EUROPEAN JOURNAL OF
ANALYTIC PHILOSOPHY

CONTENTS

ARTICLES

PUTNAM'S CONCEPTION OF TRUTH*

MASSIMO DELL'UTRI

University of Sassari, Italy

ABSTRACT

After stressing how the attempt to provide a plausible account of the connection between language and the world was one of Putnam's constant preoccupations, this article describes the four stages his thinking about the concepts of truth and reality went through. Particular attention is paid to the kinds of problems that made him abandon each stage to enter the next. The analysis highlights how all the stages but one express a general non-epistemic stance towards truth and reality—the right stance, according to Putnam, in order to develop full-blooded realism. Since the last stage combines a version of direct realism with a pluralist conception of truth, the article proceeds by focusing on Putnam's alethic pluralism, carefully distinguishing it from alethic deflationism. Finally a suggestion is made as to where Putnam's alethic pluralism may be placed within the constellation of current pluralist positions about truth.

Keywords: Truth, alethic pluralism, alethic deflationism, realism, Hilary Putnam

1. Introduction

The aim of this article is to analyse Hilary Putnam's last conception of truth, making explicit the aspects he left implicit. Since, on the one hand, one of the major traits of this conception is its being *pluralistic* and, on the other, a pluralist spirit permeates alethic deflationism, an effort will be made to appreciate the distance—however small—that separates Putnam's conception from alethic deflationism. In the course of the analysis, the question is tackled as to whether Putnam's pluralistic notion of truth renders the word "true" semantically ambiguous—a question that

seems to have been ignored in the literature so far. In order to show Putnam's last conception of truth in its own light, the article starts by following the stages that led up to it over fifty years or so of philosophical reflection.

2. Truth and Reality

The first thing to say in addressing Putnam's conception of truth is that he envisaged a strong link between the concept of truth and the concept of the world, so that an analysis of the former is unavoidably intertwined with an analysis of the latter, and vice versa.

Indeed, one of the constant traits of his thought is the conviction that "the major problem of philosophy [is] the problem of the way language and thought 'hook on' to the world" (Putnam 1983, 315), where the implicit idea is that a correct understanding of truth gives both a grasp of that "hooking" and what that hooking hooks onto. This may happen because, intuitively, when a proposition is true, truth can be taken to show the existence of a relation between that proposition and the portion of reality it is about—whatever the interpretation of truth we are willing to take on board. And not only does what we say is true highlight what intentionality amounts to, but it shows something of that portion of reality as well. Putnam himself revealed that "the problem of intentionality has been a lifelong preoccupation of mine, and […] various changes in my position were occasioned by the realization that one or another assumption about the nature of reference led to deep difficulties" (Putnam 2013a, 24).

One aspect that represents another constant trait of Putnam's thought and had an influence in the development of his conceptions of truth and reality is his *anti-positivist stance*. It constitutes perhaps the main source of his realistic attitude, since he regarded any positivist perspective as heavily slanted towards idealism. The *binary development* of Putnam's notions of truth and reality went through four stages. In chronological order:

1) Alethic correspondentism	⇔	metaphysical realism
2) Alethic correspondentism	⇔	sophisticated metaphysical realism
3) Alethic pragmatism	⇔	internal realism
4) Alethic pluralism	⇔	natural realism

Roughly, Stage 1 took place during the Sixties, and combined a view Putnam a decade later called "metaphysical realism" with a correspondence account of truth. He also made an ingenious attempt to provide a definition of correspondence, which he later deemed hopelessly flawed. The attempt was centred on the notion of "compositional mapping", and elaborated the idea according to which "a true sentence is *not* one which bears a certain relation to extra-linguistic facts, but one which bears a certain relation to extra-linguistic facts *and to the rest of the language*. (The 'correspondence' is triadic rather than diadic.)" (Putnam 1960, 82). This definition was an integral part of the metaphysical realism he subscribed to at that time, a view which also had three more assumptions: the idea that there exists (a) a fixed totality of all objects, (b) a fixed totality of all properties, and (c) a sharp line between properties we discover in the world and properties we project onto the world (cf. Putnam 1999, 183).

Stage 2 took place during the Seventies. It inherited the general framework of Stage 1, except that Putnam became aware of the impossibility of any definition of truth as correspondence. Moreover, he recognized the phenomenon of equivalent descriptions[1] (which amounted to a rejection of (c) above and made his metaphysical realism "sophisticated": cf. Putnam 1978, 51 and 131).

Internal realism triggered Stage 3. This is a stage for which there are precise starting and ending dates: 1976 and 1990, respectively. In Boston, on December 29, 1976, he delivered a talk entitled "Realism and Reason" (which was then published as the last part of Putnam 1978) where the phrase *internal realism* made its first appearance, while in the course of the Gifford Conference held at the University of St Andrews, November 23-6, 1990, in replying to the talk given by Simon Blackburn he explicitly renounced the view (cf. Putnam 1994b). Internal realism has it that "the mind and the world jointly make up the mind and the world. (Or […] the Universe makes up the Universe—with minds—collectively—playing a special role in the making up.)" (Putnam 1981, xi), so that what reality and truth really are stems from our best cognitive procedures. Truth, in particular, gets a pragmatist interpretation, in that it is seen as what can be asserted in epistemically-good-enough-conditions—an idea reminiscent of Charles Sanders Peirce's account of truth, although different in an important respect.[2]

---

[1] This is the phenomenon represented by the *cognitive equivalence* of sentences, theories, or conceptual systems which, when taken at face value, are *incompatible*: e.g., two sentences saying different things about the same portion of reality and being, nevertheless, both true (cf. Putnam 2013a, 23-24).

[2] There is no reference in Putnam to a purported ideal limit of inquiry.

Stage 4 officially opened in March 1994 on the occasion of the John Dewey Lectures Putnam held at Columbia University (cf. Putnam 1994c). In these lectures Putnam's realist attitude is influenced by William James, John Dewey, the later Wittgenstein and John Austin, and is tied to a view of perception which drops every interface between the human sensory apparatus and the world in favour of a *direct* connection between them. It is also deeply steeped in common sense. This position was termed *natural realism*, paying homage to the "natural realism of the common man" (Putnam 1999, 10). With it Putnam combined the idea that truth amounts to many different things—as many different things as many kinds of true propositions there are, and as many domains there are in which a proposition can be true: empirical, mathematical, logical, ethical, juridical, religious, and so on. In brief, truth is not one, but many.

All four stages represent an effort at showing how "language and thought 'hook on' to the world". With an important difference: Stage 3 is the expression of an *epistemic conception* of truth and reality, i.e. a conception according to which what is true and what is real are a function of our best conceptual scheme, and are therefore expressed by the propositions this scheme allows to justify, whereas Stages 1, 2 and 4 are enlivened by a *non-epistemic conception*, i.e. one to the effect that what is true and what is real may sometimes outrun justification, "because what goes on in the world is sometimes beyond our power to recognize" (Putnam 1999, 69). But, notice: the non-epistemic conception in 4 is of a different flavour to the one in 1 and 2, owing to the specific new views on reality and truth which manifest that conception—equivalently, owing to the specific new views on how "language and thought 'hook on' to the world".

The chief difficulty confronting Stages 1 and 2—the one that made Putnam shift to Stage 3—is how to account for the purported relation of correspondence linking two sharply separate elements (language/mind, on the one hand, and the world, on the other). In fact, a relation of this kind would be *external* to both the elements it puts in relation—in particular it would be external to language and mind, so that it turns out impossible for a human being to conceive it, let alone describe its nature. (Let us call this the *Kantian problem*.) Such a correspondence would only be grasped from what has been termed a "God's Eye View", i.e. a superhuman perspective which, in contexts like these, serves no useful explicative purpose. The moral is hence obvious:

> Elements of what we call "language" or "mind" *penetrate so deeply into what we call "reality" that the very project of representing ourselves as being "mappers" of something "language-independent" is fatally compromised from the very start*. […] Realism is an impossible attempt to view the world from Nowhere (Putnam 1990, 28),

where the realism in question (which Putnam was fond of writing with a capital "R") is metaphysical realism. As I hinted above it was this idea[3] that triggered his epistemic move toward Stage 3—a stage in which the interlacement of mind and the world appears at its best, vindicating the label "internal" for that kind of realism.

However—despite the term "realism" in *internal realism*—Putnam came to realize that this was not realism enough, beginning his way back to a view in which the *independence* of the world from the mind and its theoretical products is more definite. This called in turn for a novel account of the "hooking", one which would avoid the Kantian problem and the implausible account based on mind and the world jointly offering good enough epistemic conditions for what is true and what is real.

This novel account followed the realization that the "'how does language hook on to the world' issue is, at bottom, a replay of the old 'how does perception hook on to the world' issue" (Putnam 1999, 12). The traditional idea according to which we perceive a given object thanks to the myriad *sense data* giving us information about the many features of the object, so that what we are directly connected with is not the object but the sense data, raises the same epistemological difficulty in which Descartes found himself entrapped: the distinction between a mental and a physical substance that are so neatly separated to justify the hypothesis that we might after all be brains in a vat. Indeed, what could ensure that the cause of the sense data we perceive is an object existing in the world out there and not just some computer software linked to the synapses of those deluded brains?[4] Even if we were not brains in a vat, what could ensure that sense data do give us a faithful representation of the object and do not distort perception itself in inscrutable ways?

So, according to Putnam, if we keep endorsing the traditional account of perception, we will find ourselves at a loss as to how to account for the connection between mind and the world, since it appears unavoidable that we appeal to interfaces between ourselves and the world, in the form both

---

[3] It seems that this idea became a somewhat constant trait in Putnam's thought (therefore valid in his last non-epistemic stance too: see Stage 4 below). In fact, it appears that in Stage 4 there is a coexistence of two apparently contrasting beliefs. On the one hand, the belief according to which the dichotomy between properties we discover in the world and properties we project onto the world—cf. assumption c) above—is unjustifiable (a belief that, as we saw, amounts to the acknowledgment of the phenomenon of the equivalent descriptions, and that shows how Putnam was still maintaining that language and mind penetrate deeply into reality. On the other hand, the belief according to which there can be statements that are true or false and whose truth value is doomed to be beyond our ken even in principle: e.g., "There are no intelligent extraterrestrials in the Universe" (cf. for instance Putnam 2015d, 142). Statements of this kind show that, possibly, part of the world is *impenetrable* by language or mind, even in principle.

[4] For the brains in a vat hypothesis, cf. Putnam 1981, chp. 1. For a useful discussion of the hypothesis, cf. Thorpe 2017.

of sense data and conceptual schemes:

> on the "internal realist" picture it is not only our experiences (conceived of as "sense data") that are an interface *between* us and the world; our "conceptual schemes" are likewise conceived of as an interface. And the two "interfaces" are related: I saw our ways of conceptualizing, our language games, as controlled by "operational constraints" that ultimately reduce to our sense data (Putnam 2013a, 26).

Hence natural realism, which—as we saw above—is a form of direct realism. According to this metaphysical picture there is no separation between the human mind and its environment, so that the problem of their relationship does not even arise. But, Putnam hastened to clarify, it does not arise provided that we have a conception of the mind different from the traditional conception, the one inherited and revitalized by Descartes: a mind conceived of as a thing, an organ, a self-sufficient entity already endowed with all its powers. Rather, the human mind is a system of interconnected abilities that involves the world and its objects from the start:

> Mind talk is not talk about an immaterial part of us but rather a way of describing the exercise of certain abilities we possess, abilities that supervene upon the activities of our brains and upon all our various transactions with the environment but that do not have to be reductively explained using the vocabulary of physics and biology, or even the vocabulary of computer science (Putnam 1999, 37-8).

Thus, the elimination of sense data from the account of perception and the functioning of the human mind allows Putnam to discard not only "the model of the mind as something 'inside' us" (Putnam 1992, 357), but also the assumptions that remained to be discarded in the non-epistemic conception embedded in Stage 1. In particular, the notion of truth as correspondence.

Indeed, the need to appeal to a metaphysical relation of correspondence in order to give substance to the "hooking" vanishes, given that "the relation of statements to states of affairs 'out there' is too *internal* to be thought of as a 'correspondence'" (Putnam 2015b, 790). Above all, correspondence may account for some truths, but not all the truths. This has to do with what Michael Lynch has called the *scope problem*, i.e. "for any sufficiently robust characterized truth property F, there appears to be some kind of proposition K that lacks F but that are intuitively true (or capable of being true)" (Lynch 2009, 4). This is clearly a problem any correspondentist interpretation of truth must face: even if we admit the plausibility of a correspondentist explanation of the truth of empirical propositions, it turns out to be much more difficult to apply the same explanation to the truths in ethics, mathematics and the like.

In fact, Putnam's later work puts ever more stress on the irreducible and unpredictable variety that has to do with the concept of truth: the high variety of "scopes" in which truths can be stated, where the empirical scope is but one of many; the high variety of the contexts of usage of linguistic expressions within just one scope, each governed by distinct norms of rightness; the high variety of the kinds of evaluation of the propositions' truth-value. All this nourishes his latter conception of truth, a sort of *alethic pluralism* which combines the idea that there is an extendable family of uses of the terms "true" and cognates—an extendable family of ways of characterizing the answerability to reality truth consists of in new areas of discourse—with the idea of *normativity*, i.e. that "to regard an assertion or a belief or a thought as true or false *is* to regard it as being right or wrong" (Putnam 1999, 69), such that "it is a property of the notion of truth that to call a statement of any kind […] true is to say that it has the sort of correctness appropriate to the kind of statement it is" (Putnam 2013b, 97-8). But, again,

> just what sort of rightness or wrongness is in question varies enormously with the *sort* of discourse. *Statement*, *true*, *refers*, indeed, *belief*, *assertion*, *thought*, *language* […] have a plurality of uses, and new uses are constantly added as new forms of discourse come into existence (Putnam 1999, 69).

Let us try to characterize this sort of pluralism in more detail.

3. Taking into Account the Plurality of Kinds of Truths

In the current literature there seem to be just two elucidations of truth that take the plurality of kinds of truths in due account, and one may ask which of the two is Putnam's position to be ascribed to: *alethic pluralism* proper and *alethic deflationism*. Both are families of theories, rather than compact unified theories on their own.

Broadly conceived, alethic pluralism can be identified with the thesis that there are many ways of being true (cf. Pedersen and Wright 2016). Within this vast receptacle one can find both authors who maintain that there is just one property of truth which is multi-faceted, i.e. possessing many forms, and authors who think there are many different properties in virtue of which a statement can be true, combined with the thesis that the property that makes a statement true may vary from discourse to discourse. Usually, some of the properties so countenanced are "substantial", namely refer to a purported nature or substance of truth.

On the other hand, alethic deflationism is the view according to which all there is to truth are instances of the so-called *equivalence schema*, i.e. *p* is

true if, and only if, *p*—where *p* varies on one's favourite truth-bearer.[5] This is all speakers need in order to have a full mastery of the concept, according to deflationists. It follows that truth has no nature (it is not substantial) and that either the predicate "is true" is not genuine— according to the radical wing of deflationism—or it has only an expressive utility, not an explanatory one—according to the moderate wing. In Paul Horwich's words, truth is "merely a useful *expressive* device, enabling certain generalizations to be formulated—for example, 'All propositions of the form, <p or not-p>, are true', and 'A belief is correct if and only if it is true'" (Horwich 2016, 100).

