

## MODELS, REFERENCE AND REALITY: INTERNAL REALISM AND BEYOND

ROGER VERGAUWEN

### 1. *Introduction*

It goes without saying that the realism vs. anti-realism debate is at the heart of much contemporary work in analytical philosophy. ‘Realism’ and ‘anti-realism’ are, then, titles beneath which a remarkably wide range of philosophical views has been propounded.

*In the first sense*, to be a realist about some particular things or kind of things is to believe that that thing or kind of thing exists. In this sense, the antirealist believes that the thing or the kind in question does not exist (so, in this sense, philosophers may be realists about tables but antirealists about electrons (Heller 1988, 113)). Since these realists and antirealists are discussing the existence of a particular kind of entities, we might call them ontological realists or, alternatively, ontological antirealists. Their disagreement is about what the right answer to the existence question is, but they both agree on there being some right answer to the question. *In the second sense*, to be a realist about some particular object or kind is to believe that *there is a fact of the matter* as to whether that object or kind exists. Antirealists in the second sense believe that there is no fact of the matter about these objects or kinds. An antirealist in this sense might accept a fact of the matter relative to a conceptual scheme or to a background theory, but he would say that there is no non-relative fact of the matter. While the first kind of realism could be dubbed as belonging to *ontological realism* (cf. Nola 1988, 4 ff), according to which there is something which exists in a suitably mind-independent manner, whether this ‘something’ is individual objects or some kinds (such as electrons, galaxies) which are open to scientific examination, anti-realism in the second sense would belong, rather, to *ontological relativism*, the view that what exists, whether that be objects, facts, or the entities postulated in science, exists only relative to some relativizer, whether that be a person, a theory, or whatever (Cf. Vergauwen 1996, 130–131).

Ontological realism, then, amounts to the thesis that there is a world, independent of thought and language and it usually goes together with *epistemological realism*, the view that in thought we may have knowledge of this

mind-independent reality and that we may truly speak of it. The relevant versions of anti-realism can, then, be understood as involving the denial of part of or both of these theses.

The aim of this paper is to show how H. Putnam's arguments against ontological realism (which is also called *metaphysical realism* or *externalism*) and his own *internal* realism can be reconciled with a version of metaphysical realism which answers certain of the criticisms put forward by Putnam. Putnam's arguments are of a logical and epistemological nature. It is argued that both of these arguments are insufficient to rebut externalism. More specifically, Putnam's internal realism seems to point to a notion of reference which goes beyond the one it is itself committed to.

## 2. *Metaphysical Realism and its critiques*

As a theory of Reference and Meaning, metaphysical realism can be characterized by the following set of basic tenets (Cf. Putnam 1981, 49; 1988, 214, Anderson 1992).

- 1a. *The World (i.e. the world as it is, as unconceptualized) is independent of any specific representation which we may have of it.* It might be the case that we are in principle unable to represent it as it really is. There is, moreover, exactly *one true and complete description* of the way the world is (though we might never come to know it).
- 1b. *The World can be subdivided into finitely (or infinitely) many parts.* It contains finitely or infinitely many objects.
- 1c. For each language or theory there exists *a unique correspondence relation* or '*reference relation*' to the world. *Truth* involves some sort of correspondence between thought signs or words and external things.
- 1d. *Truth is radically non-epistemic.* Even an 'ideal' theory (from the point of view of 'simplicity', 'plausibility', 'mathematical elegance' or 'explanatory power') might in principle be false of the World. "Verified does not imply true on the metaphysical realist picture, even in the ideal limit" (Putnam 1978, 125).

According to Putnam such a view is seriously flawed. In order to rebut this realism, Putnam (e.g. in Putnam 1978, 1980, 1981) brings into the field a number a model-theoretic arguments designed to contradict the claims above. From Putnam's presentation of metaphysical realism especially two related doctrines come to the fore (Hallett 1994, 67). The first is that there

is a fixed and unique language-independent world and second that there is a unique reference relation between language and the world that is determined by the world itself. What, now, Putnam wants to convince us of is that there is no plausible theory of reference which will allow for the desired type of correspondence (Heller 1988, 114). Putnam's argument falls into two parts. A first part stating that externalism (metaphysical realism) requires the ability to refer to objects as they are in themselves. That is, any truth must be about the objects that really compose the world, not merely about objects posited in (scientific) theories. A second part is that we cannot have the required ability to refer at all and that we could only refer to bits of the world as it is conceptualized by us. This implies that externalism insists that independently of our theories about the world or our way of conceptualizing it, *there is a way the world really is* and that is denied by Putnam's 'internal realism' to a certain extent.

