HILARY PUTNAM HAD A GREAT FALL

A Refutation of the Model-Theoretic Argument Against Scientific Realism

Hannah Clark-Younger

Hilary Putnam's Model-Theoretic argument attempts to dispose of the view that science provides us with a literally true description of the world. It uses the Löwenheim-Skolem Theorem to show that all consistent scientific theories are true of the real world, even if they contradict each other. This is incompatible with the realist conception of truth, which allows only for one version of reality. So, Putnam rejected this conception of truth in favour of his own 'internal realism,' which claims that theories can only be true or false within themselves; within their own framework. However, I demonstrate that the Model-Theoretic argument is unsound. It relies on a false premise – that every consistent theory is true of the real world. Putnam uses a clever manipulation of model theory to attempt to convince us of the truth of this premise, but I show that he commits a fundamental error in his method.

Hilary Putnam's Model-Theoretic argument attempts to refute the Realist's 'correspondence' conception of truth, thus undermining Scientific Realism's view about the aims and successes of science. He uses the Löwenheim-Skolem Theorem in mathematical model theory, and some clever footwork, to show that every consistent theory is true of the real world. However, I will show that he makes a fundamental error when he assigns any objects to any predicates to make the theory come out true. In doing this, he is creating a new language, and thus a new theory, abandoning the original theory and in fact proving *another* theory to be true. This error means that the truth of one of the premises in his Model-Theoretic argument is compromised, so it is unsound.

The debate between Realist and Anti-realist interpretations of scientific theories and the scientific method is central to the philosophy of science, as the interpretation of the discoveries of science is of utmost importance in our discovery of the world. Generally, the debate consists of Anti-realists of one persuasion or another launching attacks on the 'naïve' pre-theoretical position of Scientific Realism, and Realists defending themselves. So, what is it all about? 'Scientific Realism' is essentially the view that science aims for a literally true account of the world, and the extent to which a theory achieves this determines the success of that theory. Necessary to this view is a reality that is ontologically independent from any human conceptual framework or ideas about the world. Science aims to make discoveries about this reality, and formulate theories which are literally true of this external reality. Successful scientific theories are those which accurately depict this real world. Also, Scientific Realists generally, but not necessarily, believe that science has been more or less successful in this aim, and so we have a rational reason to believe (tentatively) that at least the best of our scientific theories are literally true accounts of the world (Putnam 1976). Anti-realism is defined in terms of *not* being Realism, and comes in as many forms as there are aspects of Realism to be denied. Bas van Fraassen's Constructive Empiricism, for example, denies that the measure of success in science is truth, and instead claims that *empirical adequacy* is what science should aim for. According to this version of Anti-realism, there *is* a reality to the world, but it is of no consequence whether a scientific theory really is true. All that matters is that the theory can correctly predict all of the phenomena within its scope. The more phenomena that can be acccurately predicted, the better the theory (van Fraassen 1984). Putnam's version of Anti-realism, Internal Realism, denies an even more fundamental tenet of Realism – that there is one objective, independent reality to the world. This leads him to the belief that theories can only ever be consistent or inconsistent; or 'true within themselves.'

In his early career, Hilary Putnam fell squarely on the side of the Realist. To the early Putnam, along with mainstream Scientific Realism, there is one objective reality or way the world is, and truth consists of corresponding to or lining up with this objective reality. The aim of science is to discover as much of this truth about the world as possible, and science is generally successful in this endeavour (Putnam 1976). Putnam was even one of the main proponents of what many people consider to be the 'Ultimate Argument for Realism' – that if modern science was not true, then all of the success of science would be a miracle (Putnam 1976). So, Putnam believed, modern scientific theories could be considered to be literally true accounts of the external world.

In the late 1970s, when he was in his fifties, Putnam came across the Löwenheim-Skolem Theorem, which was originally formulated by Leopold Löwenheim in a paper published in 1915, and the proof modified and corrected by Thoraf Skolem in 1920. From this theorem, Putnam developed an argument that he considered so strong as to be definitive proof against Realism.

