The Model-Theoretic Argument and the Search for Common Sense Realism Hillary Putnam (trans_from_Polish_to_en by ChatGPT-3.5)

"Philosophy of Science Year XIX, 2011, No. 1 (73)"

I was asked to characterize the "realism debate" and my current approach to these issues. I am pleased to take on this task because although most of my critics and supporters have recognized that these are complex and profound issues, I believe that critics still have not fully appreciated how complex and profound they are. In any case, in the first paragraph of the article, I will present a very condensed history of the evolution of my views and briefly discuss the kind of realism that I currently prefer. During this evolution, over the past twenty-plus years, I have abandoned "anti-realism" and "verificationist semantics", which I defended in essays such as Realism and Reason (1977) and Models and Reality (1980), as well as in my book Reason, Truth, and History (1981). However, this was not due to alleged flaws in my "theory model argument" against metaphysical realism. I am now convinced that although the theory model argument was not flawed, some of the assumptions on which it was based—assumptions that were almost universal in philosophy since Descartes—must be rejected. Yet these assumptions were never questioned even by my critics.

1. Title: The Model-Theoretic Argument and the Search for Common Sense Realism. So far, the text has only appeared in German translation: "Das modeltheoretische Argument und die Suche nach dem Realismus des Common sense", [in:] Realismus, ed. M. Willaschek, Paderborn 2000, Ferdinand Schöningh, pp. 125-142. Translation made with the consent and based on the original provided by the Author. The original was supplemented by Prof. Putnam with the addition of "What I Meant by 'Common Sense Realism'", written in 2009 specifically for the needs of the Polish translation [translator's note].

In the second paragraph, the subject of consideration (and defense) will be that part of the argument which my critics have almost universally deemed erroneous, namely, the step of "even more theory". However, in the concluding paragraph, I will make some general remarks concerning the entire issue.

A Brief History of My Views

In the early seventies, I began pondering what I eventually termed the "model-theoretic argument". I describe and analyze this argument in the second paragraph of this essay, but now let me present its final result. The argument is directed against the "metaphysical realist". According to my characterization of this opponent (whom I present as my previous incarnation), the metaphysical realist is convinced that truth entails correspondence with reality, but in a decidedly post-Tarskian sense of "correspondence". The post-Tarskian "metaphysical realist" adopts a relation called reference, occurring between every proper name.

2. Of course, since language is here construed as formalized in a standard manner, only proper names with denotations are permissible.

in any language and the entity it denotes; between every one-place predicate of the language and every member of the extension of that predicate; between every two-place predicate and every ordered pair from the extension of that predicate; etc.

3. Therefore, "reference" and the relation that Tarski called "satisfaction" are mutually definable.

Truth (as a predicate asserted of sentences of that language) is to be defined here in terms of reference, using Tarski's methods. Thus, strictly speaking, in this kind of "correspondence theory", truth is defined in terms of correspondence (between predicates and elements of their extension), and is not treated as something that consists of correspondence (between sentences and "reality", or sentences and "the world", or sentences and "states of affairs").

4. I emphasize this post-Tarskian character of the "correspondence theory of truth" that I discuss in my articles because it is routinely overlooked by critics. Another common mistake is portraying these articles as entirely "rejecting the correspondence theory of truth". In doing so, it is overlooked that I reject this theory only when "reference" is to be a relation independent of the theory, some metaphysical correspondence relation, occurring between any term of any language and the corresponding subset of a fixed totality of "all beings". However, I have never denied that one can speak of correspondence between a descriptive term and the entities it refers to when the notions of "correspondence" and "reference" are given meaning according to how we use them in a specific context. Metaphysical realists generally prejudge the issue (often by simply ignoring it) whether "reference" is inherently contextual or "absolute".

Furthermore, the metaphysical realist is convinced that objects from the extensions of different predicates are mind-independent, that correspondence is exactly one (and indeed applies to all languages and theories simultaneously), and that truth and the correspondence on which truth is based are entirely non-epistemic, meaning that even a theory which is optimal according to all reasonable epistemic standards could actually be false. The model-theoretic argument, which I will present in the next paragraph, convinced me for a long time that this view is inconsistent. Roughly speaking, the result of the argument is that if a theory is epistemically ideal, then in a certain sense there are guarantees that its predicates actually refer and its sentences are "truly true" — the idea that an epistemically ideal theory could be false is inconsistent.

Under the influence of this conclusion, I was compelled to seek a more "epistemic" conception of truth, a conception in which an epistemically ideal theory is guaranteed correctness.