Deflationism and pluralism are incompatible views about truth, since the former denies "the key pluralist idea that there is a multitude of substantive properties that are alethically potent within specific domains" (Pedersen and Wright 2013b, 10), or just one substantive property susceptible to many different uses. Yet, both militate against the scope problem—pluralism solves and deflationism dissolves it. As to the latter, the "scope problem […] will be regarded by the deflationist as little more than a mildly diverting irrelevance" (Dodd 2013, 315). But why?

The fact is that deflationism about truth has a sort of intrinsic pluralist flavour. Since you can substitute whatever sentence from whatever region of discourse to the *p* in the equivalence schema, you can take the plurality of truths into account in one fell swoop. And you can do this without any appeal to a purported special truth property (or properties), according to the deflationists. Such an appeal would just be a sort of "double counting":

> it is a kind of double counting to think that [the distinctions of truths] strike at the conception of truth involved. They strike at the level of the proposition: they mark distinctions of subject matter […]. But why add to a distinction of content, another, mirroring, distinction, one only applying to kinds of truth or conceptions of truth? (Blackburn 2013, 265).

For Blackburn, there is only one counting, as it were, and it has to do with the *content* of the propositions involved. A similar idea was already expressed by another important alethic deflationist—W.V. Quine—who, arguing against the thesis that the word "true" is ambiguous, claimed that

> There are philosophers who stoutly maintain that "true" said of logical or mathematical laws and "true" said of weather predictions or suspects' confessions are two uses of an ambiguous term "true".

---

[5] I will not take a stand on this question here, and so will bracket the philosophical differences involved in taking sentences, propositions or the like as truth-bearers. I will also speak in places of the disquotational version of the equivalence schema—'*p*' is true if, and only if, *p*—without calling attention to the distinctions relevant in choosing this version over the other.

[…] What mainly baffles me is the stoutness of their maintenance. What can they possibly count as evidence? Why not view "true" as unambiguous but very general, and recognize the difference between true logical laws and true confessions as a difference merely between logical laws and confessions? (Quine 1960, 131),

namely, again, as a difference in the content of the propositions expressed in different areas of discourse.

Quine's claim was recently echoed by Charles Parsons (cf. Parsons 2013, 194), and toward the end of the last century by Mark Sainsbury, according to whom

even if it is one thing for "this tree is an oak" to be true, another thing for "burning live cats is cruel" to be true, and yet another for "Buster Keaton is funnier than Charlie Chaplin" to be true, this should not lead us to suppose that "true" is ambiguous; for we get a better explanation of the differences by alluding to the differences between trees, cruelty, and humour (Sainsbury 1996, 900).

Discounting Quine's and Sainsbury's reference to the ambiguity of truth, Julian Dodd draws the following moral regarding the plurality of truth-apt discourses:

the sorts of differences between truths described by pluralists can be construed, not as differences in *the way* these propositions can be true, but as differences in the respective *subject matters* of these propositions […] the relevant difference in the truths […] is ultimately a difference concerning the things in the world they respectively concern, not in how they are true (Dodd 2013, 305-306).

What is relevant, in a nutshell, is the extra-linguistic ontological level, not the metalinguistic one: decidedly a straightforward and beautiful way to account for the pluralism inherent in truth-talk on the part of deflationism.

4. Is Putnam's a Pluralism of a Deflationary Variety?

Now, one may be tempted to attribute this line of reasoning to the latter Putnam, maintaining that his alethic pluralism is merely a pluralism *of content*. The temptation is strong, owing to a couple of claims made by Putnam, but I think that it does not reflect his actual stance.

Here is a possible source of this temptation. In the course of his criticism towards the metaphysical realism of Stages 1 and 2, he stated that

what makes the metaphysical realist's response *metaphysical* is its

acceptance of the idea […] that our ordinary realism […] presupposes a view of truth as a "substantive property" (Putnam 1999, 55).

Given that the thesis that truth is not a substantive property is a central tenet of deflationism, this claim may give the idea that Putnam subscribes to deflationism. However, on a closer reading it is possible to realize that what he was actually criticizing is not the notion of "substantive property" per se, but the idea that there is one and the same (substantive) property in *every* case of true statement. In fact, he was criticizing the metaphysical realist for postulating

that there is some single thing we are saying (*over and above what we are claiming*) whenever we make a truth claim, no matter what sort of statement we are discussing, no matter what the circumstances under which the statement is said to be true, and no matter what the pragmatic point of calling it true is said to be (Putnam 1999, 55; emphasis added).

The stress is here on the phrase "some single thing" that—according to the metaphysical realist's rendering of the truth-talk—would be in place in *every* true statement, irrespective of its subject matter. A single thing— i.e. the property of being in a relation of correspondence with a portion of reality—that would transcend the content of what we are saying when we simply assert a claim. Putnam's denial that there is such a thing brings grist to the alethic pluralist's mill, since it is natural to think that the underlying idea is here that there are many alethically potent properties, domain by domain and, moreover, they are *embedded* in what is said, not over and above it.

All this applies also to another source of the temptation to say that Putnam is a deflationist, namely his claim that "What is right in deflationism is that if I assert that 'it is true that *p*', then I assert the same thing as if I simply assert *p*" (Putnam 1999, 56). But, again, this is too poor a basis to warrant his being deemed a militant in the deflationary camp. Actually, this claim both is a restatement of the previous idea—i.e. that truth does not go beyond the content of a statement—and shows nothing else but the *disquotational property* of truth—which in turn is a general logical trait of truth, therefore taken into account by every interpretation of the concept. That Putnam's stance is far from deflationism is apparent in the following passage:

I believe that the disquotational property of "true" is an extremely important one, [but not that it] is all there is to say about truth, which is the characteristic thesis of what is called "deflation" (Putnam 2013b, 97).

Throughout his career Putnam was a fierce critic of alethic deflationism,

and perhaps it comes as no surprise that a favourite argument of his criticism had to do with the central question of how "language and thought 'hook on' to the world". In fact, he maintained that deflationism runs the risk of *losing the world*, because of its deplorable verificationist nature (cf. Putnam 1999, 53 ff)—more precisely, because of the *verificationist account of understanding* and *meaning* it requires. The reason is clear: having deflationism banned truth from the set of the philosopher's explanatory tools, truth-conditions appear useless—they cannot explain anything, let alone meaning and understanding. Given that one of the traditional competitors of truth-conditional semantics is verificationist semantics, to ascribe the latter to deflationism requires just one step. However—and this is the linchpin of his criticism—the verificationist account of understanding ends with bracketing or downright expunging the things in the world: to put it roughly, if "what exists" is being taken as "what exists for a subject *S*" (even a collective subject *S*) thanks to her best verificationist procedures, then the *idealistic danger* of losing the world becomes obvious.

I think that Putnam's is a remark in point, even though I find the reference to verificationism unnecessary in order to show that deflationism is at risk of losing the world. Let me briefly explain this before going back to Putnam.

Rather than being tied to verificationism, I think the risk in question is inherent in the Horwichian claim that truth is "merely a useful *expressive* device", a claim heavily suggesting that truth-talk has just to do with language (cf. Dell'Utri 2016). Consider the following typical deflationary allegation:

> To explain the utility of disquotation we need say nothing about the relation between language and the world. [Our theory of the concept of truth] seems to rest only on the most general *formal* features of our language—for instance, the fact that our language has somewhat the structure of quantificational languages—the utility for us of the concept of truth seems to be a fact which is quite independent of the existence or non-existence of interesting "picturing" or referential relations between our language and the world (Leeds 1978, 44).

Deflationists correctly detect the crucial point of any elucidation of truth in the capacity to offer a plausible explanation of the "referential relations between our language and the world", and quite reasonably see that taking these relations as having a "picturing" nature may cause more of a problem. But drawing from this the conclusion that truth has nothing to do with relations of some kind between language and the world, be they referential or not, appears a self-defeating move—even granting that these relations are not strictly required when it is just the explanation of the *expressive* utility of the word "true" that is at stake. And that move is

self-defeating just because of the losing-the-world issue: if truth and the world are detached one from the other in this way, then the solipsistic picture of individuals mechanically using language as if they were robots or brains in a vat imposes itself on us.

However, coming to Putnam, it is interesting that he eventually dropped the argument based on the purported tie between deflationism and verificationism, and argued for the idea according to which deflationists run the risk of losing the world along different lines. His argument may be succinctly reported as follows. In order to function properly, the disquotational schema presupposes the notion of translation, or sameness of meaning: there are plenty of cases in which the quoted sentence in one side of the biconditional belongs to a language different from the one in which the rest of the biconditional is couched, so that in the other side of the biconditional a translation of that sentence has to appear. But there are also plenty of cases in which among the constituents of the sentence in question there are words, and a correct translation of the sentence requires knowledge of what these words refer to. So, the notion of translation presupposes the notion of reference. To quote Putnam:

> That the notion of translation is needed for disquotation and therefore needed by deflationists (since their thesis is that grasp of disquotation is all that is needed for an understanding of truth) is widely recognized. But what I have not seen discussed by deflationists, let alone taken seriously, is the thought that translating sentences presupposes knowing what their descriptive constituents refer to. It is an illusion that disquotation does not presuppose the relation of reference (Putnam, 2015a: 324).

Notwithstanding their scant regard for reference, Putnam went on to remark, and despite the formal level on which they place the analysis of truth, deflationists unhesitatingly keep uttering claims such as "electrons really exist" and the like, taking for granted that the relevant words in these claims refer to actual entities out there, just as a realist philosopher would have it—as though it was perfectly obvious that this was so. Thus the following revealing moral:

> This "semantics-free" version of realism seems to amount to the claim that to be a realist it suffices to sincerely write or utter the right realist-sounding *sentences*, regardless of the account one gives of *what one is doing* by writing or uttering them (Putnam 2013c, 125).

Since the implicit accusation is that the deflationists do not have a plausible account to offer of how a speaker manages to talk of the worldly objects and states of affair their statements are about—because of their refusal to link the issue of truth with any substantive metaphysical

issue[6]—the upshot is a picture of human beings using language as if they were in a sort of void: again, as we noticed, as if they were brains in a vat. Hence, the loss of the world, a loss that the deflationists attempt to disguise "by means of a superficial terminological conservatism" (Putnam 1999, 55).

As to Putnam's own position, it is the direct realism of Stage 4 that does the job, as it were: it is the idea that we are directly connected to the world via perception, and the idea that perception is combined with our practical, intellectual and linguistic abilities—which come in a whole and are intrinsically *world-involving*—that ensure a reference to the terms we employ. Speaking a language is not a mere syntactic manipulation of empty symbols (cf. Putnam 1999, 49), phenomenologically appearing as marks and noises that we have to associate with senses. To the contrary,

> sentences that I think, and even sentences that I hear or read, simply do refer to whatever they are about—not because the "marks and noises" that I see and hear (or hear "in my head", in the case of my own thoughts) intrinsically have the meanings they have but because the sentence in use is not just a bunch of "marks and noises" (Putnam 1999, 46).

All this distances Putnam from deflationism and cooperates to shape his pluralist conception of truth, in the wake of the later Wittgenstein:

> Instead of looking for a freestanding property of "truth", in the hope that when we find what that property is we will know what the nature of propositions is and what the nature of their correspondence to reality is, Wittgenstein wants us to *look* at ethical language (and not the kind of ethical language that only occurs in philosophy), to look at religious language, to look at mathematical language, which is itself, he says, a "motley", to look at imprecise language that manages to be perfectly "clear" in context ("Stand roughly here"), to look at talk that is sometimes nonsensical and to look at the very same sentences when they function perfectly well (talk of "what is going on in so-and-so's head" is an example of this), to look and *see* the differences in the way these sorts of discourse function, *all the very different ways in which they relate to reality* (Putnam 1999, 68; emphasis added).

From this follows "a rejection of the idea that we can speak of one single 'truth predicate' whose meaning is fixed once and for all" (Putnam 1999, 68), a rejection which far from amounting to the claim that there are many truth predicates, one for each area of discourse, suggests the anti-Quinean idea we mentioned at the end of the third section, namely that there is an extendable family of *uses* of the predicate "true". But, it would

---

[6] Cf. "truth is metaphysically trivial" (Horwich 1998, 146).

be natural to ask, does this not amount to conceding the point to Quine? "Does Putnam, in saying that 'true' has a variety of uses, mean to imply that the word is systematically ambiguous" (Lynch 2001b, 618) in its meaning, showing indirectly that Quine was after all right in his criticism?

5. On the Purported Ambiguity of "true"

I do not think so. First of all, we would be allowed to say that we have here a case of *semantic ambiguity* only if it were reasonable to assume that the use of an expression *rigidly* determines its meaning, so that even the slightest change in the use counts as a change in the meaning of that expression—an assumption that fails to have even the faintest semblance of plausibility.[7] Secondly, a clear case of semantic ambiguity is given by homonyms such as "bank" or "step", which convey different meanings in different contexts in such a way that we have to learn those meanings separately; but there is no apparent reason to think that "true" belongs to this category. Indirect evidence of this is given by Lynch who, after giving the name *simple alethic pluralism* (SAP) to the view according to which the meaning of "true" is context-sensitive, claimed "I'm not sure anyone actually advocated SAP" (Lynch 2009, 54), thereby implicitly answering his own question quoted above. Thirdly, since every version of alethic pluralism has it that we use the term "true" in many ways, if this were at all a firm sign of ambiguity in the meaning of the term, then each alethic pluralism on the scene would immediately suffer from this kind of flaw—a possibility so implausible that it is not even worth mentioning. Fourthly, even when we want to stick to the meaning-centred reading represented by SAP, it should be stressed that a far better interpretation of this kind of alethic pluralism has been put forward (but not supported) by Wright, who detaches it clearly from any ambiguity case and more reasonably sees it as a case of *stretching* the use of the word "true", exploiting the *elasticity* of its meaning in a way that "you don't have to learn each type of use separately" (Wright 2013, 126).

In sum, Quine was too drastic in suggesting that whoever believes that there is an ever-growing family of uses of the word "true" renders this word semantically ambiguous. And this implies in turn that we are not obliged to follow Quine and take the differences between truths as *just*

---

[7] It would even run afoul of the famous Wittgensteinian claim to the effect that "For a *large* class of cases—though not for all—in which we employ the word 'meaning' it can be defined thus: the meaning of a word is its use in the language" (Wittgenstein 1953, § 43).

regarding their subject matters, thereby embracing a pluralism of content[8] (as Blackburn and Dodd also urge). In particular, as to Putnam, owing to the strong link he envisaged between the concept of the world and the concept of truth, these differences have repercussions on the concept of truth itself.

6. Conclusions

The upshot of the foregoing analysis is that Putnam's conception of truth belongs to the variegated constellation represented by alethic pluralism: as we have seen, he maintained that different kinds of statement are responsible to reality in their own way. This means that different uses of "true" are allowed by different properties, and these are all genuine—*substantial*, to use the anti-deflationist jargon, where the substance is normative in character and is given by the world, in the broadest sense of the word. Beside these substantial normative properties—correspondence, warranted assertibility, coherence etc.—there is the disquotational property which, owing to its formal character, allows us to use the word "true" across the board, revealing that the latter "belongs to the family of the logical words (for example, the connectives and quantifiers), which also are used in every area of discourse" (Putnam 2015c, 559-60). It is therefore the disquotational property that gives the concept of truth its unity.