Usually, when explaining Putnam's arguments against realism, one turns to his use of the Löwenheim-Skolem theorem (Putnam, 1980). I will, however, here concentrate on an argument already present in Putnam's *Meaning and the Moral Sciences* (Putnam, 1978) (For his use of the Löwenheim-Skolem theorem, cf. e.g. Vergauwen 1993), since what is given there is sufficient for our purpose, which is to indicate why Putnam's argument doesn't really work in that it does not exclude the possibility of there being an intended model for the theory of reference that goes with metaphysical realism.

According to Putnam: "The problem that the believer in metaphysical realism has always faced involves the notion of correspondence. There are many (in fact, infinitely many) different ways of putting the signs of a language and the things in a set *S* in correspondence with one another, if the set *S* is infinite (and a very large finite number if *S* is a large finite set). Even if the 'correspondence' has to be a reference relation and we specify which *sentences* are to correspond to *states of affairs which actually obtain*, it follows from theorems of model theory that there are still infinitely many ways of specifying such a correspondence. How can we pick out any one correspondence between our words (or thoughts) and the supposed mind-independent things *if we have no direct access to the mind-independent things?*" (Putnam 1982, 143). If we suppose the reference relation as being given by a truth conditional (model-theoretic) semantics, then the understanding of a particular term in our language implies that we know to which part of the world reference is being made, or what this term is *true of* in the world. There must be a specific reference relation between terms in the language and parts or sets of parts of the world. The problem now, Putnam believes, is that a metaphysical realist considers 'the world' as existing separately from any possible representation which we have of it, in such a way that we could be

profoundly mistaken as to the true nature of the world, whatever our theory. This implies that truth is a radically non-epistemic notion for metaphysical realism. Putnam finds such a notion beyond comprehension for the following reasons (Putnam 1978, 126): Let us take a theory T1 which in a sense is ideal. It possesses all the possible properties of ‘consistency’, ‘observational adequacy’, ‘simplicity’, ‘elegance’ and whatever observational and theoretical constraints which one can possibly conceive for an ideal theory. For a metaphysical realist such a theory could, nonetheless, be false ‘of the world’ or ‘in reality’. Is this possible at all? In model-theoretic terms this means the following: Imagine that ‘the world’ can be split up into an infinite number of objects. Assuming that the theory T1 is consistent, it has an infinite number of models and, according to the Löwenheim-Skolem theorem in its ‘upward version’, there are also models in each infinite cardinality. Let us also suppose that we select a specific model M which has the same cardinal number as ‘the world’ and that we put the individuals or objects from this model in a one-to-one correspondence with the world. This means that a *satisfaction relation* SAT is created between the elements from the domain of M and (parts of) ‘the world’ via the formulae of the language in which the theory is formalized: “Since SAT establishes a correspondence between L (the language) and the world, then by virtue of this correspondence (a sentence) may now be seen as being *about the world* (rather than about the ‘artificial’ universe of M), and to say that  $\Phi$  is TRUE (SAT) is to say that  $\Phi$  is *true of the world (or really true)*” (Merrill 1980, 70).

Since each sentence of L in which T is formalized is true in M, and a sentence is true in M if and only if it is TRUE(SAT), then every sentence of T is quite simply TRUE (SAT), which is why it is said that T is true of the world. “More generally, Putnam’s argument retraces the model-theoretic proof that given any consistent theory demanding a universe of cardinality  $c$ , any given set of cardinality  $c$ , there will exist a model of the theory whose universe is that set” (Merrill 1980, 70). The consequence of the kind of correspondence as just described is that no consistent theory whatsoever which satisfies the stated operational constraints can be false of the world, because there is a model for this theory the domain of which consists of the set of objects which the world contains.