This Model-Theoretic argument convinced the Realist Putnam that his assumptions about truth were wrong. He concluded that truth was relative, not absolute. Truth is only possible *within* an interpretive framework, or a conceptual scheme; never objectively or external to human concepts. This means that there is no objective way the world is regardless of what humans happen to think about it. Any statement is only true or false within a particular broad theory or worldview if it is consistent with that theory. The consistency of the statement is what *makes* it true or false within that theory, and there is no *deeper* truth; the very idea of a theory being true or false *independent* of its conceptual framework is incoherent (Putnam 1980). To persuade Putnam himself to change his views as radically as this, the Model-Theoretic argument must be a very convincing one. Let us take a look at this indisputably solid argument.

Putnam's Model-Theoretic Argument against Scientific Realism can be summarised as follows:

1) For every theory in first-order logic, if it is consistent, then it has a countable model
2) Every scientific theory can be expressed in first-order logic

   ∴

3) For every scientific theory, if it is consistent, then it has a countable model

4) Every theory with a countable model is true of the real world

∴

5) Every consistent scientific theory is true of the real world

Putnam considered this to be a knockdown argument against what he called 'Metaphysical Realism' (Putnam 1980, p. 466). Metaphysical Realism is an essential aspect of Scientific Realism; it is the view that there is one true way the world actually is, independent of us. A true statement is one which corresponds accurately to this external world. Truth consists of lining up with an external reality, so any theory which expresses a reality different to the actual one is false. Even *ideal* scientific theories might be false; that is, theories which conform to stringent epistemological criteria such as being empirically adequate, simple, and unified. No matter how elegant, cohesive, and simple a theory, no matter how much empirical support there is for a theory, that theory may still be false if it does not accurately describe or represent the reality of the world.

Metaphysical Realism is, however, just a natural consequence of the *Realist conception of truth* – and in fact any proposition about the world, not just scientific theories, will come out true through Putnam's application of the Löwenheim-Skolem Theorem. So, it is really the Realist's 'correspondence' theory of truth that is called into question by the Model-Theoretic argument. This can be shown clearly with the following addition to the argument:

6) It is not the case that two contradictory theories can both be true of the real world
7) It is possible for two or more consistent scientific theories to contradict each other

∴

8) Some consistent scientific theories are not true of the real world

5) and 8) contradict each other, so by *reductio* Putnam concludes that 6) 'it is not the case that two contradictory theories can both be true of the real world' is false, and so contradictory theories can indeed be true of the real world. Putnam thinks the first argument is sound, thus accepting 5), so accordingly rejects 8), and the Realist conception of truth on which it depends. Scientific theories are not ever true in an objective, independent of humanity sense.

The second of the contradictory conclusions, 8), follows validly from 6) and 7). Premise 6) 'it is not the case that two contradictory theories can both be true of the real world' is a direct consequence of the fundamental tenet of the Realist notion of truth, which is exactly what is at stake in this debate, so the Realist will not reject it without being convinced of 5). Premise 7) 'it is possible for two or more consistent scientific theories to contradict each other' is uncontentious, but as an example, 'all ducks are birds' and 'no ducks are birds' are clearly contradictory, and both are falsifiable (scientific) theories about the world. It is the first conclusion, 5), which presents the challenge to Realism, so the Realist must find fault in the first argument. The argument 1) – 5) is valid, so one of premises 1) 2) or 4) must be shown to be false, if Realism is to be a defensible position. So, I will now examine these three premises.

Premise 1) 'for every theory in first-order logic, if it is consistent, then it has a countable model' is a formulation of the Löwenheim-Skolem Theorem. The Löwenheim-Skolem Theorem is a theorem in mathematical model theory which states that if a first-order theory has any infinite model (an interpretation in which it is true), then it has a countable model (Bays 2001). This means that for any theory, as long as it is translatable into first-order logic, if it has an interpretation with an infinite domain in which it is true, then it has an interpretation with a countable domain in which it is true. To break this down, first-order logic is any system of symbolic logic which contains names, predicates, connectives and quantifiers. Every (literal) English sentence can be translated into first-order logic. An 'interpretation' of a theory is a collection of sets – a set of at least one object as its domain, with subsets of objects from the domain assigned to the one-place predicates, ordered pairs of the objects in the domain to each of the two-place predicates and so on. This collection of allocations is called an interpretation. A 'model' of a theory is an interpretation in which the theory is true. 'Truth in an interpretation' means that all of the statements (in formal logic) are true within that interpretation. So, for example, $(\forall x)(Fx \rightarrow Gx)$ is true in any interpretation in which the set of objects assigned to 'F' is a subset of the objects assigned to 'G,' because in that case, every object that is 'F' is also 'G.'