In the book "Reason, Truth, and History" (1981), I proposed such a conception. I maintained that a proposition is true if and only if it would be rationally acceptable under ideal cognitive conditions;

5. See the last chapter of my book "Representation and Reality" (1988).

however, as I pointed out, "ideal cognitive conditions" cannot actually be achieved, just as one cannot actually find a "frictionless surface" in the world (Putnam, 1981, p. 55). Therefore, a proposition is true if and only if it would be rationally acceptable under sufficiently good cognitive conditions. I believe this aligns precisely with what Crispin Wright calls "superassertibility".

6. Wright, however, does not think so! (As he misinterprets Reason, Truth and History). See Wright 1992, pp. 44ff.

I began to consider this identification of truth with superassertibility somewhat incredible myself, but at that time (in the late seventies and early eighties), I did not know how — given the failure of metaphysical realism — to make sense of the concept of truth in any other way.

"Realism and Reason" (1977)

7. That was my Presidential Address at the meeting of the Eastern Division of the American Philosophical Association in 1976. For the first time, I presented my model-theoretic argument and renounced metaphysical realism. It was included in my book "Meaning and the Moral Sciences"(1978).

contained a different, entirely different argument against metaphysical realism, the argument from the phenomenon of equally correct theories (since "The Many Faces of Realism" (1987) I have used the term "conceptual relativity" for this phenomenon), which at first glance appear to be inconsistent with each other, but both are "correct". Pure mathematics is full of examples (for instance, set theory, in which functions are identified with ordered pairs of sets, and set theory in which sets are identified with "characteristic functions", are inconsistent with each other in such a sense that their combination is contradictory, yet no logician assumes that there is an "objective fact" (fact of the matter) determining which one is "truly correct"). There are also examples in the theory of space and time (the theory in which points are individuals and the theory in which points are convergent sets of spheres), and even in physics.

However, both arguments against metaphysical realism are aimed at entirely different components of this position. The model-theoretic argument was directed against the claim that even an epistemically ideal theory could be false, and hence its ultimate result was verificationist. The argument from the phenomenon of conceptual relativity was aimed at the assertion that (apart from the trivial case where a term is used in different senses in two "inconsistent" theories)

8. Sometimes it is argued—and not only by metaphysical realists—that all cases of theories that are correct but apparently inconsistent with each other turn out to be trivial in the following way: a certain term has a different meaning(in the ordinary sense of "different meaning"). I respond to this thesis in Putnam1994b.

the very nature of what we are talking about must determine exactly one truth on the matter. Using more precise language, the metaphysical realist assumes that the world (here I limit myself to empirical claims) must determine (1) the fixedness of "individual objects" and (2) the fixedness of "properties and relations" of those individual objects (and correspondingly for higher types). This means that if we imagine a language with one proper name for each individual and one predicate for each property and relation, etc. (up to a certain fixed type in the known Russellian hierarchy), then in this "infinitary" language there will be exactly one true theory of the world.

On the other hand, accepting conceptual relativity as a real and non-trivial (and non-trivializable) phenomenon implies acceptance that the world does not determine some privileged use even of such "logical terms" as "object", "property", or "exists". The whole quest for an answer to the question of the "real ontology" of the world is fundamentally misconceived. Notice that accepting this does not require acceptance of verificationism, which is the result of my "model-theoretic argument". According to my current position, conceptual relativity should be recognized as a real and non-trivial phenomenon, but the assumptions on which the model-theoretic argument was based must be rejected.

Since I had some doubts about identifying truth with superassertibility, in my Carus Lectures (Putnam, 1987), I limited my attack on metaphysical realism to defending conceptual relativity, while remaining silent about the model-theoretic argument and the identification of truth with "idealized rational acceptability". However, the final chapter of "Representation and Reality" (1988) still defended this identification, albeit emphasizing its realist aspect.

Around this time, a conference on my philosophy took place in Madrid, where Cesar Gomez, one of the world's leading mathematical physicists and a proficient amateur philosopher, was present. During the

discussion of my model-theoretic argument at this conference, Gomez made a significant observation. He suggested that my model-theoretic argument presupposes a Cartesian view of perception, according to which our perceptual contents are objects "inside" us, objects in the "theater of the mind"; he also said that "Sense and Sensibilia" by John Austin might contain a way to unravel this whole problem. At the time, I disagreed, arguing that "even if Austin were right, it wouldn't help in this matter", but after a few years, I came to the conclusion that Gomez was entirely right! (The fact that I was engaged with the philosophy of William James, who was a staunch advocate of "direct realism", played a significant role in this—see Putnam, 1996).

To grasp the significance of direct realism for the theory-model argument, we need to consider the concept of an "epistemically ideal" theory that this argument employs. An "epistemically ideal" theory is supposed to not only meet theoretical requirements such as optimal simplicity, elegance, coherence with previous doctrines, credibility, etc., but also maximally fulfill our "operational requirements". But what exactly are "operational requirements"?