Merrill (Merrill 1980, 71 ff) has shown that Putnam’s model-theoretic argument against realism is directed against a notion of reference which states that there is a ‘real’ world which objectively exists and which contains a number of objectively existing entities *without any further qualification*. The question here is whether there are indeed any realists who accept such an unqualified objectivism. What could typically be referred to as realism is that it is not only accepted that there are objectively existing entities in the world

(observable or otherwise), but that these entities also stand in specific *objective relations* to each other. In this case it may be argued that the relation between a theory and reality does not immediately imply the problem which Putnam thinks there is in determining the intended interpretation (or an *intended model*) on the basis of which he then again concludes that realism is unintelligible. Central to this is the idea of a *Structured Domain*: in formal terms, a structured domain is a triple:

2.  $\langle D, P, R \rangle$  (Merrill 1980, 72)

In (2),  $D$  is a set (the domain),  $P$  belongs to the power set of  $D$ , and  $R$  is a set of relations on  $D$ , that is,  $R$  is a set of  $n$ -place relations among the elements of  $D$ . Such a structured domain can better be considered as the model theoretic analog of the world as the realist sees it. The entities which objectively exist in the world appear (in 2) as the elements of  $D$ , their properties as elements of  $P$ , and the relations between the entities in  $D$  are expressed by  $R$ . A 'realistic' position could now consist in saying that *the world* corresponds to a structured domain the parts of which are independent of any specific representation which we have of it. The realist does not have to accept that he knows the *actual* structure of the world, simply that the world *is a structured domain*. An interpretation,  $I$ , for a language,  $L$ , which may be the language of standard predicate logic, is a function whose domain is the set of predicates and variables from this language; this function assigns an object to each variable, a (possibly empty) set of object to the one-place predicates from its domain, and a (possibly empty) relation between the objects to the  $n$ -place predicates. "We may then say that  $I$  is an interpretation of  $L$  in (the structured domain)  $\langle D, P, R \rangle$  if and only if  $I$  is an interpretation of  $L$  and the range of  $I$  is included in  $D \cup P \cup R$ . That is, an interpretation of a language *in* a given structured domain assigns to variables and predicates of the language *only* objects, sets, or relations to be found in that domain" (Merrill 1980, 73). It differs from a 'traditional' model-theoretic approach in that the role of the model is fulfilled by a combination of a structured domain and an interpretation *within* that domain. This difference is not without importance, because whereas a model normally structures a domain by means of intension and/or extension assignments, the situation is not quite the same where interpretations in structured domains are concerned. In this case the structure is given first, independently of the language, and the interpretation maps the language onto the existing structure.

There is a clear difference between *the existence* of an interpretation for a theory and the *use* which we make of it. In order to apply a (scientific) theory we must, of course, use an interpretation of its empirical terms, because it is via these empirical terms that the theory is applied, but in applying a

theory, in using it to predict or in confirming or disconfirming it, we do not need to use any (referential) interpretation of its theoretical terms and we cannot do so because (as Putnam has argued) any attempted act of reference to the theoretical entities would be unsuccessful. But given the realist's view of the world as a structured domain, even though we do not (cannot) *use* any interpretation of the theoretical terms of our language, *there are* such interpretations nonetheless. "In some cases there is a successful correspondence between the world and the language of our theory even though we can claim neither to know what this correspondence is nor to be using this particular interpretation of our (theoretical) terms in applying the theory. But in those cases where there is such a correspondence, or in which there is more than one such correspondence, our theory is *really* true ... the fact that we cannot 'single out a unique relation' between the theoretical terms of our theory and the *real objects* is of no consequence here, and it should be clear that the realist's position is not lacking intelligibility" (Merrill 1980, 75–76).

However, we could say that even with the structured domain approach there still remains a problem for metaphysical realism, if it wants to be more than just a theoretical possibility of determining the right kind of reference relation: "the problem of reference for the metaphysical realist is just that of how the terms of the language we use are to be linked up to the stuff of 'The World' to thoroughly non-theoretical entities, entities which are not in the least dependent on our language" (Hallett 1994, 74). We shall indicate in the rest of this paper, through an analysis of Putnam's *internal realism*, why this requirement is in fact too strong and in what sense metaphysical realism may still survive, even if we give up this requirement.

