variables and the null sequence satisfies it). The Criterion of Adequacy (Criterion *T*) can be generalized as follows:

(Call the result "Criterion *S*"—"*S*" for Satisfaction:)

*An adequate definition of satisfies-in-*L *must yield as theorems all instances of the following schema*:

$\ulcorner P(x_1, \ldots, x_n)\urcorner$ is satisfied by the sequence $y_1, \ldots, y_n$ if and only if $P(y_1, \ldots, y_n)$.

Rewriting (1) above as

(1′) "Electron (x)" is satisfied by y_1 if and only if y_1 is an electron—which is how it would be written in the first place in Tarski-ese—we see that the structure of the list Field objects to is *determined* by Criterion *S*. But these criteria—*T*, or its natural generalization to formulas containing free variables, *S*—are determined by the formal properties we want the notion of truth to have (this is discussed at length in my John Locke Lectures), by the fact that we *need* for a variety of purposes to have a predicate in our meta-language that satisfies precisely the Criterion *T*. (This is why we would *keep* Criterion *T* even if we went over to an Intuitionist or quasi-Intuitionist meaning for the logical connectives.)

So I conclude (as I argue in the John Locke Lectures) that Field's objection fails, and that it is correct for the realist to define "true" à la Tarski. Even though the notion of truth is derived, so to speak, by a "transcendental deduction" (the argument, which I cannot give here, that we *need* a meta-linguistic notion satisfying Criterion *T*), and Criterion *S* is only justified as a "natural" generalization of *T*, satisfaction or reference is still, viewed from within our realist conceptual scheme, a relation between words and things—and one of explanatory value, as Boyd's argument shows.

Now that I have laid out this argument, let me give a shorter and sloppier argument to somewhat the same effect:

" 'Electron' refers to electrons"—*how else* should we say what "electron" refers to from *within* a conceptual system in which "electron" is a *primitive* term?

As soon as we *analyse electrons*—say, "electrons are particles with such-and-such mass and negative unit charge"—we can say " 'electron' refers to particles of such-and-such mass and negative unit charge"— but then "charge" (or whatever the primitive notions may be in our new theory) will be explained "trivially", that is, in accordance with Criterion *S*. Given the Quinian Predicament (Kantian Predicament?) that there is a real world *but* we can only describe it in the terms of our own conceptual system (Well? We should use *someone else's* conceptual system?) *is it surprising* that *primitive* reference has this character of apparent triviality?

I conclude, dear reader, that the Tarski theory is a "correspondence theory" in any sense one could reasonably ask for *if* the logical connectives are interpreted realistically.

6. *Truth and Knowledge*

What is it to interpret the logical connectives realistically? We have seen what it is *not*: the fact that one accepts classical logic does not show that one understands the logical connectives realistically. (Nor does the fact that one rejects it show that one understands them idealistically. *Cf.* my interpretation of quantum mechanics *via* a non-standard logic.) Nor is it just a question of accepting Criterion *T* or even Criterion *S*, or a question of accepting a Tarski-style truth-definition for one's language.

What does show that one understands the connectives realistically is one's acceptance of such statements as:

(A) Venus could have carbon dioxide in its atmosphere even if it didn't follow from our theory (or even from our theory plus the set of true "observation sentences") that Venus has carbon dioxide in its atmosphere.

—and

(B) A statement can be true even though it doesn't follow from our theory (or from our theory plus the set of true observation sentences).

Now (B) follows from any sentence of the general form (A). So why do we believe (A) (and many similar sentences)? The answer is that as realists we view *knowledge itself* as the

product of certain types of causal interactions, at least in such cases as "Venus has carbon dioxide in its atmosphere". And it *follows from our theory* of the interaction whereby we learned this fact—for example, the standard causal account of perception—that we might have, for any number of reasons, made up a theory from which it *didn't* follow that Venus has carbon dioxide in its atmosphere even though Venus *does* have carbon dioxide in its atmosphere. In short, (A) is itself a "scientific" (or even a *common sense*) *fact about the world* (albeit a *modal* fact about the world). But given the obviousness and centrality to our understanding of knowledge of facts like (A), how could anyone *not* understand the logical connectives realistically? How could anyone not be a realist?

Historically, one possible tack was to *accept* (A), *accept* the "realist" (classical) account of the logical connectives, but to give an idealist account of the meanings of the descriptive terms (*i.e.*, the predicates, or at least the "theoretical vocabulary"). But with the failure of the reduction programmes of phenomenalism and Logical Empiricism, that way was blocked. The more feasible tack, if one believes that scientific knowledge does not converge, would be to argue that the *phenomenon of scientific revolutions* shows that the realist notion of reference (and hence of truth) leads to disaster (*via* the meta-induction I discussed in section 2) and so we must fall back on an Intuitionist or quasi-Intuitionist reading of the logical connectives, which would save the bulk of extensional scientific theory, and save the formal part of our theories of reference and truth, at the cost of giving up (A) and (B).

The realist, in effect, argues that science should be taken at "face value"—without philosophical reinterpretation—in the light of the failure of all serious programmes of philosophical reinterpretation of science, and that science taken at "face value" *implies* realism. (Realism is, so to speak, "science's philosophy of science".) The opponent replies (assuming *no* convergence) that science itself—viewed *diachronically*—refutes realism. But the failure of convergence is crucial to this sort of anti-realist argument. If Boyd is right in claiming that the mature sciences do "converge" (in a very sophisticated sense), and that that convergence has great explanatory

value for the theory of science, then this sort of anti-realism, "cultural relativist" anti-realism, is bankrupt.

To sum up: Realism depends on a way of understanding the logical connectives (not just "truth", not just the rejection of reductionist analyses of the descriptive terms). This way of understanding the connectives depends on taking science at "face value" in a very strong sense—counting (B) as part of science. It also depends on blocking the disastrous meta-induction that concludes "no theoretical term ever refers". But blocking that meta-induction by a theory of science which stresses the "limiting case" relation between successor theories, and which employs a "causal" theory of reference, commits us to viewing the scientific method as not given *a priori*, but as dependent on our highest level empirical generalizations about knowledge itself, construed as an interaction with the world. Both our reasons for believing in a sophisticated version of convergence, such as Boyd's, and our reasons for accepting (B), have to do with our over-all view of *knowledge as part of the subject of our knowledge.*

Idealists have always maintained that our notion of truth depends on our understanding of our theory and of the activity of "discovering" it, *as a whole*. If I am right, then this is an insight of idealism that realists need to accept—though not in the way idealists meant it, of course.

NOTES

[1] *Realism and Scientific Epistemology*, 1973 (privately circulated).

[2] In "What Theories are Not", reprinted in my *Mathematics, Matter, and Method, Philosophical Papers*, Volume 1, Cambridge Univ. Press, 1975.

[3] In "Language and Reality", in my *Mind, Language and Reality, Philosophical Papers*, Volume 2, Cambridge Univ. Press, 1975.

[4] *Cf.* his "Tarski's Theory of Truth", *Journal of Philosophy*, vol. 69, no. 13, July 13, 1972, pp. 347–375.