The discussion about alethic pluralism has been very lively for decades, and the relevant literature does not fail to pick out the different problems encountered by the many pluralist proposals. Putnam did not address this literature, and it is hard to say which proposal he was most consonant with. However, it seems that, on the one hand, he was deeply aware of some of the problems in question, primary among them the difficulty "to do justice simultaneously to the plurality of our uses of 'true' and to the logical unity of the concept of truth" (Putnam 2015c, 560), and, on the other, he would not object to the idea according to which the many properties allowing the plurality of uses of 'true' represent a sort of ground on which the property of truth may be placed, with the overall

---

[8] A word of clarification is in order here. We might say that the pluralism of content is a sort of default position, owing to the "plurality" of contents being just a plain *fact*. Indeed, it is undeniable that the many statements we utter have a multifarious content pertaining to different areas of discourse. From this default position at least three stances may ensue: (1) a *deflationary stance*, according to which truth presupposes no relation between mind-language and the world (being the question of truth *only* a linguistic question); (2) a *metaphysical realist stance*, according to which truth involves a fixed relation between any statement and the world (making truth to go beyond the content of a statement); (3) a *pluralist stance*, according to which there are many ways in which a statement can be responsible to reality (ways that, in Putnam's case, do not make truth to go beyond the content of statements); and possibly others.

result that there is

> a single property of truth, and there are many other satellite properties hanging around in its vicinity which […] are somehow doing something to service the application of the truth-property (Wright 2013, 138).[9]

But whether or not this is a plausible interpretation of the picture of truth Putnam wanted to give in Stage 4 of his philosophical life is the topic of another paper.[10]

REFERENCES

Armour-Garb, B. P. and J. C. Beall, eds. 2005. *Deflationary Truth*. Chicago: Open Court.

Auxier, R. E., Anderson, D. R. and L. E. Hahn, eds. 2015. *The Philosophy of Hilary Putnam*. Chicago: Open Court (Library of Living Philosophers vol. XXXIV).

Auxier, R. E. and L. E. Hahn, eds. 2007. *The Philosophy of Michael Dummett*. Chicago: Open Court (Library of Living Philosophers vol. XXXI).

Baghramian, M., ed. 2013. *Reading Putnam*. Milton Park: Routledge.

Bilgrami A. 2007. Pursuing an Analogy. In Auxier and Hahn (eds.), 383-408.

Blackburn, S. 2013a. Deflationism, Pluralism, Expressivism, Pragmatism. In Pedersen and Wright (eds.), 263-277.

---

[9] In a nutshell, my hypothesis is that Wright's pluralism is the right alethic context in which to place Putnam's contention to the effect that the truth of a statement does not go beyond the content of that statement, so that there is nothing *over and above a truth claim*. This means that a statement can have an "alethically potent" property (say, coherence), which is a property embedded in it, and this property in turn moulds a ground for the truth-property—embedded in the same way. Moreover, according to both Putnam and Wright truth has a normative dimension. Again, this dimension does not go beyond the content of a true statement: it does not transcend the level of language placing itself in a metaphysical sphere (it rather belongs to the dimension of inquiry Akeel Bilgrami speaks of: cf. Bilgrami 2007); to use Putnam's wording, it is "a sort of correctness that is appropriate to the kind of statement" in question and varies with that kind. Being part and parcel of the concept of truth, also the normative dimension is embedded in the statement.

[10] I would like to thank David Brett, Stefano Caputo, Pietro Salis and an anonymous referee for their comments on a first draft of this paper.

Clark, P. and B. Hale, eds. 1994. *Reading Putnam*. Oxford: Basil Blackwell.

Dell'Utri, M. 2016. Truth and the Relevance of Practice. *Rivista Italiana di Filosofia del Linguaggio*. Special Issue: 22-37. DOI: 10.4396/2015SFL03

Dodd, J. 2013. Deflationism Trumps Pluralism!. In Pedersen and Wright 2013a, 298-322.

Hill, C.S. ed. 1992, *The Philosophy of Hilary Putnam*. In *Philosophical Topics* 20: 1-408.

Horwich, P. 1998. *Truth*. Oxford: Clarendon Press.

—————. 2016. Wittgenstein on Truth. In *Argumenta* 2: 95-105. http://www.argumenta.org/article/wittgenstein-on-truth.

Leeds, S. 1978. Theories of Reference and Truth. In Armour-Garb and Beall (eds.) 2005, 33-49. Chicago: Open Court.

Lynch, M. P., ed. 2001a. *The Nature of Truth: Classic and Contemporary Perspectives*. Cambridge (MA): The MIT Press.

———. 2001b. Primitivism, Identity Theory, and Alethic Pluralism: Introduction. In Lynch (ed.) 2001a, 615-21.

———. 2009. *Truth as One and Many*. Oxford: Clarendon Press.

Parsons, C. 2013. Putnam on Existence and Ontology. In Baghramian (ed.), 182-201.

Pedersen, N. J. L. L. and C. Wright, eds. 2013a. *Truth and Pluralism. Current Debates*. Oxford: Oxford University Press.

———. 2013b. Introduction. In Pedersen and Wright 2013a, 1-18.

———. 2016. Pluralist Theories of Truth. In *The Stanford Encyclopedia of Philosophy*, ed. E. N. Zalta, URL = <https://plato.stanford.edu/archives/spr2016/entries/truth-pluralist/>.

Putnam, H. 1960. Do True Assertions Correspond to Reality?. In Putnam 1975, 70-84.

———. 1975. *Mind, Language and Reality*, *Philosophical Papers*, vol. 2. Cambridge: Cambridge University Press.

———. 1978. *Meaning and the Moral Sciences*. London: Routledge and Kegan Paul.

———. 1981. Reason, Truth and History. Cambridge: Cambridge University Press.

———. 1983. On Truth. In Putnam 1994a, 315-29.

———. 1990. *Realism with a Human Face*. Ed. J. Conant. Cambridge (MA): Harvard University Press.

———. 1992. Reply to Gary Ebbs. In Hill (ed.), 347-58.

———. 1994a. *Words and Life*. Ed. J. Conant. Cambridge (MA): Harvard University Press.

———. 1994b. Simon Blackburn on Internal Realism. In Clark and Hale (eds.), 242-54.

———. 1994c. Sense, Nonsense, and the Senses. An Inquiry into the Powers of the Human Mind. *Journal of Philosophy* 91: 445-517. Repr. in Putnam 1999, 3-70.

———. 1999. *The Threefold Cord. Mind, Body, and World*. New York: Columbia University Press.

———. 2013a. From Quantum Mechanics to Ethics and Back Again. In Baghramian (ed.), 19-36.

———. 2013b. Comments on Richard Boyd. In Baghramian (ed.), 95-100.

———. 2013c. Comments on Michael Devitt. In Baghramian (ed.), 121-26.

———. 2015a. Naturalism, Realism, and Normativity. *Journal of the American Philosophical Association* 1: 312-28.

———. 2015b. Reply to Larry A. Hickman. In Auxier, Anderson and Hahn (eds.), 788-800.

———. 2015c. Reply to Frederick Stoutland. In Auxier, Anderson and Hahn (eds.), 547-63

———. 2015d. Reply to Charles Parsons. In Auxier, Anderson and Hahn (eds.), 134-43.

Quine, W. V. 1960. *Word and Object*. Cambridge (MA): The MIT Press.

Sainsbury, M. 1996. Crispin Wright: 'Truth and Objectivity'. *Philosophy and Phenomenological Research* 56: 899-904.

Thorpe, J. R. 2017. Closure Scepticism and The Vat Argument. *Mind*. https://doi.org/10.1093/mind/fzw035

Wittgenstein, L. 1953. *Philosophical Investigations*. Oxford: Basil Blackwell.

Wright, C. 2013. A Plurality of Pluralisms. In Pedersen and Wright (eds.) 2013a, 123-53.

CONCEPTUAL RELATIVITY MEETS REALISM IN METAPHYSICS*

TOMÁŠ MARVAN

Institute of Philosophy, Czech Academy of Sciences, Czech Republic

ABSTRACT

The paper adresses the relationship between ontological realism and Putnam's thesis of conceptual relativity. The paper divides into three parts. The first part aims to reconstruct the notion of conceptual relativity, focusing on Putnam's example involving mereological principles of individuation of objects. The second part points to some major shortcomings of the mereological example of conceptual relativity and then moves to a different version of conceptual relativity, which targets objects posited by mature scientific theories. I claim that the mereological and the scientific version of conceptual relativity are different in important respects and that two main types of conceptual relativity therefore need to be distinguished. In the third part, I show that conceptual relativity is not in tension with realism. More specifically, conceptual relativity is not in tension with "realism in metaphysics" that Putnam adopted in the last decade before his death.

Keywords: Hilary Putnam, conceptual relativity, realism, optional languages

1. Conceptual relativity: The Case of Mereology

Although Putnam never ceased to be an ontological realist, believing in the realm of mind-independently constituted objects, he was relentlessly pointing out unclarities of realistic metaphysics. In this section, I will introduce the curious phenomenon of "conceptual relativity" (Putnam's coinage), which calls into question one of the assumptions of an uncritical form of realism.

First, *un peu d'histoire*. In the period from mid seventies to late eighties,

Putnam was advocating "internal realism", which was in opposition to "Metaphysical Realism" (based on the presumption of unique correspondence relation between words and a fixed "ready-made" world of external things and properties).[1] Apart from the emphasis on the epistemic notion of truth and its concomittant idea of idealized rational acceptability, internal realism had a distinctly constructivist flavour. Internal realist holds that we, human beings, "cut up the world into objects when we introduce one or another scheme of description" (Putnam 1981, 52). Objects thus conceived are scheme-dependent. They are logically mind-dependent: their existence implies existence of a mind using one or another conceptual scheme. Putnam later recanted these verificationist and constructivist notions and came to adopt a more traditional realist position. That is, he accepted that there can be truths that outstrip all our means of verifying them and rejected all constructivist talk about the mind "making" the world. After adopting "direct realism" in the philosophy of perception from the early nineties onwards (Putnam 1994a), he eventually embraced "realism in metaphysics" in his most recent writings on realism, starting with the lecture From Quantum Mechanics to Ethics and Back Again (delivered in 2007; reprinted in Putnam 2012; see also Putnam 2016a, 24–27). There is, however, one assumption of Metaphysical Realism that Putnam never accepted. It's the idea that there is only one correct and complete description of reality. Putnam's late brand of realism embraces the possibility to describe the world correctly in different ways; in a word, it embraces conceptual relativity.

Conceptual relativity, "the heart of internal realism" (Putnam 1991, 404) thus survived the collapse of other ingredients of internal realism. It reappears in Putnam's more recent writings such as Putnam (2004) or Putnam (2012). Conceptual relativity is to be distinguished from conceptual *pluralism*, another tenet adopted by Putnam. Conceptual pluralist draws attention to cases in which two quite distinct schemes, such as the scheme of particle physics and the ordinary language of chairs and tables, describe the same portion of reality – say, the furniture in a study. On this view, the talk of chairs and tables truly describes what is out there and should not be seen as inferior to the physical description of the same portion of reality; science is not the only discourse which states "the facts" (Putnam 1994b, 243; Putnam 2004, 48). Putnam, of course, admits that chairs and tables are composed of particles described by physics. These different schemes, however, deal with different "levels of reality". For this reason, everyday claims are not reducible to the statements of physics. In most radical cases of conceptual plurality, we cannot even conceive how we would go about reducing one description into another: think of physical theory and literary criticism. Conceptual

---

[1] Throughout the text, I use "Metaphysical Realism" with capitalized letters as a name for the position Putnam was critical of.

plurality thus involves irreducible but compatible descriptions of the same states of affairs. The reverse is true of conceptual relativity.

The doctrine of conceptual relativity most centrally consists in the claim that "in certain cases what exists may depend on which of various conventions we adopt" (Putnam 2004, 39). Sometimes we can describe the same state of affairs in two different ways and it's not possible to conjoin the two descriptions, because the result would be incoherent. In these cases, there is no fact of the matter as to which of the descriptions is *really right* and nothing is necessitating the choice of one of the descriptions over the other. These non-conjoinable but correct descriptions, though, are only "incompatible at face value". There is no genuine incompatibility between them, for the statements can be converted into each other. Yet, they do not preserve the same "ontology": they do not see the world as composed of exactly the same objects. Thus, conceptual relativity encroaches on basic ontological notions such as "object".

The best way to get a grip on Putnam's idea is to look at his examples of conceptual relativity. The one most frequently used by Putnam, to be found in his (1987) and various other writings, in various variants, concerns mereology and its way of individuating objects – a way that contrasts with more familiar individuating strategies. Imagine a mini world – "Carnapian world" in Putnam's terms – consisting of just three individuals, $x1$, $x2$ and $x3$. According to most of us, I guess, the Carnapian world consists of precisely three objects. However, suppose that "Polish Logician", a champion of mereology, looks at the same world.[2] In his view, it consists of *seven*, not three objects. These are, respectively:

$$x1, x2, x3, x1 + x2, x1 + x3, x2 + x3, x1 + x2 + x3.$$

Who is right, then? How many objects *really are there* in the Carnapian world? Putnam insists that these questions have no sense. We cannot determine the number of objects in the world before it is established which conceptual apparatus we are drawing on in counting the objects. Once the apparatus is fixed, the question concerning the number of objects acquires a clear meaning and can be answered. Whether we talk "atomistically" or mereologically is a matter of choice. The Carnapian world doesn't dictate to us in which scheme it should be described.

Now, the two statements concerning the number of objects certainly aren't mutual *translations* of each other in the ordinary sense in which "There are three objects in the Carnapian world" and "Il y a trois objets

---

[2] Mereology is the calculus of parts and wholes based on the principle that for every two particulars there is an object which is their sum. The "Polish Logician" is Putnam's allusion to Stanisłav Leśniewski, the author of the first formal part-whole theory. See Putnam (1990, 96).

dans le monde carnapéen" are mutual translations. But the two sentences from the alternative conceptual schemes are "in deep mutual relation" (Putnam 1987, 20). To begin with, there is a simple procedure of converting the number of atomistic objects into the number of mereological objects: if the number of atomistic objects is n, then the number of mereological objects is $2^n - 1$. The existence of such a method of "translation" is an essential feature of conceptual relativity: Putnam rejects the picture of two correct but not convertible descriptions of the world (Putnam 1983, 40). Consider another mereological variant of conceptual relativity featuring alternative *predicates*, viz. the colours of objects (Putnam 1990, 98f.). Suppose that x1 is red and x2 black. The mereologist will claim that the Carnapian world contains an object which is partly red and partly black (i. e., x1 + x2). The atomist will deny this (provided, that is, that x3 is not partly red and partly black). There is, though, a method of interpreting the Polish Logician's sentence "There is an object which is partly red and partly black" in the atomistic scheme: „There is an object which is red and a different object which is black". Both sentences correctly describe the same state of affairs. In this manner, we could reinterpret all colour attributions of one scheme in the other one. This reinterpretability of alternative descriptions generalizes across the board to all cases of conceptual relativity, whatever their subject matter.

Technically put, the two alternative descriptions of the Carnapian world are "mutually relatively interpretable". Theory T_1 is relatively interpretable in T_2 if there are formal definitions of the terms of T_1 in the language of T_2 with the property that, if we "translate" the sentences of T_1 into the language of T_2 by means of those definitions, then all theorems of T_1 become theorems of T_2. Two theories are mutually relatively interpretable if each is relatively interpretable in the other (Putnam 1983, 38). Mutual relative interpretation, though, as is clear from the above definition, concerns only the formal properties of two conceptual frameworks. In this formal fashion, two frameworks might be relatively mutually interpretable even if they deal with completely disparate domains – say, one is an axiomatic system of genetics while the other an axiomatic system of number theory (Putnam 1983, 38). This is not the case in Putnam's mereological example. The atomistic statement and the mereological statement both describe the same domain, viz. the Carnapian world. The two descriptions are thus "cognitively equivalent" in the following sense: they are describing the same state of affairs and all their predictive and explanatory powers are equal.