### *Internal Realism Revisited: Virtual or Real Causality?*

Putnam's alternative for metaphysical realism is *internal realism*. He intends his internal realism to be a "first order theory about the relation of a language (actually the speakers of a language) to the speaker's environment. From within such a story the notion of a correspondence between words and sets of things is as legitimate as the notion of a chair or of a pain" (Putnam 1979, 228). This, of course, echoes Wittgenstein's Meaning-is-Use theory which is supported by Putnam: "I am not being cute. The point is that I am not offering a reductive account of truth. In Reason, Truth and History I explained the idea thus: *truth is idealized rational acceptability*. This formulation was taken by many as meaning that 'rational acceptability' ... is supposed (by me) to be more basic than truth. That I was offering a reduction of truth to epistemic notions. Nothing was farther from my intention. The suggestion is simply that truth and rational acceptability are interdependent

notions ... To repeat: the suggestion which constitutes the essence of internal realism is that *truth does not transcend use*" (Putnam 1988, 115).

No single theory-independent reference relation allows us to describe objects or reality as they *really* are, because reference can only be effectuated within a theory or a conceptual scheme.

Objects only exist within a theory and they are, then, at least partially constructed by the theory or the conceptual scheme they are part of: "Objects do not exist independently of conceptual schemes. *We* cut up the world into objects when we introduce one or another scheme of description. Since the objects and the signs are alike internal to the scheme of description, it is possible to say what matches what" (Putnam 1981, 52). There is no cosmic exile point of view from which we could speak about things as they 'really' are: "What we cannot say ... is what the facts are independent of all conceptual choices" (Putnam 1987, 33). Consequently, theories with incompatible ontologies may both be 'true-of-the-world'. "In my picture, objects are theory-dependent in the sense that theories with incompatible ontologies can both be right" (Putnam 1990, 40). Metaphysical Realism considers this to be impossible since there can—in principle—be only one description of the world as it 'really' is. More specifically (Naumann 1993, 483), it holds that two empirically equivalent theories of the world that have different ontologies have to be considered to be different. Since empirical data do not provide the possibility of distinguishing between equivalent theories, realism contends that there has to be something (*the things-in-themselves*) transcending the empirical facts which allows, through a unique correspondence relation, to determine which of the theories is the correct one.

It is here that realism invokes *causal relations* to determine the 'right' kind of connection between language and the world. They are supposed to be the means to determine the intended model for a scientific theory, thus also providing for the correctness of the structured domain approach, since as was said before, even there the use of an intended model for a scientific theory remained problematic: Causal connections provide the additional constraint that will enable reference to be fixed. Putnam himself used to assume that causal connections were the right answer to the reference-fixing problem, but he turned away from it since adding causal connections to the theory is 'adding more theory to the theory'. 'Causal connection' is just another linguistic symbol which itself lacks determinate reference without an acceptable reference-fixer. The reasons Putnam gives for rejecting this kind of connection are twofold.

They are logical and epistemological ones. The logical reason (Putnam 1980) is that the causal predicate already presupposes some kind of a determinate relation as it is used in the several models of a scientific theory. It is

never ‘general’ but always ‘local’: “If *refers* can be defined in terms of some causal predicate or predicates in the metalanguage of our theory, then, since each model of the object language extends in an obvious way to a corresponding model of the metalanguage, it will turn out that, in each model  $M$ ,  $reference_M$  is definable in terms of  $causes_M$ , but unless the word ‘causes’ (or whatever the causal predicate or predicates may be) is already glued to one definite relation with metaphysical glue, this does not fix a determinate extension for ‘refers’ at all” (Putnam 1980, 477). The other reason is epistemological and is based on Putnam’s *Brains-in-a-vat Argument* (Putnam 1981). The idea behind this is that due to their specific situation (in which everything, including the concept of causality, is, so to speak, a computer-induced simulation of the ‘real thing’) these brains cannot have the ability to refer to the outside world. More importantly, however, Putnam deduces from this that neither ‘normal’ human beings can refer to a mind-independent reality, since they too are situated within a certain realm of abilities.

