

13 The model-theoretic argument: from skepticism to a new understanding

Gila Sher

Two well-known arguments by Putnam, the one skeptical (Putnam 1980), the other anti-skeptical (Putnam 1981b), are the “model-theoretic argument” (henceforth, MT) and “brains-in-a-vat argument” (henceforth, BIV).¹ The MT argument shows, according to Putnam, that (i) we cannot theoretically determine the reference of our words, and that, as a result, (ii) we must renounce the correspondence theory of truth and robust realism. The BIV argument shows, Putnam says, that (iii) we cannot truly believe that we are BIVs, and that (iv) Cartesian skepticism is thus undermined.

Three questions naturally arise: (a) Is there a conflict between Putnam’s skeptical and anti-skeptical conclusions? In particular: Given that the key to (iii) is Putnam’s claim that we, unlike BIVs, have referential access to the world, is there a conflict between his claims about referential access in the two arguments? (b) Is Putnam’s skeptical conclusion concerning truth and realism warranted? (c) What is the philosophical significance of the BIV thought experiment?

The answer to the first, more general, part of the first question appears to be negative. If we understand BIV as arguing against an especially extreme form of skepticism, Cartesian or Evil-Demon skepticism, and MT as arguing for a weaker form of skepticism, correspondence-truth skepticism or robust-realism skepticism, there need not be a general conflict between them. However, some ideas developed in recent articles about Putnam’s BIV (see e.g. Tymoczko 1989a, b; A. W. Moore 2011) suggest to me that on some level there might be a conflict between the two arguments. The answer to the second, more specific, part of the first question depends on whether the referential access denied by MT is of the same kind as that affirmed by BIV. But the first question is not one I will focus on in this chapter.

¹ I will use “MT” and “BIV” as abbreviations of other locutions involving the expressions “model theoretic” and “brains in a vat” as well. It will be clear from the context what I mean.

The questions I will focus on are the second and the third questions. I will offer new challenges to Putnam's skepticism with respect to robust truth and realism, and I will raise a new philosophical question about BIVs (and us).

The chapter is divided into four sections. In Section 13.1 I will challenge Putnam's MT-skepticism on the ground that it is based on a mistaken understanding of Tarskian models. In Section 13.2 I will discuss Putnam's skeptical permutation-argument. In Section 13.3 I will question Putnam's understanding of robust correspondence and realism. And in Section 13.4 I will identify a new BIV challenge, significant for understanding our cognitive access to reality.

Although Putnam's point in his MT and BIV arguments is, on the surface, semantic, I believe his deeper point is epistemic, as my own standpoint in this chapter will be. By characterizing my standpoint as epistemic, however, I do not mean what Putnam means when he talks about "an epistemic approach" to truth or realism, which he contrasts with a "correspondence" approach. On the contrary. For me, to approach truth epistemically is to ask (i) whether genuine knowledge requires something like a correspondence standard of truth, and (ii) whether a correspondence standard of truth sufficient for genuine knowledge is possible. And it is the relevance of semantic considerations to these two questions that interests me here.

In spite of my difference with Putnam on these two questions, I have much in common with him. Like Putnam, my approach to knowledge in general, and to philosophical knowledge in particular, is holistic. This means, among other things, that I do not draw sharp boundaries between epistemology, metaphysics, and the philosophy of language. Thus I see the failure of reference as posing a serious problem to knowledge in a rather straightforward way: if there is no fact of the matter about what our theories say about the world, then they cannot provide us definite knowledge about the world. Other philosophical positions I share with Putnam include rejection of Platonism, rejection of narrow naturalism, rejection of pure apriorism, and rejection of "magic" (i.e. appeal to magical or supernatural forces in philosophical explanation). Like Putnam, too, my starting point is "common-sense realism" (Putnam 1994b: 303), where by this I understand "the idea that thought and language can represent parts of the world which are not parts of thought and language" (299). I also share Putnam's view that "[t]here are many ways of describing things, some better and some worse and some equally good but simply different, but none which is Nature's own way" (302). And I agree with him that *if* we identify "world" with a Kantian "thing in

itself" and *if* we identify transcendence with God's eye view, then we must give up the correspondence approach to truth and reject robust realism. But I do not agree that robust correspondence (or realism) requires us to identify world with thing-in-itself and transcendence with God's eye view, and I disagree with his conclusion that we have to renounce the correspondence theory of truth and robust theoretical realism.

13.1 What models are designed and not designed to do – a new challenge to MT

The skeptical MT appeals to certain meta-logical results. The lesson Putnam draws from these results is that it is impossible to determine reference theoretically (as opposed to pragmatically), that is, by means of a theory that says what the referents of our words are. This, according to Putnam, leads to another conclusion, namely, that the correspondence account of truth, and with it, the only acceptable forms of robust realism, fail.