The notion of the shared domain of the two descriptions is a prerequisite of the superficial, "at face value" incompatibility of the descriptions. If the descriptions described different portions of reality, they would not be incompatible in any way and could be conjoined into a single true decription of the overarching WORLD (Lynch 1998, 29–30). But how to

secure the same domain for the superficially incompatible schemes of description?

Putnam is clearly aware of the need for a common domain of the two descriptions (see Putnam 1991, 406, and Putnam 1992, 185). At the same time, he occasionally qualifies the claim about the shared domain with scare quotes: he talks about "the 'same' world" (Putnam 1987, 20) and "(in some way) the 'same facts'" (Putnam 1987, 29). But this can't be right. The underlying reality common to both descriptions must *really* be the same, not just "as if" the same. Without this, the descriptions could not be incompatible, not even in the superficial sense Putnam has in mind. If the sentences "There are three objects" and "There are seven objects" describe domains that are in any respect different, they cease to be superficially incompatible.[3]

But how can we flesh out the idea of the common domain? There cannot be a third, neutral description of the Carnapian world which would incorporate the two competing accounts, for this would dissolve conceptual relativity. Yet, if we fail to provide some scheme-independent route to common underlying reality, how can we claim that the two descriptions capture the same facts? All that has been produced are just two different renderings of what is – without ground – claimed to be the same state of affairs. Could sensory perception help us with this problem? It could, but it doesn't seem to be necessary to fix the underlying state of affairs. In fact, I introduced the mereological example without drawing on sensory perception.[4] Therefore, I suggest that we grasp the fact that the two descriptions describe the same domain – that we just "get it" – even if the descriptions themselves not only do not state that they share a domain but they even appear to be incompatible. This is an important result. It shows that there are things we can say on the basis of the competing descriptions even if they are not explicitly contained in the descriptions themselves.[5]

As to the idea of the incompatibility at face value, one could try to unpack it in the following way: the two sentences, "There are three objects in the Carnapian world" and "There are seven objects in the Carnapian world", are incompatible only at first blush, because there is

---

[3] In more recent texts, Putnam dropped the qualifications and spoke simply of the same facts or the same state of affairs. See, e. g., Putnam (2016, 153–154).

[4] Moreover, perception would be of no use in the case of unobservable scientific entities, which figure prominently in some of Putnam's conceptual relativity arguments. We will get to these entities in the following section.

[5] In defending the common domain of the two descriptions, Putnam wants to avoid a more extravagant metaphysics according to which both "versions" of the Carnapian world literally describe two different worlds, as was famously claimed by Nelson Goodman (1978). See Putnam (1992, 122).

no such thing as "a 'proposition' which one of these sentences affirms and the other denies" (Putnam 1991, 404). When we look more closely, we realize that the word "object" doesn't have the same meaning in both descriptions. Rather, its meaning is determined with the help of the relevant contextual parameter, viz. the framework adopted (atomistic vs. mereological). Relative to the atomistic scheme, "object" means something else than what the same word, syntactically speaking, means relative to the mereological scheme. The tension between the two descriptions vanishes into thin air. Both can be true at the same time.

However, Putnam is adamant that this explanation is incorrect. It is wrong, he believes, to view the shift between the two descriptions as a shift in meaning. Rather, he suggests, what is involved is a difference in *use* of the term "object". "The ordinary notion of 'meaning' was simply not invented for *this* kind of case" (Putnam 1991, 405). The two uses of the word "object" do not deserve two separate dictionary entries (Putnam 1994a, 451f.). Putnam's favorite example of this occasion-sensitivity of discourse involves coffee: think about the sentence "There is too much coffee on the table" used on occasions when there is (1) a number of mugs full of coffee on the table, (2) a lot of spilled coffee on the table and (3) a lot of bags of coffee beans on the table (Putnam 1999, 87–88). The extension of "coffee" is somewhat different in these three instances of "There is too much coffee on the table", yet the core meaning of "coffee" is preserved in all of them. Thanks to this occasion-sensitivity, some concepts are semantically "extendable". We need these "broad-spectrum notions" when we lack more precise terms and when we try to intelligibly explain those more precise terms while introducing them (Putnam 2012, 68). According to Putnam, the notion of object is extendable in this sense. The atomist and the mereologist do not use different concepts of objects; they use the same concept in somewhat different ways.[6]

2. Beyond Mereology: Scientific Objects

The argument for conceptual relativity from mereology will only work if we accept mereological scheme as a full-blown ontological alternative to atomism. But should we? There are reasons which count against such ontological tolerance. To begin with, Putnam himself notes that mereological ontology is *profligate*. It accepts all the objects accepted by the atomistic description and adds to them a couple of weird ones. This lavishness leads to unwelcome consequences. Suppose, to use Putnam's example, that object *a* is a body of a lamp (including the bulb) and object

---

[6] Putnam claims that the notion of existence is extendable, too. He thinks that mathematical and physical entities exist in different senses and that this is also true of the mereological and atomistic objects. See Putnam (2004, 240).

b its detachable shade. Then, according to the mereological way of counting, the room contains the objects *a*, *b* and *a + b*. But since "*a + b*" stands for "bulb-containing body of a lamp plus its detachable shade", the mereologist in fact claims that there are *two* lamps in the room – a claim that can be empirically disproven. Few of us are attracted to accepting such a realm of shadowy *Doppelgängers*.[7]

Another complaint against mereological criteria of objecthood is due to Peter van Inwagen. It concerns the role of convention in ontology. Putnam's claim that what there is is in some cases partly a matter of convention sounds just incredible to van Inwagen. To postulate, by fiat, that mereological sums exist is, in his eyes, equal to postulating that Golden Mountain exists. Thus van Inwagen firmly denies mereological ontology of the Polish Logician's stripe: there just *aren't* such objects as a mereological sum of a cat and a dog, he insists (van Inwagen 2002, 192). But even if we decided to include mereological sums into our inventory of what exists, they will still be in two respects importantly different from ordinary individuals. In the first place, *sums* can hardly be imagined to exist independently of a mind that assembles them. They thus violate Putnam's requirement of logical independence, which, he claims, belongs to objects as conceived by realists. Sums do not exist in the same mind-independent way as ordinary objects. Secondly, if I read Putnam correctly, the existence of mereological sums is based on ontological convention, but the existence of ordinary objects is not. In fact, I am not sure what would be the ontological convention for ordinary objects. To call an object object? I don't believe it would occur to anyone to call this a convention. The mereological ontological convention, on the other hand, is clear enough: *for every x and every y there is their mereological sum z = x + y* (Raatikainen 2001, 172). Ontological convention thus concerns only what Jennifer Case calls "optional languages" (Case 1997). A natural language such as English can harbour any number of optional languages which are (temporarily) adopted for specific purposes. Optional language is an extension of some more basic language. In the case at hand, mereological description is a consciously adopted extension of the more basic atomistic description. Putnam endorses this view in (Putnam 1994a, 451n13); in (Putnam 2012, 57f.) he adds that the existence of mereological sums is conventional in the sense that all facts can be expressed with or without them.

Consider now yet another objection to mereological ontology, due to Smith (2004, 79): mereological sums (Smith calls them "junk particulars"), such as the sum of my nose and of the Eiffel Tower, do not instantiate universals standing to other universals in relations captured by scientific laws. Therefore, they are not in all respects equal to atomistic

---

[7] As Varzi (2000, 287) argues, a well crafted inventory must be, on the one hand, complete, but shouldn't be, on the other hand, redundant.

objects, which do instantiate universals standing to other universals in relations captured by scientific laws. Again: mereological sums are importantly different from ordinary objects and if Putnam's arguments for conceptual relativity were based solely on them, we would have good grounds to question their relevance for ontology. But mereological sums are only the most well-known example of conceptual relativity. Putnam has other examples to offer. Take geometrical points. These, Putnam tells us, can either be seen as concrete particulars occupying a portion of space-time, or as "mere limits" (for technical details, see Putnam 1992, 217n14). Similarly, a theory which represents the physical interactions between bodies in terms of action at a distance and a physical theory which represents them in terms of fields may both be right (Putnam 1990, 40); or: relative to one conceptual framework of fundamental physics, reality consist of physical particles, relative to a different framework it consists of fields (Putnam 1992, 121). In his most recent writings, Putnam favored yet another example of conceptual relativity taken from scientific practice: statements about bosons and statements about fermions can be taken as two equivalent represenations of the same quantum mechanical system – the physicists call this phenomenon "duality". Informally put, there are known ways to convert statements talking about fermions into statements about bosons, and vice versa.

Putnam claims that examples of conceptual relativity are ubiquitous in mathemathical physics (2012, 63). I will stick with bosons and fermions. A quantum field model constructed with fermions is "bosonized" (in a specified number of spacetime dimensions) when it is reformulated in terms of a model which is equivalent but constructed exclusively from bosons. By the same token, a system consisting solely of bosons can be "fermionized". From the mid seventies a number of different mathematical techniques emerged that allow for such transformations. This is surprising, given that bosons and fermions have very different properties. They, crucially, differ in their spin quantum numbers. Fermions such as neutrons and quarks have half-integer spins, while bosons (pi mesons, photons etc.) have integral spins. As a consequence of this, systems contaning fermions behave in different ways than systems containing bosons. E. g., bosons can all be in the same quantum state. Fermions can't: Pauli exclusion holds for them, and if this were not so, periodic table of elements would look very different. Fermions are matter particles while bosons are force carriers. And so on.

Despite all these differences in their ontologies and "ideologies" (i.e., the predicates used), bosonic and fermionic schemes of description preserve all observations and can account for them in fully equivalent ways. Now, as we have observed, from the possibility of construing new schemes of individuation à la Polish Logician it just does not follow that they are equally good representations of the external states of affairs as any old vocabulary. But the case of peculiar objects populating quantum physical

theories is different. Undeniably, cases such as fermion to boson conversion do exist in science. Fermions and bosons do not fall prey to Smith's objection, for they do instantiate universals standing to other universals in relations captured by scientific laws. And we cannot just dismiss either fermions or bosons in the way that van Inwagen dismissed mereological sums.

The differences between the two examples of conceptual relativity, mereological and quantum mechanical, are not accidental, but reflect a deeper, systematic dissimilarity. I submit that we need to distinguish two types of conceptual relativity. In the CR_1 type, one of the alternative languages is optional, in Case's sense. Mereological example is of this type, because the individuating scheme of mereology is adopted via a special ontological convention. In CR_2 cases, neither of the languages is optional. Bosons and fermions, together with other concepts used in physics and elsewhere, fall under this second type of conceptual relativity. Bosons and fermions are both fully self-standing, ontologically speaking. None of them is an optional extension of some other, more basic notion. Bosons are not just variants of fermions, or vice versa.[8]

There are other divergences between the two types of conceptual relativity. Most importantly, the fermion–boson duality does not conform to the *relativity-of-use* template characteristic of the mereology example (and possibly other cases of CR_1). This is the schematization of the use relativity in the mereological example: relative to the relativizer Γ (atomistic scheme of individuation), the word "object" is used in a way α; relative to Δ (Polish Logician's scheme of individuation), the same word is used in a way β. The word "object" preserves its core meaning on both occasions of use, although Γ-objects and Δ-objects somewhat differ. In contrast with this, "boson" and "fermion" are two words with distinct meanings, fixed by the physical theory. No "extendability" of concepts and no "incompatibility at face value" is at work here.

3. Conceptual Relativity and Realism

Despite the differences between the two types of conceptual relativity, CR_1 and CR_2, statements of physical theory about bosons and statements about fermions are "mutually relatively interpretable"; they are two different ways of describing the same situations. This is the core of conceptual relativity, as conceived by Putnam. We should, therefore, accept Putnam's quantum mechanical example and other cases of CR_2 as bona fide instances of conceptual relativity, even if they do not conform to the relativity-of-use template and even if there is nothing conventional

---

[8] It is possible – though I don't know how to argue for this in a principled way – that all instances of CR_2 concern scientific objects and that there are no CR_2 examples of commonsense objects.

about the existence of fermions or bosons. We have a genuine choice whether to describe the same quantum mechanical system either as composed of fermions or as composed of bosons.

This fact, on the face of it, is more disconcerting then the suggestion that objects can be individuated atomistically or mereologically – that there is no philosophically priviledged sense of "object", as Putnam put it (Putnam 1995, 303). Is it not deeply puzzling that we can swap two quite different objects for each other while describing the very same part of external reality? Isn't the physical reality itself to a certain extent *indeterminate* because of this? Not according to Putnam. He points out that according to the physicists, the possibility of boson-for-fermion exchange shows that the ontology of the quantum mechanical theory of a particular system is not the "load-bearing aspect" of the quantum mechanical scheme (Putnam 2012, 57). The quantum mechanical scheme has alternative "representations", including the fermionic and the bosonic representation. The conclusion that Putnam and the physicists draw from this is that bosons and fermions are "simply artifacts of the representation used" (Putnam 2012, 64). So the picture is this. There is an underlying quantum reality. It is constituted independently of all our observations, schemes of decriptions etc. We devise concepts in order to variously describe this independent reality. This is a constructive activity, but it does not affect what is described (Putnam 2012, 62). The concepts devised are of such a nature that in certain cases we can swap one for another while not disturbing the equivalence of the alternative descriptions and the fact that the same state of affairs is described by both of them.

What justifies Putnam in holding that *the same* quantum mechanical system is described by bosonic and fermionic scheme? This is the same worry that the mereological example of the CR_1 variety had to face. But in the case of the fermions and bosons it is, I believe, more difficult to come up with a satisfactory answer to the question concerning the sameness of domain of both quantum mechanical descriptions. The reason is that the behavior of a bosonic system is so different from the behavior of a fermionic system. Putnam didn't see this as a challenge. According to him, we just *know* that nothing in physical reality is changed when we move from one quantum mechanical description to another; all that changes pertains to the representation of the system. But do we know this? The fact of mutual interpretability of the descriptions is not a decisive argument for the sameness of their domain, for this interpretability could be purely formal: both descriptions could, in fact, describe different portions of physical reality. I don't know how to answer the common-domain worry in the case of fermions and bosons. Note, however, that since the two descriptions are not incompatible in any respect, even if the assumption of the shared domain was threatened, this would not undermine the scientific realism Putnam espouses. That is, even if we are unable to fix the common domain, this is only a problem

for conceptual relativity itself (for it needs the common domain as one of its prerequisites). It is therefore incumbent on the defenders of conceptual relativity to find a way of securing the common domain of alternative scientific descriptions.

I conclude that instances of CR_2, exemplified in this paper by bosons and fermions, do not constitute a genuine challenge to realistic metaphysics of independently constituted objects and states of affairs. We do not choose *what exists*, we only choose *how to describe* what exists anyway. This is good news for advocates of realism. We can disentangle conceptual relativity from ontological constructivism with which it was aligned when it was formulated in the eighties in Putnam's writings. Later Putnam firmly rejects the "*Internal Realist General Ontological Thesis*: The world consists of theory-dependent objects" (Gardiner 2000, 146). In one of the many retrospective summaries of his changing conceptions of realism, Putnam writes that there are aspects of reality unaffected by human interests and constructions, such as the fact that there are thousands different species of ants in the world. These aspects would have remained unaffected even if humans didn't devise the label "ant", and indeed even if no humans with their concepts ever came into existence (Putnam 1994a, 448n7). The same holds with respect to the concepts of mature science such as "fermions" and "bosons".