If "operational requirement" is a requirement in the form illustrated by the following example: "Usually, when theory T implies the sentence 'Hilary Putnam sees a cow,' Hilary Putnam will seem to himself to see a cow", then accepting the assumption that these requirements can be fulfilled even though the entities involved are "false in reality" in a realist metaphysical sense would entail acknowledging that our perceptions are largely erroneous—e.g., we are "brains in a vat" or something similar. However, one could ask someone who accepts such an assumption: "How is it that our terms are understood to refer to things like cows if what we see are not cows?" To demonstrate the inconsistency of this scenario, it's sufficient to assume that causal requirements are imposed on reference.

9. The argument against the scenario of "brains in a vat", based on the idea that causal requirements are imposed on reference, is presented in Chapter 1 of the book: "Putnam, 1981".

On the other hand, arguing (as in the "argument from permutation of individuals" in "Reason, Truth and History") that although we usually have cows before us when we utter sentences like "I see a cow right now", the mind still lacks a way to distinguish cows from all other things among the causal lineage of "cow sense data" (or "perceptual representations") leading to the acceptance of this sentence entirely relies on the following picture: mental operations start with something like "sense data" (internal representations); these representations are connected to what is "outside the mind" not through any cognitive processes (all cognitive processes occur "within" the mind), but at best through causal agentive interaction. But this picture is precisely what direct realism—or rather, common-sense realism—attacks. For the common-sense realist, there is no problem with how the human mind can isolate the referent of a term or the class of referents of a term, at least if those referents are observable. Because in the light of common-sense realism, the mind perceives both its terms and their intended referents simultaneously, it can simply say: "What's the problem?" According to common-sense realism, the relationship between our terms and their referents is inherently cognitive, not just "causal".

As a result of all these reflections, delivering the Dewey Lectures in 1994 (Putnam, 1994a), I embarked on the path that Cesar Gomez had persuaded me to take a decade earlier. I argued for "commonsense realism" and against the verificationism that my theory-model argument had led me to. I also argued that there is no hope for defending commonsense realism if we insist on the "interface" conception of perception (and likewise, concepts) inherited from Cartesianism. I believe my critics make the same mistake that I made, sticking to the idea that we can talk about "operational requirements" and "theoretical requirements" (like logical positivists!) while simultaneously defending realism.

The defense of the "yet more theory"

The main erroneous point of the theory-model argument was supposed to be the move called "just more theory move". One of the few commentators who do not consider it erroneous is Igor Douven (1999). Although, as it turns out, there are moments where his reconstruction is off the mark, he perceives the essence of the problem with exceptional clarity at key points. In this paragraph, I will present and examine his reconstruction of this argument.

Douven begins by noting (1999, p. 479) that at the heart of the doctrine of metaphysical realism, as he presents it, lies a thesis he calls methodological fallibilism (FM), which is precisely the doctrine according to which "even an epistemically ideal theory can be far from the truth". He then argues (Douven 1999, p. 479) that another thesis of the metaphysical realism under consideration is something called semantic naturalism. The essence of this thesis is as follows: "semantics is an empirical science, just like any other" (Devitt 1996, p. 48). Douven goes on to say:

> One consequence of this view is that reference theories [...] are empirical theories and therefore must be accepted or rejected on the basis of exactly the same criteria that decide the choice of theories within what we traditionally regard as natural sciences (Douven 1999, p. 479).

I will try to explain (what Douven may have been the first commentator to notice) that critics of the theory-model argument have failed to meet the requirement of semantic naturalism — a requirement that they themselves accept and praise the merits of!

In fact, Douven could have presented his line of argumentation more vividly (which I will discuss shortly) had he not made a certain mistake in an earlier article, which was not corrected in the later article currently under consideration, although this earlier article is prima facie irreconcilable with the conclusions of the later one. In this earlier article, Douven intends to show that "it would be a miracle if there were a world like ours, but reference was radically undetermined in it".

Douven's argument in the earlier article

10. "Could Reference Be Indeterminate?" was a paper presented at the Eastern Division Meeting of the American Philosophical Association in 1994, but as far as I know, it was never published [In private correspondence, Douven confirmed that it was indeed not published because it was essentially a preliminary version of a later article — trans. note].

referred to a version of the theory-model argument, which I presented in "Reason, Truth and History", specifically to my formulation of the argument aimed at showing that the fact that our evolution proceeds successfully (to the extent that "we are still here") does not lead to the rejection of the radical undetermination of reference, which according to the theory-model argument threatens metaphysical realism. Douven correctly reports that I maintain that

> If a sufficient number of our dispositional beliefs (beliefs of the form "if you want x, do y") turn out to be true in the case of a nonstandard interpretation [...], we will certainly be successful, certainly survive [...], and have offspring (Putnam 1981, p. 40).