Putnam, then, thinks that ‘causality’ is a context-dependent (theory-dependent) concept. To this, realists would tend to reply that it is not the linguistic symbol ‘causal connection’ that fixes reference, *it is causal connections themselves that do the work*: “We maintain that the use of language together with non-linguistic facts about the world (e.g. causal relations between the world and the use of language) do fix the intended interpretation of the language. Unless Putnam can discredit the foregoing claim, he has given us no reason to suppose that every attempt at a linguistic specification of a theory’s intended interpretation must fail” (Brueckner 1984, 137). Glymour further specifies this general idea: “The natural reply ... is that all intended interpretations should be replaced by talk of causally determined reference relations. Roughly, our physical and social circumstances, and sometimes our beliefs as well, determine together a set of links, connecting words and objects, and thus delimiting the admissible interpretations of our theories” (Glymour 1982, 177).

However, this leaves to the realist the burden of proof when and whether a causal connection is a non-linguistic one. Instead of trying to prove this we will try to indicate that Putnam’s argument in rejecting causation as a reference-fixer is flawed, because what he really describes is a *simulation* of the concept of causality which leaves out something in comparison with ‘real’ causation. As it will turn out, this will provide for a possible solution to the realist’s dilemma.

Putnam says: “Brains in a Vat can no more refer to what the unenvatted call “causation” than they can to what the unenvatted call “fire”. *For causal realists insist that the causal constraints that apply to our reference to any physical relation apply to reference to causation itself.* But by the

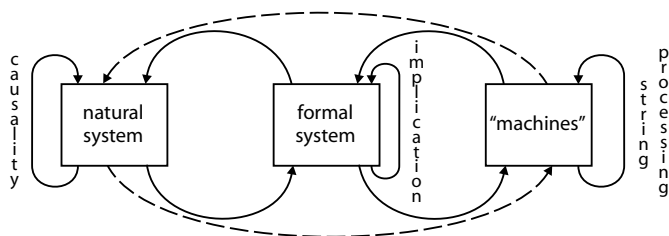


same token, we cannot be assumed to have available a notion of ‘causation’ that transcends our particular way of being situated in the world. But then, neither can we assume that we have a notion of “the intrinsic nature of mind-independent reality” —for what is *that* supposed to come to? If we say, with many scientific realists, that the intrinsic nature is given by the predicates needed, at the deepest level, for *causal explanation*, then what “intrinsic nature” refers to will depend on our situation in the world, just as the reference of “vat” depends on our situation in the world. It would seem that causal realism cannot give us the “view from nowhere” that metaphysical realism requires —or even the resources to allow that such a view is conceivable” (Putnam 1992, 362–363).

What Putnam describes (in the Brains-in-a-Vat-argument) is a concept of causality as simulated by a computer. But is it justified to transpose this to something like ‘worldly causality’ in the physical domain? Cleland (Cleland 1993) has shown that there may be a problem here. She claims that Turing-Machines (which is what computers really are) do not produce causal consequences. The computations performed by these machines are not causally linked to the instructions of the machine. They are not ‘caused’ by the instructions, they merely ‘follow’ them: this is in contrast with real causal processes (which Cleland calls ‘effective mundane procedures’): “In the case of an effective mundane procedure, the mere performance of the action-kinds specified by its instruction-list does not immediately produce the outcome; rather, the outcome is produced by a causal process which is generated by the performance of the action-kinds concerned. This is in stark contrast to the case of Turing-Machines, where nothing causal stands between the sequence of actions performed by the machine and its outcome. The sequence of actions performed by the machine literally is what produces the outcome” (Cleland 1993, 295).

This leads to the following picture of the relation between ‘natural causality’ and its simulation, which is merely ‘string processing’:

3.