For the sake of clarifying the issues, we may divide Putnam's MT into two arguments, a narrower argument and a broader argument. The narrower argument applies to theories of reference formulated within the framework of standard first-order logic. The broader argument applies to theories of reference formulated within any model-theoretic logical framework. The narrow argument relies on the Löwenheim-Skolem Theorem, the broader argument relies on what I will call the isomorphism theorem.

Before turning to these arguments, however, let me briefly introduce two terminological distinctions: (i) "model for" and "model of," and (ii) "standard" and "non-standard" first-order logic.

- (i) Model theory distinguishes between "M is a model for a language L" and "M is a model of a sentence S of L, or of a theory T formulated in L." Any model in which every sentence of L has a truth-value is a model *for* L. Any model for L in which S/T is true is a model *of* S/T. It will always be clear from the context whether by "model" I mean a model for the language or a model of S/T.
- (ii) By "standard first-order logic" I mean the kind of logical system described in common textbooks of mathematical logic (e.g. Enderton 2001 [1972]). What makes such a first-order logical system "standard" is that it limits its logical constants to the "standard" ones: a complete set of truth-functional connectives (e.g. \sim , $\&$, \vee , \supset , and \equiv), the existential and/or universal quantifier (\exists/\forall), the identity relation between

individuals (=), and any terms defined from these. Non-standard first-order logic includes additional logical constants, such as the quantifier "most."

The narrow MT argument. This argument appeals to a meta-logical theorem called "the Löwenheim-Skolem Theorem" or "LS:"

(LS) Let T be a standard first-order theory. Then, if T has any model, it has a countable model.²

It follows from LS, Putnam says, that if you formulate your theory – say, your mathematical theory of sets – within the framework of standard first-order logic, then your theory will not accurately determine the reference of its terms. For example, it will not accurately determine the reference of the first-level predicate "x is uncountable." Why? Because first-order set theory has models in which "x is uncountable" is satisfied by countable sets. In order to determine the reference of terms like "x is uncountable," Putnam says, your theory would have to single out an "intended" model or a class of models, that is, those in which "x is uncountable" has its *intended reference* or extension. But although ("uncountable" being a mathematical predicate) this only requires that your theory determine an intended model up to isomorphism, even this it cannot do: a countable model of set theory is not isomorphic to an uncountable model of set theory, but first-order set theory cannot distinguish between the two.

Putnam's conclusion is radical: it is impossible in principle to determine the reference of terms theoretically, and this applies both to theories of reference that list the referents of words one by one and to theories that determine their referents by means of general principles, for example by means of a *causal* principle of reference. In the case of theories of the second kind the problem is that they have models in which "x is a causal relation" is satisfied by non-causal relations.

Before evaluating Putnam's conclusion let us be more precise about what follows and does not follow from LS. What follows from LS is that at least some non-logical predicates have unintended denotations in some models of theories in which they appear. What does not follow from LS is, for example, that any logical terms have unintended denotations in any models (see below). We also need to indicate that the reason first-level predicates like "x is a cat" and "x causes y" have unintended denotations is due to a more general

² A countable model is one whose universe is either finite or denumerable, that is, has the smallest infinite cardinality.

meta-logical result: the isomorphism result. This brings us to the broad MT argument.

The broad MT argument. The broad Putnamian argument appeals to the isomorphism theorem (ISOM):

(ISOM) Let T be a theory/sentence formulated in any logical framework with a Tarskian model-theoretic semantics. Then, if T has any model, M , it has infinitely many distinct models, namely, all the models isomorphic to M . In other words: the notion “ X is a model of T/S ” is closed under isomorphisms.

Now, the argument says: Take a theory of reference, T , formulated within the framework of any model-theoretic logic.³ If T has a model in which “cat” refers to cats, then it also has models in which “cat” refers to dogs, trees, numbers, and so on. Similarly, if T explains the reference of words in terms of causal relations between words and objects in the world, then T has models in which “cause” refers to things other than causes: for example, to certain mathematical relations between numbers. Therefore, no theoretical (as opposed to pragmatic) account of reference is adequate.

Is Putnam’s radical conclusion justified? Is it impossible to account for reference by means of a theory that states either the specific referents of words or general principles of reference? I believe the answer to this question is negative. To understand why it is negative, let us turn to “theories.”

Theories. For Putnam, to determine reference theoretically (as opposed to pragmatically) is to formulate a standard first-order axiomatic theory of reference and use its intended model to determine what it says. Is this the right way – or an acceptable, reasonable, fruitful way – to think of a theoretical account of reference? A number of philosophers, for example Plantinga (1982), give a negative answer to this question.⁴ But most do not provide a precise explanation of why the answer is negative. Plantinga, for example, raises the rhetorical question “So what?”: “[I]ndeed number theory does have models of . . . different sorts, but so what?”, concluding that “the process of formalization is severely limited in a certain dimension” (Plantinga 1982: 59–60).