Douven's objection is short and precise: "Let's assume I speak language R, and my doppelgänger on Twin Earth speaks language R*. Of course, every true/false dispositional belief held by me will be true/false in both R and R*; thus, my doppelgänger and I will have the same number of true and false dispositional beliefs. But, for example, the dispositional belief 'If you want to survive, stay away from tigers' may have the following

content in R*: 'If you desire companionship, buy a pet.' Both are true, but it's hard to maintain that they have the same 'survival value'".

This accusation is entirely mistaken because Douven (in writing his earlier article) completely confused reference with narrow content. The argument of the theory model relies on the assumption (unquestioned by my critics, including Douven) that every mental state has what Jaegwon Kim termed an "inner core" (Kim 1993, pp. 189-190). In the case of perception, the inner core consists of the sensory data of having that perception (what these data must be if the theory is to be "epistemically ideal" is determined by "operational requirements"), while in the case of directive thought, such as "If you want to survive, stay away from tigers", the "inner core" would involve nourishing the thought whose narrow content directs us as follows: "If we want to survive, let's act in a way that makes it seem to us [that such are our sensory data] that we are staying away from tigers". More precisely, all thoughts remain unchanged both on Earth and Twin Earth, both syntactically and in every other respect, including their narrow content, but what the words with these narrow contents refer to is different in R and R*.

Douven assumed that R* attributes not only a different wide content but also a different narrow content. But that was precisely mistaken. And because it is mistaken, nothing I do or observe can determine whether, from the Divine Point of View, the "metaphysically privileged" reference relation for my language is R or R*.

Having understood this, it should not, I believe, surprise us with the correct conclusion Douven arrived at in his new article, namely, that the assertion according to which only one reference relation ("the causal relation") is somehow metaphysically privileged cannot be an empirical assertion, a theory that "must be accepted or rejected on exactly the same criteria that determine the choice of theories within what is traditionally considered the natural sciences". Metaphysical realists must violate their own semantic naturalism. However, I am beginning to get ahead of myself.

Let's now return to Douven's later article. The author begins with a brief and clear summary of the theory model argument:

> The core argument of Putnam's theory model is indeed quite simple and entirely plausible. Let's assume that T1 is an epistemically ideal theory (i.e., a theory that has every virtue except, perhaps, truth). Being ideal, it will, among other things, be non-contradictory. From an elementary theory of models, it will then follow that T1 has models of every infinite cardinality. Let's select from them some model M with the same cardinality as the world and establish the satisfaction relation (SAT) by a mutually one-to-one mapping of M onto the world. If the language of theory T1 is interpreted in accordance with SAT, T1 will turn out to be true about the world, i.e., T1 will be true-according-to-SAT or TRUE(SAT) (Douven 1999, pp. 480-481).

However, in the next paragraph, Douven makes a significant mistake when he writes: "It should be noted that, besides non-contradiction, the ideality of T1 plays no role in this argument. Thus, it is essentially an argument concerning non-contradictory theories". This peculiar statement is defended in the following way. First, Douven presents the following objection on behalf of the metaphysical realist:

> The realist acknowledges that there is an abundance of correspondence relations between our vocabulary and reality, but emphasizes that only one of them is the intended one, only one of them determines reference. So, what is the point of this argument? (Douven 1999, p. 481).

The key part of Douven's response to this question (at this stage of his argument—since he provides a different, more interesting, response later) is as follows:

> "[...] Putnam's point, roughly speaking, is that for the metaphysical realist, there is no way to distinguish between truth according to some interpretation and truth simpliciter. [...] The realist has only two options for giving substantive content to the notion of intentionality:

11. "Intentionality" obviously relates to the term "intended" in the phrase "intended interpretation".

> either define "intentionality" with respect to interpretation, so that it means either being a member of the class of F-interpretations (i.e., interpretations that make the theory true), or being a member of the class of non-F-interpretations. Since we obviously want our theories to be true, the first option seems to be the better way to define it. However, it then turns out that every non-contradictory empirical theory must be true, contrary to MM [methodological fallibilism] (Douven 1999, pp. 481-482)".

However, the assertion that "every consistent empirical theory must be true" was not part of what I was arguing for! To see what is wrong with Douven's argument presented on my behalf, let's consider a "theory" consisting of just one sentence: "I am seeing a cow in front of me".

This sentence is obviously consistent and therefore true in light of the multitude of interpretations (for example, an interpretation according to which in my case it means "I speak English"). However (at least at the moment), it doesn't seem to me that I am seeing a cow. I don't have "cow sensory data in my field of view". And although it is true that "I want my theories to be true", I want more than just that. Without assuming something that even the most fervent anti-realists or the most fervent metaphysical realists would not be able to accept, I can say that I want my theories to meet the requirement that they generally do not imply that I see something when I myself do not seem to see anything like that. I say "generally" because we can allow for times when I don't know what I see. However, accepting theories in their entirety just because they are consistent, even though their observational predictions seem false, would be a complete abandonment of empirical methodology. That's why my argument, with all due respect to Douven, is not "essentially an argument about consistent theories", but at least about theories that are consistent and (to a reasonable extent) meet our operational requirements, theories implying predictions that seem to be true.