Putnam's conceptual relativity is thus fully compatible with the principles of realistic metaphysics.[9] This annihilates the appeal of conceptual relativity for genuine ontological relativists. In a nutshell, genuine ontological relativists argue that „Soandso's exist" are to be understood as a claim that „Soando's exist relative to a particular conceptual scheme", as Nicholas Wolterstorff put it (Wolterstorff 1987, 239). From the epistemological point of view, Putnam's claim is not that we cannot ever get to the objects themselves; the claim is that sometimes we can get to them in different ways. This element of choice does not usher in constructivism: external physical reality is not an amorphous blob waiting to be cut up by us in various ways. It is ready-made, consisting of self-demarcated objects, properties and relations. Our perceptions, together with the very fact of the perfect mutual interpretability of the rival descriptions of the same states of affairs and events, confirm this inherent structuredness of reality.

We could express the same point by saying that nature dramatically limits, or controls, our ways of describing it. But wait, wasn't Putnam *criticizing* the idea of One True Description of the world? Indeed he was. However, his examples of conceptual relativity dislodge One True Description only when this idea is taken literally – as a description permitting no alternative expression whatsoever. In fact, probably no one

---

[9] Some of Putnam's perceptive readers were pointing this out for some time now: see Horgan and Timmons (2002).

is prepared to commit to such an extreme idea. Even staunch realists like John Searle (1995) and Robert Kirk (1999) admit that there might be variations in the correct descriptions of the world underlying all descriptions. A more liberal understanding of One True Description is therefore to be preferred. There is just one such decription, but some portions of it can be expressed in alternative, though mutually convertible, ways.

As to whether Putnam's arguments dislodge the idea of a *complete* description of the world, which was a part and parcel of Metaphysical Realism he was rejecting, the following observation is in order. Leaving aside the worry that the idea of a complete description of anything, let alone of all reality, is meaningless (Hacking 1983, 93), what is incomplete about, say, the atomistic description of the Carnapian world? Shall we say it is incomplete because it leaves out the mereological sums of objects? This would be confused. In its own way, the atomistic description captures all there is to the Carnapian world. According to the mereologist, the same world can be captured in a different way, but that doesn't mean that the atomistic description is incomplete, that it leaves anything out while specifying the number of objects in the Carnapian world. The same holds for the fermion/boson duality and other instances of CR_2. There might be another sense in which the decriptions are incomplete, but Putnam never specified it. I thus submit that conceptual relativity *per se* does not undermine the possibility of completeness of One True Description of reality (liberally understood).

4. Conclusion

Symbolically, Putnam's last published paper deals with the subject of realism (Putnam 2016b). I say symbollically because realism was one of Putnam's philosophical preoccupations for decades. He was unique in thinking through aspects of realism that other philosophers (and scientists) were taking for granted. His writings on conceptual relativity are a prime example of this. The central insight of these writings – that we can describe the same state of affairs in somewhat different ways – is important, although both actual and imagined instances of conceptual relativity are confined to special contexts, leaving most of both commonsense and scientific discourse untouched.

Putnam's grapplings with realism are instructive: after experimenting with various forms of realism, he finally settled on straightforward metaphysical realism, albeit one that is compatible with various forms of conceptual relativity. In closing, I will quote the words of David Lewis who, it seem to me, captured the fruits of Putnam's tenacious rethinking of realism:

It is the profession of philosophers to question platitudes that others

accept without thinking twice. A dangerous profession, since philosophers are more easily discredited than platitudes, but a useful one. For when a good philosopher challenges a platitude, it usually turns out that the platitude was essentially right; but the philosopher has noticed trouble that one who did not think twice could not have met. In the end the challenge is answered and the platitude survives, more often than not. But the philosopher has done the adherents of the platitude a service: he has made them think twice (Lewis 1969, 1).

In Putnam's case, the platitude in question was the view that external things out there really exist independently of all our perceptual and cognitive contributions. Over decades and even centuries, this platitude proved suprisingly difficult to defend, with many clever thinkers advocating its exact opposite. In the thesis of conceptual relativity, Putnam put his finger on what is worthwile in the complaints of some of the critics of straightforward realism, while cutting off the philosophical deadwood.[10]

REFERENCES

Case, J. 1997. On the Right Idea of a Conceptual Scheme. *The Southern Journal of Philosophy* 35 (1): 1–18.

Clark, P. and B. Hale, eds. 1994. *Reading Putnam*. Oxford: Blackwell.

De Caro, M. and D. Macarthur, eds. 2012. *Philosophy in an Age of Science*. Cambridge, Mass.: Harvard University Press.

De Caro, M., ed. 2016. *Hilary Putnam: Naturalism, Realism, and Normativity*. Cambridge, Mass.: Harvard University Press.

Gardiner, M. Q. 2000. *Semantic Challenges to Realism. Dummett and Putnam*. Toronto: University of Toronto Press.

Goodman, N. 1978. *Ways of Worldmaking*. Indianapolis: Hackett.

Hacking, I. 1983. *Representing and Intervening*. Cambridge: Cambridge University Press.

Horgan, T. and Timmons, M. 2002. Conceptual Relativity and Metaphysical Realism. *Philosophical Issues 12, Realism and Relativism*: 74–96.

Kirk, R. 1999. *Relativism and Reality. A Contemporary Introduction*. London and New York: Routledge.

Lewis, D. 2002. *Convention: A Philosophical Study*. Second edition. Oxford: Blackwell Publishers.

Lynch, M. 1998. *Truth in Context*. Cambridge, Mass.: The MIT Press.

Putnam, H. 1981. *Reason, Truth and History*. Cambridge: Cambridge University Press.

———. 1983. *Realism and Reason. Philosophical Papers Volume 3*. Cambridge: Cambridge University Press.

———. 1987. *The Many Faces of Realism*. La Salle: Open Court.

———. 1990. *Realism with a Human Face*. Cambridge, Mass.: Harvard University Press.

———. 1991. Reply to William Throop and Catherine Doran. *Erkenntnis* 34 (3): 404–408.

———.1992. *Renewing Philosophy*. Cambridge, Mass.: Harvard University Press.

———.1994a. Sense, Non-sense and the Senses. *The Journal of Philosophy* 91 (9): 445–517.

———.1994b. Simon Blackburn on Internal Realism. In Clark and Hale (eds.), 242–254.

———.1995. *Words and Life*. Cambridge, Mass.: Harvard University Press.

———.1999. *The Threefold Cord: Mind, Body and World*. New York: Columbia University Press.

———.2004. *Ethics without Ontology*. Cambridge, Mass.: Harvard University Press.

———.2012. *Philosophy in an Age of Science*. In de Caro and Macarthur (eds.).

———.2016a. *Naturalism, Realism, and Normativity*. In de Caro (ed.).

———.2016b. Realism. *Philosophy and Social Criticism* 42 (2): 117–131.

Raatikainen, P. 2001. Putnam, Languages and Worlds. *Dialectica* 55 (2): 167–174.

Searle, J. 1995. *The Construction of Social Reality*. Harmondsworth: Penguin Press.

Smith, B. 2004. Beyond Concepts: Ontology as Reality Representation. In *Formal Ontology and Information Systems. Proceedings of FOIS 2004 – Third International Conference on Formal Ontology and Information Systems,* eds. Achille Varzi and Laure Vieu. Amsterdam: IOS Press: 73–84.

van Inwagen, P. 2002. The Number of Things. *Philosophical Issues 12, Realism and Relativism*: 176–196.

Varzi, A. 2000. Mereological Committments. *Dialectica* 54 (4): 283–305.

Wolterstorff, N. 1987. Are Concept-Users World-Makers?. *Philosophical Perspectives 1, Metaphysics*: 233–267.

ARGUING ABOUT REALISM: ADJUDICATING THE PUTNAM-DEVITT DISPUTE*

JADE FLETCHER
University of Leeds, United Kingdom

ABSTRACT

In this paper I want to adjudicate the dispute between those philosophers who do and those who do not think that the philosophy of language can illuminate metaphysical questions. To this end, I take the debate between Devitt and Putnam as a case study and diagnose what I take to be illuminating about their disagreement over metaphysical realism. I argue that both Putnam and Devitt are incorrect in their assessment of the significance of the model theoretic argument for realism. That, whilst Devitt is entitled to claim that truth does not have anything to do with realism, Putnam's challenge can still gain traction and seriously call into question our ability to engage in realist metaphysics. I argue that even if a completely semantically neutral conception of realism can be successfully articulated, doing so has the potential to bankrupt the methodology of metaphysical realism. Having taken this debate as a case study, I then offer some brief remarks on how to understand the philosopher who claims that realist metaphysicians should care about discussions of metasemantics and truth. Whilst I want to be cautious about generalising on the basis on this case alone, I think there are important lessons to be learned about the way in which considerations to do with language can shed light on the concerns of metametaphysics.

Keywords: *Metaphysics, Metametaphysics, Reference, Indeterminacy, Truth, Hilary Putnam, Michael Devitt*

1. Introduction

Putnam claims that his model theoretic argument "has profound implications for the great metaphysical dispute about realism which has always been the central dispute in the philosophy of language." (Putnam

1980, 464) Putnam's claim captures the kind of view about the relationship between metaphysics and language that I interrogate in this paper. Specifically: the view that the question of realism is not exclusively the remit of metaphysics, but rather is a problem in the philosophy of language. If this were the case, then philosophical problems pertaining to truth, semantics, and our systems of linguistic and mental representation would indeed be relevant to discussions of metaphysical realism. There is a tendency in some contemporary discussion of realism to think that this kind of view is outdated, and that it merely derives from a conflation of the philosophy of language and metaphysics. Such philosophers hold that considerations to do with language cannot determine or limit the enterprise of metaphysics: that reflecting on problems in metasemantics cannot deliver any insights about metaphysical realism.

In this paper I want to progress the dispute between those philosophers who do and those who do not think that the philosophy of language can illuminate metaphysical questions. To this end, I take the debate between Devitt and Putnam as a case study and diagnose what I take to be important about their disagreement. I argue that both Putnam and Devitt are incorrect in their assessment of the significance of the model theoretic argument for realism. That, whilst Devitt is entitled to claim that truth does not have anything to do with realism, Putnam's challenge can still gain traction and seriously call into question our ability to engage in realist metaphysics. I argue that even if a completely semantically neutral conception of realism can be successfully articulated, doing so has the potential to bankrupt the methodology of metaphysical realism. Having taken this debate as a case study, I then offer some brief remarks on how to understand the philosopher who claims that realist metaphysicians should care about discussions of metasemantics and truth. Whilst I want to be cautious about generalising on the basis on this case alone, I think there are important lessons to be learned about the way in which considerations to do with language can shed light on the concerns metametaphysics.

The paper proceeds as follows. In section two, I present Putnam's model theoretic argument. I argue that the argument, if successful, is capable of motivating a radical metametaphysical thesis: that realist metaphysics cannot be sensibly pursued (§ 2). I engage with Michael Devitt's criticisms of Putnam's attack on realism. I entertain the possibility that Putnam's model theoretic argument is not in good standing as his characterisation of metaphysical realism is incorrect. Thus, given a more apt characterisation of the commitments of metaphysical realism, Putnam's arguments are unsuccessful in establishing the robust metaphysical conclusions he has in mind (§ 3).

I then evaluate Devitt's criticisms of Putnam along two dimensions. First, I briefly consider the possibility that Devitt's own characterisation of

realism cannot escape Putnam's attack, as the notion of dependence that Devitt espouses could be understood in semantic terms (§ 3). Second, I argue that even if Devitt is right about the effects of the model theoretic argument on metaphysical realism (if this is understood as a thesis about the independence of what exists and that is all), there may be an equally dangerous threat to realism. The final section, therefore, is devoted to explaining this new challenge. I argue that even a metaphysics-first approach to metaphysics is not immune to Putnam's argument, as even though the argument may not be able to motivate a rejection of realism, it could motivate a rejection of our realistic theorising. Therefore, the challenge posed by the model theoretic argument, when understood how I suggest, would be one which could totally undermine the notion of accurate representation in our theorising (§ 4). I conclude by offering some brief remarks what I take to be the metametaphysical and methodological significance of the proceeding discussion (§ 5).[1]

2. Putnam's attack

Putnam's model theoretic argument aims to advance a devastating challenge to metaphysical realism. It aims to show that given plausible considerations from model theory, we are misguided in believing that our language can be hooked up with a mind-independent world in right way; where "right way" is to be understood as "the way the metaphysical realist purports they are hooked up". Putnam thinks this motivates a move away from metaphysical realism towards his preferred 'internal realism'.[2]

Before turning to the argument itself, I need to outline how Putnam characterises the commitments of metaphysical realism. On his view, the realist is not committed to one thesis but three interrelated theses. First, a claim about the world: "THE WORLD" is to be characterised as independent, in the sense of independent of any representation of it. Second, a claim about language: that the reference relation for the realist is one of correspondence; that is, that there is a determinate reference relation which holds between expressions of our language and the parts of the world to which they refer. Third, a uniqueness thesis: that there is a *singular* correspondence between the terms of a theory and the objects and the properties to which they refer. (Putnam 1981, 49)

Expressing the uniqueness thesis requires some subtlety. Putnam claims that there is a singular correspondence. We might think this is too strong a requirement and as such Putnam is overstating the realist commitments.

[2] For some excellent discussion of Putnam's position see Button (2013).

There are some forms of semantic indeterminacy that do not seem, prima facie, to be a problem for metaphysical realism. There is a prevalent phenomenon of indeterminacy in natural language: vagueness. It is not particularly controversial to claim that natural language is vague: reference is not always determinant.[3] Paradigm terms that exhibit this indeterminacy would be, 'tall', 'heap', 'bald', etc. Such terms admit borderline cases; cases where it is not clear whether the predicate applies. (Keefe 2000, 6) Such terms apparently lack clear extensions.[4] Given that there are vague expressions in a language, it seems that we need to make sense of the reference of a term not being wholly determinant. In offering an explanation of this phenomenon, a semantic approach has been historically most popular: theorists have tried to explain this vagueness in terms of features of the language, and have offered a semantics and a logic for the suspect expressions. (Merricks 2001, 146) We might think, therefore, that the existence of vague predicates suggests that a one-to-one correspondence is too committal. Consider your total theory of reality. Perhaps you quantify over colours. There could be two properties, red and red*, which the word "red" in the theory could pick out. If our total theory quantifies over "red", and yet there are two equally good candidate referents for "red", this might seem like enough to suggest that there is not a singular correspondence, but rather that there may be multiple equally good ones.

I think that the realist could make two moves. First, perhaps she could retreat. A one-to-one correspondence is too committal, and as such the realist accepts that in some instances there is not a singular correspondence. The theory can correspond to greater and lesser degrees, and the realist prefers those theories which correspond more. However, there can still be a degree of flexibility, especially in those cases of mun-

---

[3] There is a distinction between vagueness and indeterminacy. As Van Inwagan clearly puts it, "Vagueness is a special case of indeterminacy—semantical indeterminacy. It may be indeterminate whether a sentence is true or false, indeterminate whether a term denotes a certain object, and indeterminate whether a given set is the extension of a certain predicate. I take the word 'vague'— my universe of discourse here comprises only linguistic items—to be entirely appropriate only in application to predicates and certain of their constituents. A predicate is vague if it is indeterminate, or, at any rate, possibly indeterminate, which set is its extension—or if it is possible that, for at least one object, it is indeterminate whether that object belongs to the extension of that predicate." (Van Inwagan 2009, 1)

[4] It is worth noting that the view that extensions are not determinant only fits with some explanations of vagueness. For example, according to Williamson's epistemicism about vagueness (1994), the referents of vague terms are determinant; the source of the vagueness is in our own ignorance.