³ By “model-theoretic logic” I mean, in this chapter, simply a logical system that has a Tarskian model-theoretic semantics and is subject to ISOM.

⁴ Such negative answers are aligned with claims by philosophers of science that it is unfruitful to think about scientific theories as theories formulated within the framework of some model-theoretic logic or think about scientific models as models in the model-theoretic sense. (For discussion, see Frigg and Hartman (2012 [2006]) and references there.)

I think we can go further and pinpoint the reason models and logical frameworks are not appropriate tools for determining reference. The key question is: "What are models and logical frameworks designed to do and what, as a result, are they incapable of doing?"

Models and logical frameworks.⁵ If we turn to the paper in which Tarski explained his motivation for constructing a model-theoretic semantics (Tarski 1936), we see that the designated task of this semantics was to identify *logical consequences* correctly. One way to identify logical consequences correctly, Tarski realized, was to formulate the language whose logical consequences we wish to identify within a *logical framework* and construct an apparatus of models that *takes into account the reference of the logical terms of this language while disregarding the reference of its non-logical terms*. That is how model-theoretic semantics is built.⁶ Logical terms are semantically "pre-fixed," while non-logical terms are "highly variable" (Sher 1991: chapter 3).⁷ To determine whether an arbitrary sentence S_2 is a logical consequence of an arbitrary sentence S_1 we have to overlook the full truth conditions of S_1 and S_2 and in particular the "intended" reference of their non-logical terms. And we must focus on their logical form and its significant constituents – the logical constants. It is paramount that model-theoretic semantics get the reference of the logical constants of a given language right (in every Tarskian model "=" must denote the identity relation, " \exists " the second-level property of non-emptiness, " \forall " the property of universality, the logical connectives must denote certain Boolean operators, and the same holds for other logical constants, if there are any). But it is also paramount that the semantics abstract from the particular reference of the non-logical vocabulary. For that reason it is misguided to think of model-theoretic semantics as a general theory of reference, or as a tool for identifying non-logical reference.

Consider, once again, the uncountability predicate. If we build the uncountability predicate, or the membership relation in terms of which it can be defined, as a *logical constant* of a first-order logical framework of set theory (i.e. as a logical quantifier denoting a second-level cardinality property in the case of uncountability, and as a relational quantifier whose first argument is an individual and whose second argument is a property or a set in the case of the membership predicate), then first-order model-theoretic

⁵ For an earlier discussion of some of the points in this subsection see Sher (2000).

⁶ Although models were used in logic before 1936 (e.g. by Skolem), the capacities and limits of model-theoretic semantics are best understood, in my view, by reference to Tarski's 1936 principles.

⁷ For other aspects of the model-theoretic apparatus and an explanation of how it identifies logical consequences and why it is suitable for this task, see Sher (1991, 2013a).

semantics would get its reference conditions right. In Tarski's words: "[I]f we treat \in like a logical symbol . . . interpreted as signifying membership, we will, in general, not have a denumerable [hence, countable] model" (cited in Skolem 1958: 638).⁸ That is, we will not have a countable model of "There are uncountably many things" or "There is an uncountable set." If, however, we build the uncountability predicate as a non-logical constant, then it is incumbent on model-theoretic semantics to abstract from its reference, hence assign it "unintended" referents in some models. And, given the role of models in determining logical consequences, it is essential that model-theoretic semantics treat all models on a par, that is, not distinguish "intended" from unintended "models." The question is, then, why would anyone use a tool designed to overlook non-logical reference to identify (determine, specify, account for) this very kind of reference?⁹

Turning back to LS and ISOM, I think that a reasonable conclusion to draw is that model-theoretic semantics is not a suitable tool for determining reference and that therefore we should not think of a general theory of reference as a theory that determines the reference of our terms by identifying an "intended" model or a set of such models. Thinking of a general theory of reference in this way is counter-productive.

Putnam, however, presented a version of his skeptical argument that does not appeal to model-theory at all, the so-called *permutation argument*. Let us now turn to this argument.

13.2 The significance of invariance under permutation

The permutation argument appeals to a variant of the isomorphism thesis. We may call it the "permutation thesis" (PERM):

(PERM) Let T be a theory, and let A be the ontology of T . Then, if there is one determination of reference for all terms of our language under which all the sentences of T are true, there are many such determinations of reference, obtained from the original determination by permutations of the given ontology. We may say that T is closed under permutations.

⁸ A denumerable model is one which is infinite yet countable. The translation of Tarski's sentence (from French) is mine.