However, that's not enough because (there are countless examples) sometimes we reject a theory that doesn't imply false predictions but also doesn't imply any true predictions that wouldn't be implied by a simpler, more elegant, or stronger theory contradicting it. Such a theory violates what I termed in "Realism and Reason" (1977) the "theoretical requirement". A theory that doesn't violate any operational or theoretical requirement can be called "epistemically ideal". My argument was precisely about epistemically ideal theories. It is entirely mistaken to assume, as Douven did, that I argued that

> there is no feature or set of features F* other than F that differentiates all existing correspondence relations between the words of the language of theory and things or sets of things in the world (Douven 1999, 482).

My argument was based on the fact that although some correspondence relations may be excluded—because theories made true by them turn out to be epistemically unacceptable, even though they are consistent—a metaphysical realist cannot point to a non-epistemic feature that differentiates all existing correspondence relations between the words of the language of theory and things or sets of things in the world, which—to use Douven's phrase—would "meet the standards of a metaphysical realist".

Fortunately, the argument Douven presents (on my behalf) in support of the latter claim I can accept, despite the error I just pointed out (assuming that epistemic features F* are irrelevant to my argument). He formulates

this claim (it's his claim "3") as follows: "(3) There is no such feature F* that would meet the standards of a metaphysical realist" (Douven 1999, p. 482).

Let's continue following Douven's reconstruction. He writes:

> (3) Many realistic opponents of Putnam were puzzled. Typically naturalistically inclined realists adopt some version of the causal theory of reference (CTR), according to which reference is constituted causally.

12. The critique of the concept of "causality" assumed by such theories is contained in my considerations in the book "Renewing Philosophy" regarding Fedorov's version of the causal theory (Putnam 1992). The fact that what we refer to as the "cause" of an event depends on our interests, and therefore is not something pre-established and independent of our intentionality, is completely ignored by proponents of the so-called causal theory of reference!

> In other words, they argue that there is an acceptable property F*—causality—sufficient to distinguish reference from the tangle of word/world relations. The realist response to Putnam's model-theoretic argument is therefore most likely this: since there is no guarantee that SPEL corresponds to the causal relations constituting reference, there is no guarantee that T1 [epistemically ideal theory] is true (rather than simply being TRUE(SPEL)) (Douven 1999, p.483).

As Douven explains, I anticipated this response and quickly dealt with it (probably too briefly, as seen in hindsight), by writing:

> Notice that the "causal" theory of reference is not (would not be) of any help here, because in the picture of metaphysical realism, how an expression can unambiguously refer to something as "causing causally" is just as puzzling as how the word "cat" can refer (Putnam 1977, p. 486; reprinted, p. 126).

Douven writes that this remark is typically interpreted in the following way: assuming that the CTR is indeed an empirical theory (as required by semantic naturalism) and that it does not violate any empirical requirements (because if it did, it would be subject to empirical or methodological disconfirmation), the CTR is part of a epistemically ideal theory as a whole.

> However, now the model-theoretic argument can, due to its generality, be repeated [for the entire theory, including CTR]. It is entirely obvious—according to the subsequent stage of this interpreted argument—that if we assume the problem faced by the realist has already been solved for that part of language in which the CTR is formulated, then this theory can probably be adopted to solve that problem for T1 (just like any other theory to which the model-theoretic argument applies). However, "in the picture of metaphysical realism, how an expression can unambiguously refer to 'causing causally' is just as puzzling as how the word 'cat' can refer". In other words, the CTR is part of the problem, not the sought-after solution. Assuming otherwise would prejudge the matter (Douven 1999, pp. 483-484).

This interpreted argument has been widely criticized as being "incorrect and unacceptable"(Hale, Wright1997, pp.440nn.) and as "simply prejudging the matter to the disadvantage of the realist"(Devitt1991, p.227). However, as Douven points out, in such an interpretation, the fact that the CTR is assumed to be an empirical theory plays no significant role. If we add a purely metaphysical(i. e., empirically untestable) CTR to an ideal empirical theory, or any other untestable theory of reference, this argument will work just as well.