-dane indeterminacy considered.[5] The second option is stricter. This type of realist may object to the putative problem, and they will say that no metaphysical theory would quantify over colours simpliciter. Perhaps this is because they think that colours are not the types of things that belong in the fundamental description of reality, or perhaps it is because the language we use to talk about colours in metaphysics do not contain the predicates of everyday English but rather precisified technical predicates. Thus, this realist maintains the one-to-one correspondence, by eradicating instances of mundane indeterminacy by some means. Putnam seems to be levelling his argument against realists of this second variety. For our purposes, what is important is the thought that there should be an intended interpretation for our theory (allowing to the refinements above). The realist contention is that the intended interpretation is the one that matches the terms of the theory to the way the world is.

And so, we return to Putnam's characterisation of realism. Once we have these three theses, it should be clear that it is possible that we may be unable to represent THE WORLD at all. To put it in Putnam's terms: truth is, for the metaphysical realist, radically non-epistemic. (Putnam 1977, 485) The theory which is epistemically ideal, in the sense of meeting all our theoretical virtues, might still be false. Of course, it might be the case that ideal theory *is* true; the claim is not that ideal theory must fall short of truth. Rather the claim is that, for the realist, idealness does not constitute or guarantee truth. The realist must hold that there is a gap between ideal theory and true theory. At the heart of Putnam's attack, therefore, is a distinction that realist must be committed to, but, according to Putnam, cannot be maintained:

> Here again, the realist – or, at least, the hard-core metaphysical realist – wishes it to be the case that *truth* and *rational acceptability* should be *independent* notions. He wishes it to be the case that what, e.g., electrons are should be distinct (and possibly different from) from what we believe them to be or even what we would believe them to be given the best experiments and the epistemically best theory. Once again, the realist – the hard-core metaphysical realist – holds that our intentions single out "the" model, and that our beliefs are then either true or false in "the" model *whether we can find out their truth values or not*. (Putnam 1980, 472)

Now, let's turn to the argument. I should note that *the* model theoretic argument doesn't strictly speaking pick any one thing out. Putnam forwards several arguments which are similar in spirit, and, whilst they

---

[5] Perhaps this could be metaphysically spelled out in terms of a commitment to ontic vagueness. This would not be to move away from realism as we can be realist about what is metaphysically indeterminate. For some illuminating discussion please see Barnes and Williams (2011).

differ in precise target and technical machinery, they all aim to forward the same conclusion.[6] Some of his discussion focuses on mathematical language and employs the Löwenheim-Skolem theorem,[7] whereas other discussion employs permutation models and focusses more on ordinary language.[8] Here, I focus on the so-called permutation argument.

The model-theoretic conception of a theory is one whereby there is a language L with a given interpretation function I, which maps the expressions of L onto a world of objects and properties. Putnam places some methodological constraints on what can fix the intended interpretation. First, there are what Putnam terms 'theoretical constraints'. These include the standard axioms of set theory, as well as principles and theories from other branches of science. Second, there are the 'operational constraints'. These are the various empirical observations and measurements that we make in the course of scientific investigation. There is a dual constraint on the assignment of extensions to the subsentential components of the sentences of the language. First, this must be constrained by facts about usage, understood in a suitably naturalistic way. I take it that by "naturalistically acceptable" Putnam intends to acknowledge his Quinean heritage and thus this means taking a somewhat behaviouristic view of the data for semantic theorising. Second, assignments must be constrained by getting the coarse-grained truth conditions of whole sentences correct. The data which must constrain our semantic theorising is the semantic values of sentences, and the success conditions for a theory of reference are to fit this data set. Importantly, if there are multiple theories which both equally fit the data, then it is indeterminate which theory is correct.

The argument then shows that for every theory T1, it is possible to find a permutation function K, such that each item in L is interpreted "in violently different ways, each of them compatible with the requirement that the truth value of each sentence in each possible world be the one specified". (Putnam 1981,33) The mechanism functions as the "crazy assignments of reference to names can be 'cancelled out' by a compensating assignment of extensions to predicates, so that, overall, the truth value of sentences is unaffected." (Williams 2007, 369)

What we find is therefore worrying. There are multiple interpretations, all of which meet the requisite constraints. Although these considerations only apply to the set-theoretic language in question, Putnam argues that an argument such as this be generalised: it can apply to all languages.

_____

[6] This claim is contingent on how finely you individuate the conclusions. When I say similar in spirit, therefore, I mean Putnam uses both arguments to attack metaphysical realism.

[7] See for example, Putnam (1980).

[8] See for example, Putnam (1981) and (1977).

Different interpretations of a sentence can produce the same truth value at every possible world, but assign different extensions to the subsentential expressions. That is: in both the intended and permuted interpretations, 'cat' is true of radically different things, but true in exactly the same circumstances nonetheless. Given how we defined the constraints on our metasemantic theorising, neither interpretation can be said to be any better than any other, and it is therefore indeterminate which is correct.

This is deeply in conflict with the commitments of Putnam's realist. Given the realist's commitment to uniqueness, we would have hoped that there would be a single true interpretation function which takes us from terms of the theory to the world. However, we find that reference is *radically* indeterminate. This is not the innocuous sort of indeterminacy we considered at the start. It is the claim that even with all the constraints on reference assignment in place, and even for precise languages like mathematical language, reference is still indeterminate. These considerations put pressure on another supposed commit of realism: the non-epistemic nature of truth. Putnam's realist needs it to be the case that there is an independence of ideal theory and truth. However, for any consistent theory, that theory has a model, and Putnam's argument shows us that if there is one truthmaking model, then there are infinitely many permuted variants. The realist wants to be in a position to say that of all these models, one of them is getting it right and the others are getting it wrong. But even with all the constraints on reference assignment in place, they are still not able to discriminate. So, if the realist is going to persist in maintaining the distinction between ideal theory and truth, then they need to invoke some kind of magical relation. They cannot say anything principled about why one model is getting it right and the others are getting it wrong. Putnam claims that these considerations are sufficient to undermine the prospects for robust metaphysical realism.

If we were to try to resist such an argument, in which direction would we go? Perhaps a more sophisticated theory of reference could show that we were wrong? Putnam claims that this cannot save our intuitive notion of reference as whatever additional constraint we bring in to save some sensible realist notion of reference, could be likewise subjected to a permutation.[9] That is: the new constraint itself needs interpretation and each model will interpret it in different ways. The point of Putnam's argument is that no first order theory can, by itself, determine its own objects up to the point of isomorphism. Putnam states the point thus: "The problem as to how the whole representation, including the empirical theory of knowledge that is a part of it, can determinately refer is not a problem that can be solved by developing a more and better empirical theory." (Putnam 1980, 477) There are some philosophers who hold that

---

[9] Again, for more in depth discussion of this aspect of Putnam's argument see Button (2013).

the arguments for radical indeterminacy theses must not be in good standing as they are self-defeating. For example, Scott Soames claims that if the conclusions to the arguments were true, we couldn't even state them. (Soames 1998, 213) There are other philosophers who have argued that the arguments are not in good standing as they are too stringent in what they allow to be taken as the constraints for providing a theory of reference. David Lewis, for example, argues that we need not be worried as some interpretations are more eligible than others and so the type of indeterminacy in question does not arise. (Lewis 1984, 227)

There is of course a great deal more to be said about these responses. For my present purposes, however, I want to precede by assuming that Putnam's argument for radical indeterminacy is in good standing; that the metamathematics of Putnam's argument is correct.[10] I am interested in working out the conditional: *if* an argument like Putnam's is right, what can that tell us, if anything, about metaphysical realism? I now turn to assess a response to Putnam according to which even if we allow that Putnam's conclusions about reference and truth are correct, this does not show us anything about the prospects for metaphysical realism.

3. Devitt's response

Michael Devitt asks this question: "What does truth have to do with realism?" (Devitt 1983, 292) An obvious initial answer to Devitt's question: it depends who you ask. First, let's consider Putnam. Putnam tells us that a central commitment of metaphysical realism is that it is possible for ideal theory to be false: there is independence of ideal theory and true theory. He tells us that the notion of truth that we are interested in as metaphysical realists is one which matches up with reality, in an appropriate way. Thus, if we find out that we have a crazy metasemantics, then given that some of the central commitments of realism are semantic ones, we are in trouble.

Devitt, however, tells us we do not need to worry. His answer to the question of what truth has to do with realism is "nothing at all". (Devitt 1983, 292) Devitt captures the notion of realism he is interested in, and according to him the one metaphysicians should be interested in, in terms of two commitments. These commitments run along two dimensions. First: existence. Realism commits us to a view about what exists. If I am a realist about some entity, or class of entities, I am committed to those things existing. The second commitment is trickier: "words that frequently occur in attempts to capture the second are 'independent', 'external' and 'objective'. The entities must be independent of the mental;

---

[10] This view has been defended at some length. For some good discussion see Button (2011).

they must be external to the mind; they must exist objectively in that they exist whatever anyone's opinions." (Devitt 1983, 292)

The doctrine of realism according to Devitt should be construed thus: "Common sense, and scientific, physical entities objectively exist independent of the mental." (Devitt 1983, 292) Construed as such, I can see the temptation to claim that truth does not have anything to do with realism. It makes no mention of semantic notions at all and it does not seem clear, at least prima facie, how this characterisation might implicitly rest on the sort of semantic commitments required to get Putnam's attack going. Devitt claims that the type of indeterminacy Putnam espouses only has metaphysical significance if you endorse a language first approach to metaphysics.

Perhaps we can view Devitt's criticism of Putnam as a plea not to conflate metaphysics and the philosophy of language. It is only by conflating the two that one might be inclined to think that considerations from the philosophy of language could impact upon the realist status of the world. Of course, language is a tool for theorising. We do theorise *using* language, but we must not let this confuse us. The *content* of our theorising has nothing to do with language, truth or any kind of semantic concerns, and so we should not be alarmed by Putnam's putative challenge. Metaphysics, realist metaphysics, is concerned with the world, not language.

Additionally, perhaps we can see some immediate appeal to Devitt's position. There is intuitive appeal; what we can and cannot say, what we can and cannot theorise about, should not effect what actually is the case. That there is a world, and that that world exists mind-independently, should not be effected by any concerns pertaining to model theoretic languages. In misrepresenting the commitments of realism Putnam gets away with far more than he should. Whilst we may still have cause for concern about the significance of Putnam style arguments in the case of metamathematics or metasemantics, this does not force us to any conclusions about the viability of metaphysical realism. Realism, so construed, has minimal commitments (i.e. an existence thesis and an independence thesis), and neither of these commitments are effected by the semantic concerns forwarded by Putnam.

There are two responses to Devitt I want to consider. The first is a direct concern about his explicit separation of realism from any semantic notions. Specifically: do we need any semantic notions to make sense of his independence thesis? The question to ask at this stage is: exactly how should we understand mind-independence? This is of course a well-rehearsed subject matter: there has been much discussion, especially in recent years, about dependence. But in the face of the challenge from Putnam, it is worth thinking about exactly what notion is at stake.

There might be some who chose to reduce their notions of dependence to

semantic notions, and perhaps we can see the motivation for this. Historically, philosophers have been very cautious about using metaphysically primitive notions. You find a lot of engagement with reductive projects which aim to analyse away the metaphysically primitive in favour of some more familiar notions. Rather than dependence being something that exists between things (in the world) perhaps it could better (and more parsimoniously) be characterised as a relationship between propositions? Analysing dependence in terms of, for example, the truth of propositions, we can do away with primitive metaphysical ideology. And perhaps we can see motivation for this: many metaphysicians want to think of some truths holding in virtue of some other truths.[11] But the bearers of truth are propositions, not objects. However, thinking about things in this way, a concern emerges: if we decide to explicate dependence in such terms and place at the centre of our understanding of realism a claim about the relationship between true propositions, then Putnam's argument can gain traction even with Devitt's characterisation of realism.

However, we can circumvent such problems by keeping dependence metaphysical. Exactly what would this consist in? To say that something exists mind independently is to say that the thing does not require the existence of minds for its existence. This requires some clarifications. First, there is a sense in which the laptop in front of me is mind dependent as, if it were not for the existence of minds, if human beings had not designed and built my laptop, it would not exist. This is not a metaphysically interesting sense, and not what we are concerned with when we are concerned with realism. Second, talk of "minds" needs some unpacking. Whose minds exactly? Is it my mind? Your mind? Some kind of collective consciousness? When talking of minds, as it concerns debates about dependence, we are (mostly) concerned with not just our actual mental lives, but rather with the mental lives of any being with a "finite extension of our cognitive powers". (Jenkins 2005, 199)

Given these two clarifications, there are still two ways we could read a claim of mind independence. Jenkins makes a distinction between modal independence and essential independence. First, we could characterise mind independence as a modal thesis. According to this construal, something is mind independent just in case, "there is a possible world where that thing is the case although our mental lives are not such that …". (Jenkins 2005, 200) This ellipsis will be filled in differently depending on the nature of domain under consideration. For example, if we were considering moral properties, we would say that moral properties exist mind independently is there is a possible world at which moral properties are instantiated even though no minds exist at this

---

[11] For example, in "Fundamental and Derivative Truths", Williams uses a "truth-in-virtue" relation to characterise the fundamental-derivative dependence relation (2010).

world. The second option characterises mind independence as an essentialist thesis. According to this second sort: "p's being the case is independent of our mental lives iff it is no part of what it is for p to be the case that our mental lives be a certain way". (Jenkins 2005, 200) Whilst I do not hope to settle which of these characterisations of mind independence is preferable, they both provide Devitt with the requisite resources to disarm Putnam. Neither of these types of independence theses seem to immediately require any semantic notions and as such we might hope that the Devitt response is in good standing.

For the purposes of argument, then, let's suppose that Devitt's characterisation of metaphysical realism is correct and not susceptible to criticism.[12] Is Putnam well and truly defeated? I think there is a larger, perhaps more pressing, worry which faces realist theorising in light of Putnam's attack. By way of a second response, therefore, I pursue a more indirect route in the next section. Perhaps we concede the point to Devitt: we should not take realism to consist in any commitment to a particular view of truth or reference, and therefore Putnam's criticisms are inert. But, against what are his criticisms inert? Devitt's rebuttal of Putnam's attack on realism was supposed to preserve the world. Any considerations from metasemantics should have no impact on whether there is a mind independent existing world. Fine, we will allow this. But now we can perhaps reconstruct a Putnamian attack: we may have got the world, but can we theorise about it?

4. Adjudicating the dispute: realism and representation

Suppose Devitt is right. Suppose that Putnam's arguments cannot gain any traction in the realism debate, as realism is concerned with the nature of reality, not representing reality, and that further these two concerns are discrete. What I want to suggest in this section is that we may be able to mount another challenge in the Putnamian spirit. The challenge I forward here accepts Devitt characterisation of realism, and, as such, holds that Putnam's arguments cannot have *metaphysical* significance. However, I argue that they carry great significance for the prospect of realist *theorising*. It should be clear, therefore, that the picture I propose is not one that endorses any kind of anti-realism; I am not making any claims about what exists or the mind-independent/dependent nature of what exists. Rather my point is that Putnam's arguments can generate trouble for any metaphysician who aims at faithful representation.