A 'countable' domain means a domain containing objects which can, in theory, be counted up. For example, the set of real numbers (1, 1.00001, 1.00002...) is *not* countable, because there are an infinite number of real numbers between any two given real numbers. One could start, for example, at '1,' but there is no *next* real number. Between 1 and 1.1, there is 1.01 to 1.09, between 1 and 1.01 there is 1.001 to 1.009, and so forth. In contrast, the set of *natural* numbers is a countable set, because there is an obvious progression from each natural number. Starting anywhere, say at 22, there *is* a number which comes next, in this case 23. So, the Löwenheim-Skolem Theorem tells us, every first-order theory that has a true interpretation in which the domain contains an infinite number of objects, must also have a true interpretation in which the domain is countable.

The proviso of the domain being infinite is trivial, because if a theory has an interpretation with a finite number of objects in the domain, then that same interpretation will work with the domain being expanded to include more (even an infinite number of) objects, as long as nothing else changes. For this reason, 'a theory which has an interpretation in which it is true with an infinite domain' amounts to 'a theory with at least one possible interpretation in which it is true,' which in turn amounts to 'a theory which is *consistent*.' So, provided Löwenheim and Skolem knew their model theory, premise 1) 'for every theory in first-order logic, if it is consistent, then it has a countable model' is true.

Premise 2) 'every scientific theory can be expressed in first-order logic' is, for our purposes, straightforward. The purpose of any first-order logic is to be able to symbolise language in order for propositions be clearly analysed. Any statement about the world can be symbolised into first-order logic. Scientific theories are statements about the state of the world, so scientific theories can be symbolised into first-order logic. Premises 1) and 2) together validly guarantee 3). All this is saying, so far, is that any consistent scientific theory can be modelled by a countable set such as the natural numbers.

Premise 4) 'every theory with a countable model is true of the real world' is doing all the work in creating a problem for the Realist definition of truth, and it expresses the point that Putnam devotes most of his article *Models and Reality* to defending (Putnam 1980). Putnam gives an account which attempts to convince us of the truth of 4). Suppose we write down our scientific theory in first-order logic. We then take the natural numbers (a countable set) as the domain of an interpretation and assign a natural number to each of the names in the theory, assign a set of natural numbers to each one-place predicate, a set of ordered pairs of natural numbers to each two-place predicate, and so on. We do this in such a way that the theory comes out true in this interpretation. We have shown the theory to be consistent – that is, it is not *impossible* to construct a true interpretation. This is an unproblematic way of testing the consistency of theories, and demonstrates 1) – 3) in the first argument. Now, Putnam urges, if we make a new domain – one which is made up of real-world objects – we can assign each of the numbers from the first interpretation onto an object from the second. Keeping the equivalent sets of objects in each predicate the same, we now have a theory that is *true of the real world* (Putnam 1980).

An example will illustrate this point. Take the scientific theory 'all cats are fish.' If this is consistent – that is, not contradictory – we can make an interpretation with the natural numbers as the domain. Let the domain be the natural numbers, let us assign the predicate 'is a cat' to refer to the set {1, 2, 4, 6, 7} and the predicate 'is a fish' to refer to the set {1, 2, 3, 4, 5, 6, 7, 8}. The statement 'all cats are fish' is true in this interpretation. We have shown 'all cats are fish' to be a non-contradictory theory. Now, to demonstrate Putnam's method, we match the numbers up to objects in the real world; for example like this:

1: a table

2: a watch

3: a cat

4: a monkey

5: a woollen jersey

6: a grain of sand

7: a goldfish

8: a scarf

Now, the predicate 'is a cat' refers the set which consists of the table, the watch, the monkey, the grain of sand and the goldfish. All of these items ('cats') now fall into the set of objects assigned to the predicate 'is a fish,' so the theory 'all cats are fish' is true of the real world.