According to Douven's reconstruction, I compel the metaphysical realist to confront a problem that crucially depends on their ambition to be a naturalist in semantics (when writing "Reason, Truth and History", I argued that in this form the problem does not arise for something I called a "magical theory of reference"—which is not, of course, an argument for adopting a magical theory!). Douven develops this in an extraordinarily imaginative way, relying on his mistaken belief that the model-theoretic argument can be applied to any consistent theory, but his idea is easily freed from this defect. Here is my own corrected version of Douven's argument:

Let's assume that the CTR is indeed an empirical theory. There are two possibilities: either (1) the CTR is deduced from the non-semantic part of the ideal theory T1, or (2) the CTR is logically independent of the non-semantic part of theory T1. If we assume (1), then the metaphysical realist owes us an explanation of how a semantic theory can be deduced from non-semantic empirical facts. (They would have to provide an analytical definition of "reference" in non-semantic terms!) As far as I know, no metaphysical realist accepts that this is possible. Therefore, we are left with (2).

Because the CTR is a theory with empirical content, independent of the non-semantic part of the ideal science, there exists a logically possible world in which the non-semantic part of the ideal science is true, while the CTR is false.

13. The metaphysical realist will not be helped by the claim that even if such a world were logically possible, it is not "metaphysically possible". This notion, I believe, is hopelessly problematic. (For a detailed critique, see Putnam 1990).

If such a world is actual, then—since no one has suggested any other naturalistic requirement imposed on the set of permissible reference relations other than "causal relations", and causal relations do not single out a privileged reference relation—if the CTR is false in such a world (as argued by Douven), then the model-theoretic argument is correct: any reference relation that makes the ideal theory true is a permissible reference relation, and none of the many permissible reference relations is privileged. Because the entire ideal theory, including the expressions expressing the CTR, is true for all these reference relations, it follows that in the world we have just imagined, the words we use to express the CTR are true (true within all permissible reference relations), even though the CTR is not "actually" true. Therefore, it turns out that even in worlds where the CTR is false, we must still accept the words expressing the CTR as true, as long as the non-semantic part of the ideal theory is not empirically refuted! In short (assuming that the non-semantic part of our theory is settled), the CTR violates MM (methodological fallibilism). In a sense, the situation is similar to the one I presented in my argument of brains in a vat in the first chapter of "Reason, Truth and History" (Putnam 1981): the words expressing the CTR are semantically unstable because what they refer to depends on the world we are in. However, unlike the words "We are brains in a vat", which changed reference in such a way that they always turned out to be false, regardless of the world in which the person uttering them was, the CTR changes the conditions of reference in such a way that it remains true even when metaphysical realism is false.

That's an incredibly clever argument. However, I don't intend to claim that it was my argument in "Realism and Reason". I presented a much simpler argument there, leading to the same conclusion: that the question of which reference relation is "intended" or "metaphysically privileged", as well as the question of whether a given unique reference relation is "intended" or "metaphysically privileged", are not empirical questions, and therefore the metaphysical realist cannot be simultaneously a semantic naturalist. It's this simpler argument that Douven mistakenly rejected in his earlier article. Regardless of whether one believes there is one intended reference relation or that all SPEL relations making T1 true are equally permissible and none is privileged, as I argued in "Realism and Reason" and detailed in "Reason, Truth and History", it has no bearing on the

predictions that will be made about anything. These differing metaphysical views are empirically indistinguishable. This means that the CTR is a "semantics" that does not adhere to the maxim that "semantics is an empirical science, just like any other".

Of course, the metaphysical realist can respond to Douven's argument by claiming that even if the CTR is irrefutable, as long as we regard the non-semantic part of our theory of the world T as settled, it still undergoes revision in a broader sense, as there are certain parts of T—certain non-semantic beliefs we currently hold and also consider part of the ideal theory T1—such that their empirical refutation would lead to abandoning the CTR. The metaphysical realist can reasonably argue that the requirement for the CTR to be falsifiable even when the non-semantic part of T1 is taken as settled is an untenable interpretation of methodological fallibilism (because in reality, it requires that every logically independent part of an empirical theory be independently testable). It is sufficient if there are conceivable empirical discoveries that would require us to abandon the CTR, even though they would also require us to change the non-semantic part of T1.

If the metaphysical realist takes this path, they are forced to argue that there are possible non-semantic discoveries that could refute the CTR. However, as far as I know, no metaphysical realist has suggested or signaled anything of the sort, nor have they indicated how such an experiment would look like!

And one more observation. A philosopher who accepts the model-theoretic argument and thus is convinced that there is no single "intended" reference relation may also adopt the meta-metalanguage MML, in which they will say about a given object language L that a certain interpretation of L (in the metalanguage ML) is the "intended interpretation of L". (They may even accept the CTR). However, in another metalanguage MMML, they will then state that the meta-metalanguage MML itself has "more than one intended interpretation". As I wrote in "Realism and Reason": "the question of whether a theory has a unique interpretation does not have absolute sense" (Putnam 1977, p. 494; reprint, p. 136). "Having a unique intended interpretation" turns out to be a property relative to the theory. (Assuming, which I myself no longer accept, that the conclusion of the model-theoretic argument is valid).