⁹ It is true that sometimes we can use a tool to do something it was not specifically designed to do. But in the case of Tarskian models and determination of non-logical reference this is especially problematic.

The skeptical argument can now be stated as follows: Let T be an adequate theory of reference for our language. Then, presumably, T assigns to "x is a cat" and to "x is a causal relation" a set of cats and a set of causal relations, respectively, as their referents/extensions. But the truth of T is preserved under permutations. So T remains true under some reference-determinations which assign to "cat" a set of non-cats and to "causal relation" a set of non-causal relations. This renders T an inadequate theory of reference. Contradiction. Conclusion: Determining the reference of our words theoretically, that is, by a theory T, is impossible.

In response to this argument let me make two comments:

1. Permutation does not change the reference of *all* words in our vocabulary. Take, for example, the second-level predicate "non-empty" (the existential quantifier of first-order logic), understood as "X is a non-empty property of individuals," or "X is a non-empty set of individuals." Now, the (correct) extension of "non-empty" is the set of all non-empty sets of individuals in the world. Let us call this set "NE." It is easy to see that the image of NE under any permutation of the individuals in the world ("universe of discourse" of our language) is NE itself. That is, the reference of the second-level predicate of non-emptiness does not change under permutations. The same holds for the first-level relations of identity and non-identity, the second-level predicates "is universal," "has cardinality α " (for any cardinal α), "is symmetric" (i.e. "is a symmetric relation"), and so on. Elsewhere (e.g. Sher 2013a) I suggested that we characterize all the predicates that are preserved under all isomorphisms – hence, all permutations – as *formal*.¹⁰ Under this characterization, the reference of formal terms, unlike the reference of terms like "cat" and "is a causal relation," which are not formal, is preserved under all permutations. So the claim that PERM undermines the possibility of a theoretical account of reference is at most partially correct, that is, correct for that part of the theory of reference that deals with non-formal terms.
2. PERM does not undermine the possibility of a theoretical account of reference for non-logical terms as well. PERM shows that if by a theory you understand just a pattern of truth-values assigned to sentences of our language (those assigned "true" belong to the theory, those assigned "false" do not), then a theory in this sense cannot account for reference. It also shows that if by a theory you understand a set of uninterpreted sentences considered as axioms and closed under the relation of logical consequence,

¹⁰ All logical terms are formal in this sense.

then again, a theory in this sense cannot account for reference. A mere pattern of truth values is something formal, hence invariant under permutations. And by paying attention only to patterns of truth values what we get is a theory of truth-functional operators, not of reference. An uninterpreted axiomatic theory, too, cannot convey non-formal content. But these are not the only ways to think of a theory of reference. A theory of reference is itself an interpreted theory, and its basic terms (e.g. "is a causal relation") must all be interpreted.

Someone might complain that such a theory cannot determine the reference of its interpreted terms. For example, the reference of "causal relation" must be determined outside the theory. This is true in some sense, but not in a sense that is problematic for me, or that should be problematic for Putnam, since we both share a holistic outlook on theories. As a holist, I believe that, for a theory to be significant, at least some of its words must come with a prefixed meaning. We cannot get something from nothing, and even our theories of meaning/reference must assume some meanings/references. But I also believe that we do not have to hold the same meanings/references fixed in all our theories or at all times. To develop a general theory of reference we may take the term "cause" from other theories, where this term was defined, explained, characterized, and its satisfaction conditions were at least partly determined. Then, using the theory of reference we have just developed with the help of this term, we might revisit the theories we used to fix its reference in the first place, reexamine their use of this term in light of what our newly developed theory of reference (which has other resources besides the prefixed meaning/reference of "cause") teaches us about reference, in light of what other theories teach us about causes and things related to causes, and so on. And based on this we decide whether to leave the reference of this term as it is or revise it.¹¹ We may even account for the reference of "cause" within the theory of reference by holding other terms fixed (relying on other theories, other discoveries, other contacts with the world, and so on). There is nothing skeptical in this conclusion: A theory of reference is an ongoing project just like any other human project.

Putnam, however, claims that in the case of the correspondence theory of truth, in the (closely related) case of robust realism, and in the case of the

¹¹ Other possibilities are discarding the term or replacing it by another, which will then lead to revisions in our theory of reference.

theory of reference they require, we are doomed to failure. Here, the only way to proceed is to appeal to magic (supernatural cognitive powers). Let us now turn to this claim.