---

[12] I could imagine someone who is sympathetic with Putnam's anti-realist project thinking that I am being too concessionary; that I am conceding too much ground to Devitt. If my primary concern was how we should characterise realism then I think such a charge is reasonable. However, my present concern is to show that even when conceding this ground to Devitt, realism still isn't out of the Putnamian woods.

I take it that realist theories aim at providing a representation of reality. The notion of representation carries with it some notion of correctness. We can get it right, and we can get it wrong. So, what does getting it right or wrong consist in; what grounds 'getting it right'? You might think that getting it right merely requires truth. This still does not tell us enough. There are many different conceptions of truth, and not all of them are going to be compatible with realism. A realist theory is one that aims to represent reality *as it is*. The realist has a robust notion of representation. When thinking about realist representation, therefore, we want metaphysical perspicuity; the theory needs to be true for the right reasons. The elements of the theory correspond to the elements of reality that the theory aims to represent.

This presentation of the requirements of successful representation seems compatible with the practice of realistically minded metaphysicians. Let's consider two examples as a means of illustrating the point. Let's suppose we are realist metaphysicians. We want to offer a metaphysical theory about the nature of, say, time. As a crude toy example: we have the A theorists and the B theorists; let's say a presentist (who thinks that only the present exists) and a moving spotlight theorist (who thinks that all times are equally real, and the present is a matter of perspective) are having a discussion. The thing that they are trying to offer a theory of (i.e. the nature of time) is the same, irrespective of the explanation they offer. So, let's say the presentist has a go first. Time is like thus and so. Then the moving spotlight theorist has a go. Time is like this and that. Both theorists, in aiming to explain and account for time, offer a representation of what they think that aspect of reality is like. They think it is this way, I assume, for good reason. They have data and arguments which urge us to adopt their side. But, only one of these 'gets it right'.

This description of theorising seems like a reasonable characterisation of what many realist metaphysicians seem to be doing. What appears to be a key notion for the metaphysical realist, is that we have a secure notion of accurate representation. Even if we do not want to build a view about the intended interpretation of a theory into the definition of realism, the realist metaphysician must be committed to a view about this.[13] If this is not what they are concerned with, it seems challenging to say exactly what is the concern. By way of contrast, consider second toy example: an instrumentalist about science. Their theories do not aim to accurately represent the way the world is; they do not try to explain the mechanisms which govern natural processes. A successful theory need not say anything true or false about the world. Rather, they aim to provide a tool of prediction and as long as they have a theory which is empirically

---

[13] Whether the realist wants to develop their preferred view of the structure of theories in terms of models or sets of sentences, I take it that the point equally well applies: we need to be able to pick out an intended interpretation. Thank you to an anonymous referee for pushing this point.

adequate, this is good enough. This conception of the purpose of theorising seems in stark contrast to the realist.

Theorising about the world (which is what I take it is we are doing when we engage is realist metaphysics) is a practice of theory construction. This theory, as a representation of that world, has semantic properties; properties such as content, reference, truth conditions, truth values, etc. It is these semantic properties which Putnam's argument can make trouble for as the model theoretic argument shows that the extension of the terms of the theory are radically indeterminate.

The model theoretic argument was supposed to damage realism by showing that there is mass indeterminacy in interpretation. Devitt's criticism of Putnam's argument aimed to show that metaphysical realism is safe from any such attack, as realism is not about interpretation, it is about the world. What I have considered in this section is the possibility that realism is not so safe. If realism is understood as a thesis about what exists and the independence of what exists, a criticism can still be mounted. The same old Putnamian criticisms now attach themselves to the theories that the realist wants to give about what the world is like. The realist's theories come out true (as all Putnam's permuted models come out true), but not for the right reasons. If we wished to resist this what might we say? Perhaps we could argue that theorising does not require sematic notions. I do not see how this can get off the ground. Some minimal notion of accuracy in our theorising is needed, and, in as much as this is the case, Putnam can gain traction. I conclude this section therefore by noting that Devitt does not win by shifting the goal posts. Putnam's argument can yet present a challenge, even if not a *metaphysically* anti-realist one. This challenge, I have argued, bankrupts the realist's method for theorising about the world.

5. Some Concluding Remarks

Let's take stock. The primary aim of this paper was to assess whether we could derive any substantial metametaphysical conclusions from Putnam's argument for semantic indeterminacy. That is: if we accept the permutation argument, must we reject metaphysical realism? The answer to this question is no. If realism is understood as a claim about the nature of reality, then the argument is not capable of showing anything of metaphysical interest about reality.[14] However, the discussion raises a different concern. The significance of Putnam's argument for semantic indeterminacy arises at the level of representation. If we aim to accurately represent reality in our metaphysical theories, then the foregoing arguments seem to make this task impracticable.

_____

[14] Except perhaps something about the factuality of meaning.

I claimed that this debate between Putnam and Devitt speaks to wider issues about how to understand the relationship between language and truth, and metaphysical realism. It would require a much longer study to fully taxonomise the contemporary debates pertaining to this issue, and that is sadly beyond the scope of this paper. Whilst I want to be cautious about over generalising the significance of my present discussion, I want to close with a brief remark on what I think is the take home message from the disagreement between Devitt and Putnam.

When thinking about the relation between metametaphysics and the philosophy of language, I urge that we separate two different kinds of issues. That is: the idea of the philosophy of language as having methodological, if not metaphysical, import. There is a question about what it takes for a position to count as realist. If I claim to be a realist metaphysician I take it that there are two methodological presuppositions which are necessary for me to engage in theorising in the first place. First, I must think that I have some kind of access to the world. This could be seen as an epistemological presupposition. Second, I must think that I am capable of representing the world in my theories. This could be viewed as a metasemantic presupposition. Whatever I take the *metaphysical* characterisation of realism to be, it seems plausible that I must hold these two presuppositions as given if my metaphysical theories are to do what any realist would want them to do. I take the upshot of my argument against Devitt to be that there are certain kinds of arguments and problems in the philosophy of language that are capable of calling into question the legitimacy of these presuppositions. In conclusion, therefore, we must be careful and prudent if we want to try and dismiss the significance of the philosophy of language to realism. Merely claiming that realism is about the nature of reality and not language or truth is, in many cases, not sufficient to rebut attacks on realism which come at it from this angle. Whilst the scope of the significance of such attacks might not be what some philosophers have taken it to be historically, there is a substantial sense in which assumptions about the nature of truth and representation ground the enterprise of realist theorising.

REFERENCES

Button, T. 2011. The Metamathematics of Putnam's Model-Theoretic Arguments. *Erkenntnis* 74(3): 321-349.

————. 2013. *The Limits of Realism*. Oxford: Oxford University Press.

Devitt, M. 1983. Realism and the renegade Putnam: A critical study of meaning and the moral sciences. *Noûs* 17 (2): 291-301.

Keefe, R. 2000. *Theories of Vagueness*. Cambridge: Cambridge University Press.

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

Merricks, T. 2001. Varieties of Vagueness. *Philosophy and Phenomenological Research* 63: 145–157.

Putnam, H. 1977. Realism and Reason. *Proceedings and Addresses of the American Philosophical Association* 50.

————. 1980. Models and Reality. *Journal of Symbolic Logic* 45 (3): 464-482.

————. 1981. *Reason, Truth, and History*. Cambridge: Cambridge University Press.

Soames, S. 1998. Skepticism about meaning, indeterminacy, normativity, and the rule-following paradox. *Canadian Journal of Philosophy* 23: 211-50.

Van Inwagen, P. 2009. Indeterminacy and Vagueness: Logic and Metaphysics. *European Journal for Philosophy of Religion* 1 (2): 1-19.

Williams, J. R. G. & Barnes, E. 2011. A Theory of Metaphysical Indeterminacy. In Bennett and Zimmerman (eds.), 103-148.

Williams, J. R. G. 2010. Fundamental and Derivative Truths. *Mind* 119 (473): 103-141.

————. 2007. Eligibility and inscrutability. *Philosophical Review* 116 (3): 361-399.

Williamson, T. 1994. *Vagueness*. London and New York: Routledge.

DEFERENCE AND STEREOTYPES*

ANDREI MOLDOVAN
University of Salamanca, Spain

ABSTRACT

In this paper I discuss Hilary Putnam's view of the conditions that need to be fulfilled for a speaker to successfully defer to a linguistic community for the meaning of a word she uses. In the first part of the paper I defend Putnam's claim that knowledge of what he calls "stereotypes" is a requirement on linguistic competence. In the second part of the paper I look at two consequences that this thesis has. One of them concerns the choice between two competing formulations of consumerist semantics. The other concerns the notion of deference, and in particular the question whether deference can be non-intentional. Although the standard view is that deference is intentional, it has also been argued (Stojanovic et al. 2005) that most common forms of deference are not. I argue that deference is best understood as intentional, given the possibility of failures of deference. Cases in which the requirement that the speaker know the stereotypes associated with a particular word is not fulfilled are examples of unsuccessful attempts to defer.

Keywords: *deference, Putnam, stereotypes, intention, default deference*

1. Introduction

Semantic externalism, broadly conceived, is usually understood as a meta-semantic, or presemantic (Almog 1984, 482), account of natural language. In Robert Stalnaker's (1997, 535) words, semantic externalism is a *foundational*, as opposed to a *descriptive*, theory of meaning. The aim of a descriptive semantic theory is to characterize the meaning of certain expressions. The aim of a foundations theory is to give an account of the facts that need to obtain in order for those expressions to have the meaning that they have. According to Stalnaker, a foundational theory

tells "what it is about the capacities, customs, practices, or mental states of a speaker or community of speakers that makes it the case that an expression has the semantic value that it has" (Stanlaker 1997, 542).

The paradigmatic externalist view concerning proper names, as formulated by Saul Kripke (1980) in his causal picture of reference, explains why proper names have the referent they have by appeal to three kinds of facts: 1) an initial baptism, in which the name is associated to the individual it is meant to name; 2) a chain of uses that preserves the reference over time; 3) and finally, certain facts about the way the speakers use the name (in particular, the intention with which they utter the word).

These meta-semantic considerations have been extended to natural kind terms such as 'water' and physical magnitude terms such as 'temperature' in Kripke (1980), Putnam (1975), Devitt (1981) and others. The idea is that the reference of these terms is fixed by mechanisms that have a social and historical dimension, in ways that are analogous in important respects with the case of proper names. Moreover, as Joseph Almog argued, the externalist considerations are even wider in scope. Externalism about proper names sets the basis for a socio-historical theory of linguistic *meaning*, and not only of reference:

> The historical chain preserves the linguistic meaning of any expression. In the case of names, all there is to this meaning is to stand for the given referent. *Ergo*, the chain preserves the fact that the name stands for that referent. (Almog 1984, 482)

If Almog is right, an externalist meta-semantic story can be told about linguistic meaning in general, both in case of typical referential expressions, as well as for other expressions.

Keith Donnellan (1993, 155) argues that in characterizing externalism it is important to distinguish two theses that are both proposed Putnam's "The Meaning of 'Meaning'" (1975). One of them might be called "physical externalism" (as in Wikforss 2008), and it concerns the semantics of natural kind terms. This is the thesis that the extension of such terms is set by the underlying nature of a set of paradigmatic instances of that natural kind. The rule for the use of the term sets the paradigmatic instances of that kind (e.g., water), and sets as extension anything that has the same internal structure as the paradigm instances, but does not identify the properties that form the underlying nature of these paradigmatic instances (e.g., having the chemical structure H_2O). The latter might be, and usually is, unknown to the speaker, and even to the whole linguistic community.

A second thesis that is to be found in Putnam (1975, 228) might be called "social externalism". Putnam formulates it as the Hypothesis of the Universal Division of Linguistic Labour:

> Every linguistic community… possesses at least some terms whose associated 'criteria' are known only to a subset of the speakers who acquire the terms and whose use by the other speakers depends upon a structured cooperation between them and the speakers in the relevant subsets. (1975, 228)

Not all members of a linguistic community possess all the criteria for the correct use of all the words. There are those members whom Putnam calls "experts" (1975, 228), who do have the capacity to discriminate instances of, say, water, from those that are not. Putnam writes: "the way of recognizing possessed by these 'expert' speakers is also, through them, possessed by the collective linguistic body, even though it is not possessed by each individual member of the body" (1975, 228). If 'arthritis' has a meaning in the linguistic community then, in accordance to Putnam's Hypothesis, some users must have good recognitional capacities of the instances of that kind that are presented to them in normal conditions. This is the case not only for natural kind terms, but also for words that refer, for instance, to artifacts, such as 'carburetor', Donnellan (1993, 162) notes. Although I might lack the capacity to discriminate carburetors from similar devices, the mechanic does possess this capacity, and, in virtue of this, the word has a meaning in the linguistic community. The individuation of meaning includes factors that are external to the speaker, and concern the existence of a linguistic practice of using that expression in the speaker's linguistic community. As Donnellan (1993, 162) explains, this idea is similar to Burge's (1979) "anti-individualism" and his claim that it is the linguistic community that provides the standard by which the speaker's *mastery* of a word is to be judged. According to Burge, the conventional linguistic meaning "may vary with the individual's environment, even as the individual's activities, individualistically and nonintentionally specified, are held constant" (Burge 1986, 273). The relevant environment includes the linguistic community to which the speaker belongs.[1]

The distinction between the physical and the social varieties of externalism has parallels in what concerns linguistic competence. Concerning the former, competence with proper names and natural kind terms does not require access to facts about the physical constitution of the individual or kind in question. The speaker needs not possess a uniquely identifying description of the essential properties of the referents. Similarly, social externalism supports the idea that an individual can achieve competence in using an expression even if she

---

[1] Burge (1979, 1986) initially defends anti-individualism with respect to the content of cognitive states, and later, in Burge (1986, 1989), extends the view to semantic properties. He proposes an externalist view of the linguistic meaning of a large class of expressions applicable to empirical objects, stuffs, properties, and events. Burge (1989: 283) makes reference to expressions such as 'tiger', 'water', 'mud', 'stone', 'tree', 'bread', 'knife', 'chair', 'edge', 'shadow', 'baby', 'walk', 'fight'.

does not have access to a correct description of the relevant social facts on which the meaning of the expression depends. The speaker needs not know when or where the initial baptism took place, or how exactly the word was introduced in the language, and she needs not know much about the chain that connects her causally to that initial moment. Social externalism also entails that the competent user of an expression needs not be able to provide a uniquely identifying description of the linguistic practice, i.e., an account of the semantic rules that members of the relevant social community follow. As Burge puts it, speakers need not have the ability to offer a correct "explication" of the meaning of the words. A speaker's explication of meaning, or "what the individual would give, under some reflection, as his understanding of the word" (1989, 282), may be far from correct. Possessing the information needed for perfect mastery of the word is not required for meaningful use of a word (although a minimum of knowledge is, arguably, required, as we will see later). In other words, in order for a speaker's utterance of a word-form[2] to have the linguistic meaning that it has in the relevant linguistic community (that is, to be a *word* with that form) the speaker needs not have access to a complete characterization of the correct use of the expression.