FINAL REMARKS

I emphasized that we are compelled to accept the theory-model argument, along with its verificationist view that an epistemically ideal theory cannot fail to be true only when we accept the assumptions on which the argument is based. These assumptions include not only the idea of "operational requirements", with its implicit reference to sensory data, but also what I myself have termed the "verificationist concept of understanding". Is it surprising that by adopting so much from logical positivism, one is later forced to accept positivistic conclusions? I suspect that the very critics who regarded the conclusion of the theory-model argument as a surrender to positivism failed to criticize the positivistic assumptions of the argument because they accept those assumptions—in fact, these assumptions seem to be imposed by the Cartesianism associated with materialism, which nowadays often passes for "cognitive science". And yet, as Cesar Gomez first pointed out, it is precisely at this point—where the theory-model argument should be criticized.

In my Dewey Lectures (Putnam 1994a), I presented a way to avoid the entire argument by abandoning the "interface conception of perception" and the "interface conception of conceptualization", which we have held onto for three centuries. Stating this does not carry an "anti-scientific" connotation. As I wrote in the Dewey Lectures:

> In our common-sense realism concerning both perception and conceptualization, there is nothing that would be "unscientific" in the sense of erecting barriers to serious attempts to provide better models,

> both neurological and computational, of the brain processes on which our perceptual and conceptual abilities depend — processes about which we still know so little. Furthermore, it is a huge mistake to equate true science with Cartesianism associated with materialism, which for three centuries has been trying to masquerade as science (Putnam 1994a, p. 494; reprinted, p. 48).

What do I understand by 'common-sense realism'?

14. I make use of two or three paragraphs from my intellectual autobiography, contained in: The Philosophy of Hilary Putnam, ed. R. E. Auxier, Chicago, Open Court (in preparation).

Many philosophers have acknowledged that they are confused about what I actually meant when talking about a return to "commonsense realism" in Dewey's lectures. In short, "commonsense realism", in my sense of the term, contains a negative element, namely the rejection of the view that truth cannot exceed verifiability,

15. See my exchange with Crispin Wright (Wright 2000; Putnam 2001) on this topic.

and two positive elements: a return (as much as possible) to "naive realism" concerning perception and a disquotational conception of truth, similar to what I perceive in Wittgenstein. It differs from what I once called "metaphysical realism" because it rejects what I consider fantasizing about one ultimately true and complete ontology, but, of course, it is in its own way both metaphysical and realistic. Here is a brief explanation of these two positive elements.

Naive realism

In my book "The Threefold Cord: Mind, Body, and World" (Putnam 1999), partially influenced by John McDowell's views in "Mind and World" (McDowell 1994), I focused on the need to avoid thinking of either perceptual experiences or thoughts as "screens" between us and the world. Today, part of the position I argued for then, namely the claim that during normal veridical perception of objects in our environment (e.g., a bunch of red roses), the objects of our awareness are not mental objects or qualia but the external objects themselves (the red roses), has become orthodox among philosophers of mind. The picture of our minds as gazing at an "inner movie screen" is decidedly unfashionable. Almost all philosophers of perception want to return, at least partially, to "naive realism" (a term that currently seems fashionable to denote what I called "natural realism" in the mentioned book). However, the return to naive realism is more complicated than I presented it in "The Threefold Cord". One sign of the complexity of the issues involved is that alongside the positions I discussed in my Dewey and Royce Lectures (in two series of lectures later published in "The Threefold Cord"), such as the traditional sense-datum theory and "alternativism" (e.g., John McDowell), there are now also "phenomenalists" (e.g., Ned Block) and "representationalists" (e.g., Michael Tye), and each of these schools has its factions that differ on important issues. Moreover, most (though not all) of these philosophers, including myself, are "broad functionalists", meaning they understand our successful perceptions as the result of world-involving functional states, although not generally computational states. (This last fact is, of course, related to the abandonment of the picture of perception as scanning an internal "movie screen" and to the fact that all these philosophers want to do justice to what is right in "naive realism").

Disquotationalism

However, there is another element of my "commonsense realism". The second element on the path to the philosophical defense of "natural realism" that I presented in the Dewey Lectures is the "disquotational

conception of truth". To explain this, I will briefly describe a certain family of (twentieth-century) truth theories. This family is commonly referred to using the term I just used: "disquotational" conceptions of truth. In the Dewey Lectures, I argued that there are at least three very different versions of "disquotationalism": Frege's version; the "deflationary" version, represented by Carnap (1949); and the version defended by me (and attributed to late Wittgenstein).