13.3 Do correspondence and robust realism require magic?

The key to evaluating Putnam's claim is understanding what he means by "the correspondence theory of truth" and "robust realism." Let us begin by seeing what adjectives he uses to describe the realism he rejects. They include "*moderate*" (Putnam 1980: 464), "*metaphysical*" (Putnam 1981a: 49), "*externalist*" (49), "modern" (51), "*hard-line*" (Putnam 1983b: viii), "causal" (xi), "naive" (xi), and "classical metaphysical" (xviii)¹²

Next, let us look at some of the ways Putnam characterizes this realism. This realism, according to Putnam, "seeks to preserve the centrality of the classical notions of truth and reference *without postulating non-natural mental powers*" (Putnam 1980: 464). "[I]ts favorite point of view is a *God's Eye point of view*" (Putnam 1981a: 49). It is a realism which holds that "the truth of a theory . . . consist[s] . . . in its **corresponding to the world as it is in itself**" (50). It adheres to "*the notion that our words 'correspond' to determinate objects* (where the notion of an 'object' is thought to have a determinate reference which is independent of conceptual scheme)" (Putnam 1983b: viii). It is a realism that **requires "an absolute¹³ relation to the world"** (xi). It assumes "that we can think and talk about things as they are, independently of our minds" (Putnam 1983a: 205). This realism "claims that **there is¹⁴ a 'transcendental' reality in Kant's sense, one absolutely independent of our minds**, that the regulative ideal of knowledge *is* to copy it or put our thoughts in 'correspondence' with it, *but . . . we need no intellektuelle Anschauung¹⁵* to do this: *the 'scientific method' will do the job for us*. 'Metaphysics *within the bounds of science alone* might be its slogan" (226). It involves the "idea of a coherent **theory of the noumena**; consistent, systematic, and **arrived at by 'the scientific method'**" (226).

Now, a careful look at these characterizations shows that Putnam sometimes views the realism he rejects as an *enlightened* realism, one held by

¹² Here and below I use bold in citations to direct the reader's attention to things I will focus on later. Because I would like to point a potential tension in Putnam's view of the realism he rejects, I will use two styles of bold: regular and italics. In the above citations, "moderate" is already italicized in the original text, though for a different purpose.

¹³ In the original text, "absolute" is italicized for emphasis.

¹⁴ In the original text, "is" is italicized for emphasis.

¹⁵ In the original text, too, "*intellektuelle Anschauung*" is italicized, but not for emphasis.

enlightened philosophers – enlightened in the sense of being moderate and rejecting supernatural powers. (See italicized bold expressions.) At other times Putnam views this realism as an *unenlightened* realism, in the sense of being extreme and assuming a radical, outdated conception of reality and a radical, outdated standpoint: *thing-in-itself* and *God's eye view*. (See regular bold expressions.)¹⁶

So, is the realism Putnam criticizes an enlightened realism, in which case his criticism might be very significant for the viability of robust realism and the correspondence theory of truth, or is it an unenlightened realism, in which case his criticism would be less significant? One clue to understanding Putnam's conception of this realism is his statement that "it is, unfortunately, the *moderate* realist position which is put into deep trouble by the Löwenheim–Skolem theorem and related model-theoretic results" (Putnam 1977: 2). This suggests that Putnam thinks that robust realists who support science and reject magic believe they are enlightened, but in fact they are not, since their realism is committed to *noumenal reality* and a *God's eye view*.

Putting aside the question of what kind of robust realism scientifically-minded contemporary philosophers actually support, the question I would like to raise here is whether an enlightened robust realism, one that does not appeal either to things-in-themselves or to a God's eye view, is possible. If the answer to this question is positive, this, together with my earlier challenges to Putnam's meta-logical arguments, would challenge his entire skeptical argument.

I believe that the answer to the above question is, indeed, positive. To show why it is positive, I will briefly delineate a philosophical methodology that renders robust correspondence and realism possible without rendering commitment to thing-in-itself, a Godly standpoint, or magic, possible. The discussion below is based on Sher (1999, 2004, 2010, 2013a, b, and 2015).

The methodology I have in mind was already introduced earlier in this chapter. It is the *holistic* methodology, or rather a specific version of holism, which in Sher (2013a) I call "foundational holism." My claim is that if we realize holism's potential, we can develop an enlightened yet robust correspondence theory of truth, one that requires substantive correspondences

¹⁶ Of course, it is an open question whether in every place identified here as indicating an unenlightened conception of realism this accurately captures Putnam's intention in this particular place. But what is important for the purpose of the present chapter is that (i) in order for Putnam's criticism of robust realism to make sense he must attribute to robust realists views of realism involving noumena (things-in-themselves) and a God's eye view, and that (ii) there are many places in his writings in which he either attributes these views to them explicitly or uses expressions that, given his criticism, are naturally interpreted as representing these views.

between truths and reality yet eschews the problematic commitments feared by Putnam. My discussion will be divided into three parts: (a) foundational-holistic approach to knowledge; (b) reality and correspondence without things-in-themselves; and (c) transcendence and correspondence without God's eye view.