(Rosen 1988, 531)

When we consider causality as described by a formal system (a logic) which can in its turn be simulated by the actions of a Turing-Machine computing its theorems, what happens is a ‘translation’ of something semantical into a purely syntactical mechanism. In doing so, information is lost (since syntax cannot adequately simulate semantics, as Gödel’s incompleteness Theorems show) and it is highly probable that Church’s Thesis would not be physically true: “Thus, in formal systems, we already find that a purely syntactical encoding will in some sense lose information. The information lost must then pertain to an irreducible, unformalizable semantic component in the original inferential structure. By changing the encodings, we can shift to some extent where this semantic information resides, but we cannot eliminate it. By itself, this result of Gödel does not bear on the physical truth of Church’s Thesis, since it is purely formal result. But it is in fact suggestive of how the physical form of Church’s Thesis might be verified or falsified... Formal models of material systems are then perfectly good formal systems, whose inferential structures by definition reflect causal processes in the natural system being modeled. Thus, if a model, arising in this fashion, should fall within the purview of Gödel’s argument, this would at least be strong evidence that Church’s Thesis is false as a physical proposition. Stated another way, there would exist physical processes which could effectively compute nonrecursive functions. It would also mean that Natural Law cannot be expressed entirely in syntactical terms” (Rosen 1988, 533).

The crux of the argument, following Cleland is, then, that string processing is computable (Turing-computable) *while causality contains a non-computable (non-algorithmic) element* (Cleland 1993, 307 ff). So, if Cleland is right, Putnam is comparing ‘algorithmic causality’ (simulated string-processing) to ‘real causality’ (non-algorithmic) and is really comparing apples and oranges. At first sight this would not constitute an argument in favor of realism, for why would non-algorithmic causality be ‘more real’ than its simulated (algorithmic) counterpart? And is it, then, less theoretical? However, as will become clear soon, the aforementioned difference may have important consequences when it comes to the basic tenets of metaphysical realism.

### *Beyond Internal Realism?*

Internal Realism dismisses the idea of there being a unique correspondence relation between language and the world and that truth is radically non-epistemic. We have already shown earlier that it does not follow that the

notion of an intended or standard model for a theory as used by the metaphysical realist would therefore collapse into incoherence and unintelligibility.

Moreover, in indicating that the real concept of causality seems to be non-algorithmic rather than algorithmic we have already indicated a way out for the realist. That way out may hinge on Gödel's arguments for Platonism (realism) in mathematics.

In his *Gibbs Lecture* (Gödel 1995 [1951], 304–323) Gödel discusses some of the philosophical implications of his (incompleteness) theorem(s). The thesis he wants to defend in this lecture is in fact two-fold. He wants to show that mathematics is 'incompletable' or 'inexhaustible' by which he means that the resources of e.g. set theory allow a continuous 'growth' of sets which cannot be fully captured by any set of axioms and second that there will always be true mathematical statements that cannot be proven on the basis of given axioms. Gödel concludes from this that "either the human mind (even within the realm of pure mathematics) infinitely surpasses the powers of any finite machine, or else there exist absolutely unsolvable diophantine problems of the type specified (where the case that both terms of the disjunction are true is not excluded)" (Gödel 1995 [1951], 310). By an absolutely unsolvable problem Gödel means that it is undecidable, not just within some particular axiomatic system, but by any mathematical proof the human mind can conceive. In a second move Gödel, on the basis of these considerations tries to defend a realist or platonist position in the philosophy of mathematics which he defines as follows: "what is wrong, however, is that the meaning of the terms (that is the concepts they denote) is asserted to be something man-made and consisting merely in semantical conventions. The truth, I believe, is that these concepts (Mathematical concepts, R.V.) form an objective reality of their own, which we cannot create or change, but only perceive and describe" (Gödel 1995 [1951], 320). It is *mathematical intuition* which somehow brings us in contact with this world of mathematical concepts.

As Gödel puts it in his 1964 paper on Cantor's Continuum-problem: "despite their remoteness from sense experience, we do have something like a perception of the objects of set theory, as is seen from the fact that *the axioms force themselves upon us as being true*. I don't see any reason why we should have less confidence in this kind of perception, i.e. in mathematical intuition, than in sense perception, which induces us to build up physical theories and to expect that future sense perceptions will agree with them, and moreover that a question not decidable now has meaning and may be decided in the future" (Gödel 1990 [1964], 268).