What these conceptions have in common is the emphasis they place on what Michael Dummett calls the Fregean "principle of equivalence" (which is an obvious anticipation of Tarski's "Convention T") (Dummett 1973 and 1981; Tarski 1956). This principle asserts that to assert or judge that something is true, such as that snow is white, or that murder is wrong, or that two is the only even prime number (or any such example), is equivalent to judging that—respectively—snow is white, or murder is wrong, or two is the only even prime number. If one accepts that judging that something is true, such as that snow is white (etc.), is the same as judging that the sentence (in what Tarski called the "object language") "Snow is white" is true (this being one of the issues that distinguishes Tarski from Frege), then the principle of equivalence can be expressed by writing (according to Tarski):

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

When I say that all versions of disquotationalism "emphasize" some version of the principle of equivalence, I mean the following things:

(1) Disquotationalists maintain that when the word "true" is used to assert a proposition S that is directly given, it is eliminable. For example, in some places, Frege seems to maintain that a meta-statement such as "It is true that snow is white" and the corresponding object-level statement "Snow is white" express one and the same judgment. If we were to use the word "true" only in sentences of the form "S is true", where S is a quotation of a sentence, or in sentences of the form "It is true that p", where p is a sentence, then the word would be unnecessary.

(2) According to disquotationalists, the reason we need the word "true" and its synonyms in our languages is that we need to be able to use the word "true" in sentences of the form "x is true", where "x" is a quantification variable, and not—let's say—a sentence in quotation marks. For example, if I say: "At least one of the sentences John wrote on page 12 is true" [in predicate calculus notation: (∃x)(x is written by John on page 12 and x is true)], I can "eliminate" the word "true" if and only if I know which sentences John wrote on page 12. For instance, if I know that the only sentences John wrote on page 12 are: "Snow is white", "Murder is wrong", and "Two is the only even prime number", then I know that this statement has the same logical value as: "John wrote 'Snow is white' on page 12 and snow is white, or John wrote 'Murder is wrong' on page 12 and murder is wrong, or John wrote 'Two is the only even prime number' on page 12 and two is the only even prime number, and these are the only sentences John wrote on page 12", where this long sentence does not contain the word "true". However, if I do not know which sentences John wrote, then I would not know how to construct a materially equivalent sentence that does not contain the word "true".

In summary, all disquotationalists agree that the predicate with the logical property of truth ("disquotational property") is necessary for logical reasons, not descriptive ones.

The differences between various versions of disquotationalism are, roughly speaking, as follows. For Frege, the word "true" is a predicate asserted of "thoughts" (Gedanken), and thoughts do not consist of words. Whether they are Platonic objects or extend beyond the Platonic-non-Platonic dichotomy is a matter of dispute among Frege scholars; in any case, there is a metaphysics of "thoughts" presupposed by Frege's version. Moreover, thoughts about empirical realities are about them by their very nature; for Frege, thoughts are not images

painted with "mental paint" that somehow "correspond" to the world; they state something about the world. For Tarski, on the contrary, the word "true" is directly predicated of sentences, and sentences are ordinary sequences of symbols on paper. (Tarski does say that he assumes these signs have "concrete meanings", but this requirement is not part of the definition of truth, which makes truth simply a property of signs, not a property they have when they have (undefined) "concrete meanings"

- The passage from Tarski's dissertation to which Putnam refers reads: "Needless to say, we are not concerned here at all with 'formal' languages and sciences in a certain specific sense of the term, namely those sciences in which no intuitive sense is attributed to the signs and expressions occurring in them; with respect to such sciences, the problem posed here loses all sense of being and ceases to be simply understandable. We always attribute quite concrete and understandable meanings to the signs occurring in these languages, which are the subject of the following considerations" (Tarski 1956, p. 166-167; original Polish, p. 17; reprint, p. 33) [translator's note].

). Tarski's formulated "definitions of truth", contrary to what is often claimed, neither assume nor lead to a "correspondence" concept of truth. For Carnap, who fully accepted Tarski's theory, we understand a sentence, say "Snow is white", by understanding its verification procedure, and we understand the corresponding meta-sentence "The sentence 'snow is white' is true" by knowing that it has exactly the same verification procedure. (This is precisely the "deflationary" concept I rejected in the Dewey Lectures). For Wittgenstein (and myself), "true" is a predicate asserted of sentences used in certain ways—meaning it is asserted of objects that are neither purely syntactic (like Tarski's "sentences") nor independent of the world of use of syntactic objects in a particular linguistic community. Similar to Frege's conception, this leads to the understanding that a statement such as "Snow is white" is not an image painted with mental paint that somehow corresponds to the whiteness of snow; the use of this sentence involves snow and whiteness. If there is a "correspondence relation" here, it is an internal relation, not an external or contingent one.

Translated by Krzysztof Czerniawski and Tadeusz Szubka.