- (a) *Foundational-holistic approach to knowledge.* Following Quine (1951), many contemporary philosophers classify themselves as holists, though different philosophers understand different things by holism. Here I am not interested in the historical roots of holism, its current sub-classifications (e.g. meaning holism versus confirmation holism), or who supports (objects to) what kind of holism. Rather, I would like to briefly delineate a conception of holism that is pertinent to the question at hand and that, I believe, opens up more options in developing a correspondence theory of truth, a realistic epistemology, and a theory of reference. The holism I have in mind, *foundational holism*, may best be understood by comparison with *foundationalism*. Foundational holism shares with traditional foundationalism the view that a substantive grounding of knowledge in reality is both required and possible, and it goes further than most versions of foundationalism in making this requirements truly universal, that is, applicable to highly abstract disciplines like logic as much as to largely empirical disciplines like biology. Unlike foundationalism, however, it does not impose a rigid ordering on our system of knowledge (collection of disciplines and subdisciplines) according to "basicness," and it does not require that in grounding a given discipline (subdiscipline) in reality we limit ourselves to resources produced by more basic disciplines (subdisciplines). The grounding process, like knowledge itself, is an ongoing process, involving back and forth motion, change and revision, and setting demanding, yet not impossible, requirements on both the construction and the grounding of theories.
- (b) *Reality and correspondence without things-in-themselves.* The foundationalist methodology, as I understand it here, aims at an absolute foundation for knowledge, a grounding of knowledge in absolute reality, an account of how terms of our language refer to absolute objects, and how sentences correspond to absolute facts or states of affairs. In short, the foundationalist thinks of reality as *noumenal reality* or *thing-in-itself* in Putnam's sense, something which is absolutely and completely independent of the way the human mind conceives reality or cognitively accesses it. In contrast, the foundational holist thinks of reality in

non-absolutist terms. When I say that my standard of truth for "Snow is white" is a correspondence standard I mean that for this sentence to be true it has to be the case that the worldly, non-noumenal, stuff snow has, in the world, the worldly, non-noumenal, property of being white. This kind of standard is a robust correspondence standard,¹⁷ but it has nothing to do with noumenal reality. Similarly, when I say that my standard of truth for first-order arithmetic sentences is a correspondence standard, I mean that such sentences have to be systematically connected to certain facets of reality in order to be true, but not that they have to be systematically connected to noumenal reality or even to Platonic reality. This point is worth elaborating on.

From a holistic point of view there is considerable flexibility concerning the form correspondence takes, namely, what pattern the connection between true sentences and reality takes in different fields. In particular, there is no need to insist that the correspondence relation is a copy or an isomorphism relation, or that it is a direct relation or the same relation in all fields. This is especially important in developing a correspondence account of mathematical truth, since the traditional way of thinking about mathematical correspondence involves the requirement that there are mathematical individuals in the world and the mathematical vocabulary refers to these individuals, a requirement that has turned out to be highly problematic (see e.g. Benacerraf 1965, 1973).

From a holistic perspective, it is an open question how our words, including our abstract words, are connected to reality. The key to answering this question is Wittgenstein's injunction: "Look and see!"¹⁸ Don't decide in advance how words are related to objects, but look and see, search, find out. The fact that humans use 0-level expressions like "one" to state the laws of finite cardinalities, for example, does not automatically mean that for their statements to be true – correspondence true, true about the world – the term "one" must be connected to a numerical individual in the world, let alone directly connected to such an individual. Perhaps there are no numerical individuals in the world, but there is something else in the world that "one" is connected to and in virtue of

¹⁷ I can make my standard of truth for this sentence still stronger by setting precise scientific conditions on something to be snow and equating the condition of being white with some non-trivial physical condition in terms of wavelength.

¹⁸ Wittgenstein (1953: §66), where it is written "look and see." Unlike Wittgenstein, however, I do not contrast looking and thinking. On the contrary I see some forms of thinking as falling under "looking."

which mathematical statements are true or false about the world, true or false in a correspondence sense.