The analogy between mathematics and natural science is very prominent in Gödel, who even considers that perception of physical bodies as the basis

of science is quite analogous to the perception of mathematical concepts: “it seems to me that the assumption of (set and mathematical concepts) is quite as legitimate as the assumption of physical bodies and there is quite as much reason to believe in their existence. They are in the same sense necessary to obtain a satisfactory system of mathematics as physical bodies are necessary for a satisfactory theory of our sense perception” (Gödel 1990 [1944], 128). When he says that the axioms of set theory *force themselves upon us as being true* it is precisely as a result of a perception-like contact with the reality of mathematical structures (sets, numbers) that their truth is ‘revealed’ to us.

Gödel maintains that mathematics and science (physics) are analogous and that both kinds of knowledge (physical and mathematical) are in a sense comparable and complementary: “The fact is that only laws of nature together with mathematics (or logic) have consequences verifiable by sense experience. It is, therefore, arbitrary to place all content in the laws of nature. What mathematics adds to the physical laws, it is true, are not any new properties of physical reality, but rather properties of the concepts referring to physical reality—to be more exact, of the concepts referring to combinations of things. But ... such properties are something quite as objective as properties of physical reality and even verifiable by sense experience under the hypothesis that certain laws of nature which can be confirmed independently of mathematics proper, hold good” (Gödel 1995 [1953/9], 348–349).

Gödel’s realism is to a large extent based on his *Incompleteness Theorems* (Gödel 1986 [1931], 144–195) which, among other things, claim that there are, for any suitably formalized system of elementary arithmetic, undecidable but true formulas, expressing true properties of natural numbers. It seems fair to say that Gödel’s theory is the best possible theory for at least part of mathematics in the sense that it delimits the axiomatizability of (certain parts of) mathematics.

Could it then still be true that one of the tenets of metaphysical realism—viz. *the thesis that even our best possible theory might still be false*—would still hold? Can the reality of numbers as described here be totally different from our best possible theories? I think not. But for Gödel, what was in a sense affirmed by the Incompleteness Theorems, is *that the reality of (natural) numbers cannot fully be captured by any axiomatic system*. In other words, the theorems say that (mathematical) reality is (necessarily) somehow different from any possible axiomatization, and that is why any axiomatization is necessarily incomplete. *So, the theorem says (or rather ‘shows’) that there is a fact of the matter about how numbers are*, independently of the way in which they are described or captured formally. The reality of numbers could not be totally different from what Gödel’s theorem says it is. Nevertheless, human beings are able to know the truth of the (undecidable) Gödel-sentence. So, strangely enough, while contradicting one of the tenets

of metaphysical realism, Gödel's theorem seems to somehow affirm its truth, in the sense that reality is different from any possible description (be that an axiomatization) which we may come to have of it. At the same time two further tenets of metaphysical realism are denied, viz. *that there is exactly one true and complete description of the way the world is*, and that *truth is radically non-epistemic*.

Indeed, though there may be exactly one true description of the way the world (of numbers) is, every possible axiomatization is clearly incomplete, while from the fact that human beings can intuit the truth of the Gödel-sentence and so have a capacity to capture or 'intuit' (at least in part) what numbers are really like it seems to follow that metaphysical realism in this sense does not require truth to be radically non-epistemic, since this kind of truth clearly belongs to the epistemic capacities of human beings. It also follows that *truth is non-algorithmic*. Given the further tenet of metaphysical realism that truth involves some sort of correspondence between thought signs or words and external things, this implies that reference is a non-algorithmic process and any theory which does not recognize this must fail.

Both causality and truth come out to be non-algorithmic and they are both connected to the concept of Reference. It remains to be seen whether this idea of non-algorithmicity is in general a good argument in favor of realism (for someone like Penrose, e.g. Penrose 1990, 1994, it certainly is). What it does show, however, is that internal realism may be overcome in the sense that one can still be a (metaphysical) realist and accept many of the criticisms leveled against it by internal realism.

Institute of Philosophy  
Center for Logic, Philosophy of Science and Philosophy of Language  
Catholic University of Leuven  
Belgium  
E-mail: Roger.Vergauwen@hiw.kuleuven.ac.be

#### REFERENCES

- Anderson, D. (1992) 'What is Realistic about Putnam's Internal Realism?', *Philosophical Topics* 20: 49–83.
- Brueckner, A. (1984) 'Putnam's Model-theoretic Argument Against Metaphysical Realism', *Analysis* 44: 134–140.
- Cleland, C. (1993) 'Is the Church-Turing Thesis true?', *Minds and Machines*, 3, 283–312.