Of course, the intuitive first response is to say: what is meant by 'is a cat' and 'is a fish' is fixed – these terms must refer only to cats and fish respectively. So, Putnam says, we write this extra part of the theory down – 'is a cat' refers to object 3, and 'is a fish' refers to object 7. However, as this is now part of the theory, we can make another interpretation where these propositions *also* come out true (Putnam 1980). Let us see how it works. We now have:

All cats are fish

The term 'cat' denotes object 3

The term 'fish' denotes object 7

To symbolise into first-order logic, we will use the following notation:

Cx = 'x is a cat'

Fx = 'x is a fish'

Tx = ' x is the term 'cat''

Sx = 'x is the term 'fish''

Ox = 'x is object 3'

Bx = 'x is object 7'

Dxy = 'x denotes y'

And now our theory can be symbolised as follows:

$(\forall x)(Cx \rightarrow Fx)$

$(\forall x)(\forall y)((Tx \ \& \ Oy) \rightarrow Dxy)$

$(\forall x)(\forall y)((Sx \ \& \ By) \rightarrow Dxy)$

But now, Putnam could give us the following interpretation:

Domain: {1, 2, 3, 4, 5, 6, 7, 8....}

C: {1, 2, 4, 6, 7}

F: {1, 2, 3, 4, 5, 6, 7, 8}

T: {1}

S: {2}

O: {5}

B: {6}

D: {<1,5>,<2,6>}

The new, total theory, including the constraints on denotation, is true in this interpretation. $(\forall x)(Cx \rightarrow Fx)$ is the same theory as above, and is satisfied in the same way, by the items assigned to 'C' being a subset of those assigned to 'F.' The second statement, $(\forall x)(\forall y)((Tx \ \& \ Oy) \rightarrow Dxy)$, is satisfied by T: {1}, O: {5} and D(1,5). In every set of two objects that contains one which is 'T' and another which is 'O,' the 'T' one denotes the 'O' one. Similarly, $(\forall x)(\forall y)((Sx \ \& \ By) \rightarrow Dxy)$ is satisfied by S: {2}, B: {6} and D(2,6). Again, Putnam can make each number represent an object in the real world just as before, and so our intuitive objection does not work. This is how premise 4) is justified. It does not matter what the original theory is, or how many constraints on denotation there are; as long as it is consistent, it can be modelled by a countable set such as the natural numbers, and then by Putnam's method there is a true interpretation *in the real world*. If we accept Putnam's method of applying the Löwenheim-Skolem theorem, then premise 4) is true, 1) – 5) is sound and so 5) 'every consistent scientific theory is true of the real world' is true.

However, Putnam's method is flawed, and the initial objection to it *does* work. Premise 4) is false. As David Lewis pointed out, an interpretation in which the theory plus the constraints on reference are *true* is not the same as an interpretation which *conforms to* the constraints on reference (Lewis

1984). Saying that 'cat' refers to object 3 is not just adding more theory. Rather, it is an *overarching* reference scheme which the final interpretation must conform to. It is a constraint *not* on the consistency of 'all cats are fish' but on the *truth* of it. What we really meant was, 'in any interpretation where C: {object 3} and F: {object 7}, all cats are fish.' *This* is clearly false. Objects 3 and 7 are *names,* not *predicates*. They are *members of the domain* – they cannot have items in the domain assigned to them like predicates do – that is a category error. 'The term 'cat'' and 'the term 'fish'' are *not* separate predicates from 'is a cat' and 'is a fish.' They are not predicates at all; they also cannot have items from the domain assigned to them. We did *not* mean 'denotes' to be a two-place predicate, but rather a constraint on which object can be included in the sets of 'C' and 'F,' when testing the *truth* of the theory. These constraints on denotation do not apply to the test of *consistency* which is carried out in the language of first-order logic. The situation is actually: $(\forall x)(Cx \rightarrow Fx)$ has an interpretation with a countable domain in which it is true, which demonstrates that it is consistent. However, when the interpretation includes C: {object 3} and F: {object 7} (which is just a way of formalising the English language), $(\forall x)(Cx \rightarrow Fx)$ is false.