Thus, consider the observation that objects and properties in the world have properties of many kinds, and these include formal properties like cardinalities. Suppose there are no numerical individuals in the world, but properties of individuals in the world have cardinality properties. For example, the first-level property *is-a-moon-of-Earth* has, in the world, the second-level property of having cardinality ONE. Suppose, further, that for some reason having to do with our cognitive capacities we prefer to construct first-order theories of cardinalities rather than second-order theories. Suppose, for example, that we, humans, get confused when we deal with higher-level properties and their relations, but see our way clearly into these relations when we translate them into relations between individuals. That might result in our having a language with a 0-level term, "one," which *refers* to the second-level cardinality property ONE. How would it do that, and how would first-order arithmetic statements be true of second-order cardinality properties? Well, we, humans, might exercise our "epistemic freedom" and create a 0-level model of second-level cardinalities. Our word "one" would be systematically connected to our 0-level "posit" one, which, in turn, would be systematically connected to the second-level cardinality property ONE. In this way first-order arithmetic statements could be true, in a "composite" correspondence sense, of certain formal facets of reality, namely those involving second-level finite cardinalities. (For more details, including how this account can be extended to theories of infinite cardinalities, see Sher (2013b; 2015).) Mathematics, then, even higher mathematics, would be true of our world, true about some abstract facets of our world, without requiring any noumenal or Platonic world to be true of.

- (c) *Transcendence and correspondence without a God's eye view.* A correspondence standard of truth requires a three-level setting: Level W (world), level L (language), and level T (transcendence). On level L there is a theory, statement, or thought S. S says something about the world (where "world" is broadly understood). In the simplest case, S attributes a property to an object in the world, that is, S attributes a property of level W to an object of level W. Now, to apply a correspondence standard of truth to S we must transcend L to a level T from which we can see both S and that portion of the world that S has to be systematically connected to in order to be true. That is, our correspondence standard for S lies in T, a level that has both L and W in view. A paradigmatic example of level T is that of a Tarskian meta-language.

Considering a pair of levels $\langle W, L \rangle$, where objects of level L refer, or attribute properties, to objects of level W , and characterizing level L as a *human level of thought*, we can in principle distinguish two types of transcendence (cognitive movement) from level L to a level transcendent to it, T . We may call these *Human–Human (HH)* and *Human–God (HG)* transcendence:

HH-transcendence: We transcend a human level of thought to a different human level of thought.

HG-transcendence: We transcend a human level of thought to a different, non-human or Godly level of thought.

Now, according to Putnam, a correspondence standard of truth requires HG-transcendence. But from a holistic perspective, a correspondence standard of truth requires only HH-transcendence. According to Putnam, a Tarskian meta-language is a Godly language; according to us, it is a human language. It is not surprising that for Putnam level T of thought is a Godly level: if level W is the level of thing-in-itself or noumenon, and if level T is a level from which we can access both level W and level L , then level T must be a Godly level. Only a God is capable of seeing which things-in-themselves have which properties-in-themselves. But if level W is the level of *regular* reality, the reality studied by *humans* (by our physical, biological, psychological, and mathematical theories), then level T is a level of *regular* cognition, that is, *human* cognition. It is a more powerful level of cognition than level L , as we have learned from Tarski. We can see things from level T that we cannot see from level L . But level T is just as human as level L . To set a standard of truth for human theories (statements, thoughts) there is no need to be God. This is something that humans do, can do, and have always done. It is also something that we can do better or worse, something that it takes quite a lot of work to be better at (figuring out ways to access physical aspects of the world that are not accessible to us at the moment, replacing common-sense methods of accessing the world by rigorous methods, creating sophisticated meta-languages, and so on). But it does not require either being Gods or using supernatural resources.

I can sum up my challenge to Putnam by saying: Putnam believes that robust realists and correspondence theorists must renounce their positions because (i) they cannot give an account of reference of the kind required by their positions, and because (ii) their positions unavoidably carry problematic commitments. But the meta-logical considerations that lead Putnam to conclude (i) are irrelevant to a robust realist/correspondence account of reference. And the considerations that lead Putnam to conclude (ii) are incorrect:

there are ways to be robust realists and correspondence theorists, including ways that are friendly to many aspects of Putnam's philosophical outlook, that do not involve the problematic commitments rejected by Putnam.

I have said earlier that Putnam is a holist rather than a foundationalist. How does Putnam's holism differ from foundational holism? A full answer to this question is not possible in this chapter. But one significant difference is a difference in the *balance of pragmatic and veridical considerations*. Both from the point of view of Putnam's holism and from the point of view of foundational holism, pragmatic considerations are integral to the project of knowledge in all fields, including philosophy and, within it, epistemology and semantics. But while Putnam gives more weight to pragmatic goals, considerations, and methods than to veridical ones, especially in philosophy, foundational holism gives more weight to veridical methods, goals, and considerations. This explains why, for foundational holism, a correspondence standard of truth, which focuses on veridicality requirements, is crucial for knowledge in all fields, while for Putnam a pragmatic standard, like acceptance of theories at "the ideal limit," is sufficient.

This is connected to the "foundational" aspect of foundational holism. While holism is usually viewed as an *antidote* to foundational interests, foundational holism is holism put *in the service of* just such interests. Foundational holism says that the holistic rather than the foundationalist method is the preferred method for foundational investigations, including those aiming at understanding the relations between mind and world, theories and their targets, words and the objects they refer to.