- Glymour, C. (1982) 'Conceptual Scheming or Confessions of a Metaphysical Realist', *Synthese* 51: 169–180.
- Gödel, K. (1986 [1931]) 'On formally undecidable propositions of Principia Mathematica and Related Systems,' Repr. in: S. Feferman (a.o. eds.) (1986) *Kurt Gödel Collected Works, vol. 1*. New York: Oxford University Press, 144–195.
- Gödel, K. (1990 [1944]) 'Russell's Mathematical Logic'. Repr. in: S. Feferman (a.o., eds.) 1990, 119–141.
- Gödel, K. (1990 [1964]) 'What is Cantor's Continuum Problem?' Repr. in: S. Feferman (a.o. eds.), 1990, *Kurt Gödel. Collected Works, vol. 2*, New York: Oxford University Press, 264–270.
- Gödel, K. (1995 [1951]) 'Some basic theorems on the foundations of Mathematics and their implications' (the Gibbs Lecture). In: S. Feferman (a.o., eds.) (1995), *Kurt Gödel. Collected Works, vol. 3*. New York: Oxford University Press, 304–323.
- Gödel, K. (1995 [1953/9]) 'Is Mathematic Syntax of Language?' In: S. Feferman (a.o., eds.) 1995, 334–356.
- Hallett, M. (1994) 'Putnam and the Skolem Paradox', in: P. Clark and B. Hale (eds.) *Reading Putnam*, Oxford: Blackwell.
- Heller, M. (1988) 'Putnam, Reference and Realism', in: P. French *et alia* (eds.) *Midwest Studies in Philosophy* Vol. XII, Minneapolis: University of Minnesota Press, 113–127.
- Merrill, G. (1980): 'The Model-theoretic Argument against Realism', *Philosophy of Science* 47, 61–81.
- Naumann, R. (1993) *Das Realismusproblem in der Analytischen Philosophie*. München: Alber Verlag.
- Nola, R. (1988) 'Introduction', in: Nola R. (ed.) *Relativism and Realism in Science*, Dordrecht: Kluwer, 1–35.
- Penrose, R. (1989) *The Emperor's New Mind*, Oxford: Oxford University Press.
- Penrose, R. (1994) *Shadows of the Mind*, Oxford: Oxford University Press.
- Putnam, H. (1978) *Meaning and the Moral Sciences*, London: Routledge and Kegan Paul.
- Putnam, H. (1979) 'Reference and Understanding', in: A. Margalit (ed.) *Meaning and Use*, Dordrecht: Kluwer, 199–271.
- Putnam, H. (1980) 'Models and Reality', *Journal of Symbolic Logic* 45, 464–482.
- Putnam, H. (1981) *Reason, Truth and History*, Cambridge: Cambridge University Press.
- Putnam, H. (1982) 'Why there isn't a ready-made World', *Synthese* 51, 141–167.
- Putnam, H. (1987) *The many Faces of Realism*, La Salle (Ill.).

- Putnam, H. (1988) *Representation and Reality*, Cambridge MA: Harvard University Press.
- Putnam, H. (1990) *Realism with a Human Face*, Cambridge MA: Harvard University Press.
- Putnam, H. (1992) 'Reply to David Anderson', *Philosophical Topics* 20, 361–369.
- Putnam, H. (1994) 'Model-Theory and the 'Factuality' of Semantics'. In: J. Conant (ed.), *Words and Life: H. Putnam*, Cambridge MA: Harvard University Press, 351–376.
- Rosen, R. (1988) 'Effective Processes and Natural Law', in: Herken, R. (ed.) *The Universal Turing Machine: A Half-Century Survey*, Oxford: Oxford University Press, 523–537.
- Vergauwen, R. (1993) *A Metalogical Theory of Reference*, New York-London: University Press of America.
- Vergauwen, R. (1996) 'The view from Somewhere in Nowhere'. In: I. Douven and L. Horsten (eds.), *Realism in the Sciences*. Leuven: Leuven University Press, 130–145.