There is a trivial sense in which Putnam is right, however. What he is actually doing is creating a new language, English II, where English words make up the vocabulary, but they have different meanings than in English. So, 'all cats are fish' *is* true, but only in English II, where 'cat' in English II refers to the set of 'the table, the watch, the monkey, the grain of sand and the goldfish,' and 'fish' in English II refers to the set of 'the table, the watch, the cat, the monkey, the woollen jersey, the grain of sand, the goldfish and the scarf.' In this (albeit bizarre) language, 'cats' are a subset of 'fish,' so 'all cats are fish' *is* true. However, the original theory was postulated in *English,* not in English II, so if we are to have any chance at meaningful discourse, we must be consistent and use English throughout our entire analysis, including in the final interpretation.

Agreeing to speak in the same language does not imply that English words are somehow 'stuck' onto objects with some kind of 'magic glue,' nor that the mind needs to have any 'mysterious' non-natural power to 'directly grasp' Platonic forms (Putnam 1980). It just means that we have a mutually agreed upon *reference* scheme. Using a common language is not part of a theory about the world. Using the term 'cat' to refer to cats is not part of a theory – there need not be any ontological connection between the word and the objects in order for there to be a semantic one. It does not even necessitate any real significance or similarity between the objects which belong to each group. Saying that 'cat' denotes cats is not saying anything about the reality of the world. It is, instead, agreeing (largely for pragmatic reasons) to use a certain word to refer to certain objects, so that we can discuss them with each other. Language is merely a tool, used for exchanging ideas. The truth of our theories is not *reliant* on this reference scheme (language). The world is the way it is, regardless of which words we use to describe it. However, we do need to use *some* words to describe the world when we discuss it with each other. Whatever we agree them to be, we must be consistent from the start to the end of the analysis of the theory.

This reference scheme is also not merely an 'interpretive framework,' within which theories can be true or false. It is not more of the theory. It is just, as I have stressed, an agreed system of reference; that is, which words we use to refer to which objects. The truth or falsity of scientific theories is determined by reality of the actual objects in the real world. Words are just our tools of discourse.

Premise 4) 'every theory with a countable model is true of the real world' is false – it relies on faulty reasoning. So, the conclusion to the first argument, 5), is not supported by this argument. Realists can safely believe that 6) 'it is not the case that two contradictory theories can both be true of the real world,' and the correspondence theory of truth which this principle derives from, without fear of contradiction by the Löwenheim-Skolem Theorem.

Hilary Putnam commits a fundamental error in his reasoning when he attempts to show that all consistent scientific theories are true of the real world. His method is to assign objects to each predicate in the theory, regardless of what the objects and predicates are, in such a way as to make the theory come out true. In doing this, he creates another language, English II, and so is not testing the truth of the English theory any more. Rather, he is testing the truth of a *different* theory, one presented in English II. Because of this error, one of the premises in the Model-Theoretic argument is false, rendering it unsound. The Realist conception of truth does not need to be abandoned, and we do not need to live in Putnam's Humpty Dumpty world.

- *University of Otago*

BIBLIOGRAPHY

Bays, Timothy. (2001) On Putnam and His Models. *The Journal of Philosophy* 98 (7), 331-350.

Lewis, David. (1984) Putnam's Paradox. *Australasian Journal of Philosophy* 62 (3), 221-236.

Merrill, G. H. (1980) The Model-Theoretic Argument Against Realism. *Philosophy of Science* 47 (1), 69-81.

Putnam, Hilary. (1976) What is Realism? *Proceedings of the Aristotelian Society* 76, 177-194.

Putnam, Hilary. (1980) Models and Reality. *The Journal of Symbolic Logic* 45 (3), 464-482.

van Fraassen, Bas. (1984) To Save the Phenomena. In: Leplin, Jarrett (ed.) *Scientific Realism.* California, University of California Press, 250-259.

van Fraassen, Bas. (1997) Putnam's Paradox: Metaphysical Realism Revamped and Evaded. *Nous* 31, 17-42.