Another distinctive characteristic of foundational holism is its approach to humans' cognitive resources. One resource, in particular, is viewed very differently by the foundational holist and by other epistemologists, holists as well as foundationalists: intellect. This brings me back to Putnam's BIV argument, which focuses on such resources.

13.4 Intellect and the BIV challenge

BIV skepticism is a modern version of Cartesian, Evil-Demon skepticism, and Putnam's anti-skeptical argument is a modern counterpart of Descartes's anti-skeptical argument. But the two differ in substantial ways. In particular, while Descartes's argument is a rationalist argument, Putnam's argument is to a large extent an empiricist argument. This difference is crucial to the success of Putnam's argument. While Descartes's argument does not distinguish between creatures that share the same intellectual capacities but not the

same sensory capacities, Putnam's argument does. It is the fact that we stand in direct causal relations to the physical world, mediated through our sensory organs, that distinguishes us from BIVs, according to Putnam. And it is this difference that makes it impossible for us to entertain the skeptical possibility that rather than humans with cognitive access to the world we are BIVs controlled by a computer run by a (possibly malevolent) scientist or by an indifferent nature.

Now, one feature of this anti-skeptical argument is that it neglects to examine the possibility that, under certain conditions, intellect, as a *natural* resource, is also capable of forging significant cognitive routes to the world. What I am talking about is *not* the specific possibility of purely *a priori* knowledge which, as a holist, I tend to be skeptical of. What I have in mind is the more general possibility that intellect plays a significant role in our cognitive access to the world, either in isolation from our other cognitive resources or in cooperation with these resources. I think that exploring this possibility is an important task for contemporary holists who, like Putnam and myself, reject both radical empiricism and radical rationalism.

Furthermore, one way to orient ourselves toward this investigation is to pose a *new* BIV question. Not "Could I truly believe (or seriously entertain the possibility) that I am a BIV?" but: "Given that it is possible that BIVs exist, is it possible that under certain conditions *they* will have some genuine knowledge of the world, and if so, what kind of knowledge and under what conditions?" Although this question does not directly concern our own cognitive access to the world, it could potentially teach us something about our own cognitive access, and, in particular, about the possible role of human intellect (which, by construction, is the same as BIV intellect) in this access.

One source of inspiration for this question is Kant's observation that while there is "no doubt that all our knowledge [cognition] begins with experience[,] . . . it does not follow that it all arises out of experience" (Kant 1787 [1781]: B1). Kant himself went on to ask whether there was purely *a priori* knowledge, but the possibility that human knowledge arises partly out of intellect does not require *a priori* knowledge. Nor does it require intellectual intuition akin to perception. A more suitable paradigm of intellect's ability to access reality is "figuring out," where this might be a lengthy process, involve use of background knowledge, and so on. Figuring out is a natural activity, one found in animals as well as in humans, and in humans it occurs in all ages and in all contexts, from everyday practical endeavors to the most abstract mathematical endeavors. This is how I would formulate the "new" BIV question: "Could BIVs figure out some things about the

world? What kind of things? Under what conditions?" Could *we* figure out some things about the world in a similar way?

Another source of inspiration comes from real-life examples of humans who overcame severe sensory limitations, such as Helen Keller. BIVs are severely limited in their physical ability to access the world. But they have some physical contact with the world – their brains are physically connected to a physical object, a computer, which, at least under some conditions, is connected to a person who herself might have sensory contacts with the world. Under what conditions are such contacts sufficient (necessary, instrumental) for BIVs obtaining indirect knowledge of the world?

A third source of inspiration is Plato's parable of the cave. Is it possible that one of the BIVs will suddenly, or not so suddenly, "awaken from its slumbers," "stand up," "walk out" (cognitively speaking) and "see" the light? Not Platonic forms, but regular things, the kind of things we want our own theories to teach us about? Is it possible that a BIV will outsmart its handler? Could it cooperate with its handler? Could it see through attempts to deceive it? Could it find ways to exploit the computer "controlling" it? Quine said that science has taught us that our contact with the world is *purely* sensory. But has it? Has science given us a full and conclusive account of our cognitive access to the world, our ability to figure out things about the world, the role of intellect in accessing the world? These questions call for "looking and seeing," and "figuring out" is part of it. The question of what BIVs could and could not do and under what conditions calls for the same. And if there are conditions under which BIVs could figure out some things about the world, are we as different from them as Putnam thinks we are? Is it absolutely irrational to entertain the possibility that we are them, that we are at least a little bit like them?

The Brain in a Vat

Edited by
Sanford C. Goldberg

 **CAMBRIDGE**
UNIVERSITY PRESS