2. The notion of deference

The thesis that semantic competence with proper names and natural kind terms does not require uniquely identifying descriptive knowledge of either the extension or the correct use does not yet tell us what competence *does* require. According to the standard externalist view, at least one condition must be fulfilled: the speaker must *defer* to the linguistic community for the meaning of the word. In relation to the notion of deference, it is important to distinguish, following Donnellan (1993, 163), *semantic* deference, which consists in relying on the linguistic community (in particular, on the "experts", in Putnam's sense of the word) for the exact meaning of a particular word from *epistemic* deference, which consists in deferring to a particular specialist, or to a community of experts (in the more usual sense of the term "expert") for the justification of a particular claim. Given that I have an incomplete mastery of the word 'arthritis', I defer semantically to the community of medical experts for the exact sense and correct use of the term. On the other hand, in uttering the sentence 'I do not have arthritis' I may defer, epistemically this time, to the physician that gave me this diagnosis, as I lack direct evidence in favour of this claim. If the physician is wrong about my condition my utterance is false, but still meaningful, in virtue of

---

[2] By 'word-form' I mean a phonetic entity or mark on a paper that is a possible candidate to be an expression of a language, and which is non-syntactically and non-semantically individuated.

having semantically deferred for the meaning of the word.

In order to use the word with its customary meaning, the speaker must rely on the community of competent speakers that engage in the linguistic practice on which the meaning of the expression depends. In order to do so, he must form a particular *intention*, which Kripke characterizes as follows: "When the name is 'passed from link to link', the receiver of the name must, I think, intend when he learns it to use it with the same reference as the man from whom he heard it." (Kripke 1980, 96)[3] Given that, as Burge (1979) puts it, it is the linguistic community that possesses the standards by which the speaker's *mastery* of a word is to be judge, semantic deference also requires the *disposition* to stand corrected in face of criticism coming from those he recognizes as having a better mastery of the word.

But is it sufficient to *intend* to use the word correctly, and to have the disposition to stand corrected? Semantic deference, at least in some formulations of it, requires more than the mere intention to use the word correctly. Kripke's formulation of the condition in the passage quoted presupposes that the speaker be aware of the semantic function that the expression serves in the language, i.e. that he be able to correctly identify the expression as a proper name, or at least as a singular referential expression. Otherwise he cannot "intend when he learns it to use it *with the same reference* as the man from whom he heard it." (Kripke 1980, 96; emphasis added) Scott Soames mentions such a requirement on competence with proper names explicitly:

> In order to be a user of a name n of an object o, two things are required. (i) One must have acquired a referential intention that determines o as the referent of n. […] (ii) One must realize that to assertively utter *n is F* is to say of the referent, o, of n that it 'is F'. (2002, 65)

For linguistic competence with names one must, apart from having the right deferential intention when using n, realize that n is a singular referential expressions.[4]

According to Putnam (1975), these conditions are sufficient for competence with proper names, but not for natural kind terms. He writes:

---

[3] As Kaplan (1989: 602), among others, points out, and *pace* Soames's formulation in the passage quoted below, it is not the speaker from whom one learned the word that one must defer to, but the linguistic community as a whole. That speaker from whom one learned the word may have used it in non-standard ways, or even with non-standard intentions.

[4] In a plausible interpretation of Soames' second condition, this does not require that the speaker have *the concept* of a referential expression, or that of a proper names, but rather that he must treat the expression *n* as a referential one in his linguistic and non-linguistic behaviour.

we don't assign the standard extension to the tokens of a word W uttered by Jones no matter how Jones uses W… One can use the proper name 'Sanders' correctly without knowing anything about the referent except that he is called 'Sanders' – and even that may not be correct... But one cannot use the word tiger correctly, save per accidens, without knowing a good deal about tigers, or at least about a certain conception of tigers. (1975, 246-7)[5]

Putnam (1975) proposes that the speaker must also associate with the natural kind term a minimum set of descriptive information, which he calls "stereotypes".[6] Stereotypes, he writes, are "conventional ideas, which may be inaccurate" (1975, 249), and which have come to be associated with the kind term due to a variety of historical contingencies. These are claims commonly held to be true about the instances that belong to a specific natural kind, but which are not analytically true, and might not even be true at all. Although they include criteria for recognizing if a thing belongs to the kind, they need not include the best criteria that the community has, or the criteria that the experts use to recognize typical members of the class. In our culture, Putnam (1975, 230) writes, the stereotype for 'elm' might just be that it is a common deciduous tree, and that for 'molybdenum' might be that it is a metal.

In the above passage Putnam mentions knowledge of stereotypes as a requirement on correct use of the word. Later on he writes that this is also a requirement on *word acquisition*:

We shall speak of someone as having acquired the word 'tiger' if he is able to use it in such a way that (1) his use passes muster (i.e. people don't say of him such things as 'he doesn't know what a tiger is', 'he doesn't know the meaning of the word "tiger"', etc.); and (2) his total way of being situated in the world and in his linguistic community is such that the socially determined extension of the word 'tiger' in his idiolect is the set of tigers. (1975, 247)

In order for one's use of 'tiger' to mean tiger, i.e. for one to acquire the word 'tiger', one must (2) rely on the linguistic community for the meaning of the word, but also (1) count as "knowing the meaning" of the word. For (1) to be fulfilled, one must conform at least minimally to the correct practice of using the word. That requires a minimal level of knowledge about how people use the word 'tiger', which, in turn, requires knowledge of the associated stereotypes. If one lacks this know-

---

[5] An anonymous reviewer pointed out that it is strange that Putnam does not use single quotes when he refers to the word 'tiger', unlike in other passages in the same paper. This is also inconsistent with his use of quotes in the same passage in the case of the proper name 'Sanders'. Most probably this is a typo.

[6] Devitt (1981: 197-199) and Platts (1997: 288), among others, disagree with Putnam, as they see no reason why the conditions should be different for natural kind terms from those for proper names.

-ledge, then the intention does not guarantee success in word acquisition, that is, in using the word with the meaning it has in the community. Although Putnam does not put it in these terms, one could express this condition by saying that the *intention to defer* is a necessary but not a sufficient condition for successful deference. Only if the speaker also knows the stereotypes associated with the natural kind term she *successfully defers* to the linguistic community for the meaning of the word.[7]

According to Putnam's argument, as I understand it, knowledge of stereotypes is a requirement for minimally correct use of the word, and the latter is a requirement on word acquisition, i.e. successful deference. Let us focus on the last step of Putnam's argument. Why is correct use a requirement on successful deference? I find no answer to this question in Putnam's discussion of stereotypes, but a plausible one might be that a minimal level of mastery is itself needed for the possibility of rational interpretation of one's utterances, and the latter is required for successful attribution of meaning to the word-forms one utters. So, the justification of the condition concerning knowledge of stereotypes rests on the consideration of certain requirements on the general form of a theory of *interpretation* of linguistic behaviour. As Mark Platts (1997) writes:

> The aim of a theory of interpretation is to make sense of a person's linguistic behaviour as part of making sense of him. The constraint upon any such theory is that it makes sense to say of the person that he was then and there saying what our theory represents him as saying. (1997, 288)

Applied to semantic interpretation, this is the following principle: if we are not disposed to attribute to the speaker the speech act of having *literally said* what the word-forms literally mean within the linguistic community to which the speaker is intending to defer then deference is unsuccessful, and the word-forms fail to obtain that literal meaning. Such a case obtains when the speaker's ignorance of the correct rules of application of the terms is manifest. So, minimally correct use is a requirement on successful deference because it is a requirement on semantic interpretation, i.e. ascription of literal meaning to the speaker's utterance. If the speaker has no idea how to apply the word 'tiger' we will refrain from interpreting her as making assertions about tigers, and so as using 'tiger' with its customary meaning.

Let us consider a concrete example to illustrate this claim. Suppose J is a student of English at beginner level, and completely lacks knowledge of

---

[7] Kim Sterelny (1996: 104-106) maintains a similar view about natural kind terms: apart from the deferential intention, a speaker must have a cluster of beliefs about the causal powers of the kind, and "the term will be grounded only if most of those beliefs, or perhaps certain central ones, are true". The speaker must also possess recognitional capacitates: she must be able to "discriminate, reasonably reliable, members of the kind".

the associated stereotypes for any of the expression-forms he is pronouncing and spelling. He is merely repeating words he has just heard and which he assumes to be meaningful in English when uttering 'I have seen tigers in the street today'. It would be incorrect, or at least misleading, to say that J is literally talking about tigers, even if he does have the intention to use the words with the standard meaning. As a result, semantic interpretation cannot proceed, and the sounds he made cannot be counted as meaningful tokens of the language he intends to speak. A child in the early stages of language acquisition, who imitates the sounds she hears, provides a similar example.

Another case relevant to our purposes is one that Burge (1979) discusses, involving a grossly nonstandard use of 'orangutan'.[8] He writes (1979, 191-120): "If a generally competent and reasonable speaker thinks that 'orangutan' applies to a fruit drink, we would be reluctant, and it would unquestionably be misleading, to take his words as revealing that he thinks he has been drinking orangutans for breakfast for the last few weeks." It would also be misleading to take him as having *asserted* (or said) that about orangutans. This case is different from the previous one in that now the speaker does have an identifiable communicative intention (e.g., he has orange juice in mind all the time when he uses the word-form 'orangutan'), and so we can retrieve a *speaker* meaning. We can make sense of his linguistic behaviour, and attribute to him the *intention to say* that he has been drinking orange juice for breakfast, but not the speech act of having literally said so. It was not orangutans that she was talking about. Whenever the use is systematically and grossly non-standard we count the speaker as having failed to say anything that might be judged literally true or false.[9]

Cases such as the above show that the intention to participate in a practice does not suffice; when the speaker's use of the word radically departs from standard use the intention to defer is frustrated. To use a common analogy, one can play a game even if one does not know all the rules and all the definitions of technical notions (i.e. one does not have perfect mastery of the game); but one cannot play a game if one believes the rules of the game are very different from what they actually are. And the same applies to a 'language game'. So, the speaker who systematically uses 'orangutan' in non-standard ways cannot count as playing the same language game the members of the community to which she intends to defer play. The deferential intention to play the language games of words such as 'tiger' and 'orangutan' is not sufficient.

---

[8] Burge (1979) uses this example to make a different point, one concerning attribution of mental states on the basis of the literal meaning of that the words uttered, and not concerning attribution of literal meaning to the words.

[9] If his use of 'orangutan' to mean orange juice becomes standard over time, then we might say that 'orangutan' obtains a new meaning within that new linguistic community, in the same way in which 'Madagascar' does, in Evans' famous example.

The language game analogy suggests that in order to use a word with its customary meaning some minimal level of familiarity with the basic rules of the language game is needed. But what is this level, and what are these basic rules? The English student example suggests that one cannot defer if one lacks knowledge even of the semantic properties of the word (e.g., being a common noun, in this case).[10] The 'orangutan' example suggests that one cannot defer if one *has false believes* about what the stereotypes are. But what if one identifies the semantic category of the expression, does not have false belief about stereotypes, but still does not have knowledge of the associated stereotypes? Would that allow for successful deference? Suppose I am a fluent speaker of English who has never heard the word 'carburetor' before. Does my utterance of it have the customary meaning when I ask what a carburetor is? Intuitively it might seems so, and it that case, whatever successful deference requires, it does not require any knowledge of stereotypes.[11]

However, there are good reasons to resist this conclusion. There is, indeed, a sense in which I *use* the word 'carburetor' even if I lack any knowledge about what the word stands for, and I am completely unfamiliar with how others use it. In the same sense, one might count as *playing* badminton when one is merely practicing the movement of the hand in hitting the shuttlecock as part of the learning process, even when one knows nothing else about badminton. But those are very loose senses of "using" the word and "playing" the game, respectively. The sentences in which I am disposed to use 'carburetor' are mainly questions about how the word should be used or what it means. I am aware that if venture to use it in other contexts I run the risk of using it in radically nonstandard ways, as our speaker in the 'orangutan' case does. Deference is unsuccessful in that case because use is radically nonstandard. Deference is unsuccessful here because use reduces to very little.[12]

To sum up, the speaker's expression-forms acquire meaning from a pre-existing linguistic practice only if the speaker *participates* in that linguistic practice, and so, only if she uses the expression-form in mini-

---

[10] We have seen that Soames (2002) requires such knowledge for successful deference with proper names.

[11] According to such a position a better analogy than that of the language as a game would be that of language as a tool: a word has a meaning when I utter it in the same way in which, say, the blacksmith's tools I am contemplating in his workshop have specific purposes and uses for which they were designed, even if I am not knowledgeable of what they are. If this analogy were correct, the student of English in our first example, who lacks any syntactic or semantic knowledge of the use of the words she is pronouncing, would still defer successfully. If she does not it means the language game analogy is a better one.

[12] Consequently, one might view this case as one in which the word is not used, but mentioned, even if no quotation marks are used. In the same sense, in 'My name is Andrei', 'Andrei' is mentioned and not used.

-mal accordance with the rules of that linguistic practice. A minimum is required in terms of knowledge of the semantic rules that govern the use of that expression, and, in turn, this requires acquaintance with the associated stereotypes. As Putnam comments,

> This idea should not seem too surprising. After all, we do not permit people to drive on the highways without first passing some tests to determine that they have a minimum level of competence… The linguistic community too has its minimum standards, with respect both to syntax and to 'semantics'. (1975, 248-249)

3. Discussion: what is consumerist semantics exactly?

In what follows I explore a couple of consequences that Putnam's requirement on deference has with respect to closely related issues within the externalist picture of meaning. A direct consequence concerns the general formulation of the externalist view of meaning and meaning acquisition, sometimes called "consumerist semantics". Almog (1984), in the passage quoted at the beginning, notes that Kripke's account of initiation into the use of a proper name is an instance of a more general principle that is not concerned only with the preservation of reference, but also with the preservation of meaning in general. Kaplan (1989), who calls this view "Consumerism", offers the following formulation of it:

> we are, for the most part, language consumers. Words come to us prepackaged with a semantic value. If we are to use those words, the words we have received, the words of our linguistic community, then we must defer to their meaning. Otherwise we play the role of language creators. (Kaplan 1989, 602)

It is useful to distinguish, following Kaplan (1989), between a *producer/ creator* and a *consumer*. Acquiring competence as a consumer is a very different business from acquiring competence as a producer. A consumer of a pre-existing meaning of a word in a language is someone who successfully defers for the meaning of that word to the respective linguistic community. In contrast to a consumer, a producer is someone who does not rely on a pre-existing linguistic practice, but introduces a new word in the language.

Gareth Evans (1982) also uses the consumer-producer dichotomy.[13] The fact that "individual speakers exploit general practices" (1982, 387), he writes, is true of many other semantic properties besides that of referring

---

[13] Evans (1982) gives a slightly different sense to 'producer': for him, in the case of proper names a producer is not only the person who introduces the name in the language, but also that user of the name who can identify the referent of the name.

and transmitting the reference of a referential term. He formulates the following "general principle" for being a consumer:

> if a speaker uses a word with the manifest intention to participate in such-and-such a practice, in which the word is used with such and such semantic properties, then the word, as used by him, will possess just those semantic properties. This principle has as much application to the use by speakers of words like 'agronomist', 'monetarism' and the like as to their use of proper names. (Evans 1982, 387)

Notice that, while Kaplan states a necessary condition, Evans formulates his "general principle" as a *sufficient* condition for one's use of an expression-form to have the linguistic meaning it has in a certain linguistic practice. So do other authors, such as, for instance, Adèle Mercier, who writes:

> Consumerism is the view that, so long as language users intend to defer to the linguistic community in matters of usage, their idiolect and the meanings of their words are individuated by reference to that community. (Mercier 1994, 500)

According to both Evans and Mercier, the *intention* (or the "manifest intention", as Evans puts it) to defer to a particular linguistic community is sufficient for *successful* deference. But if Putnam is right about natural kind terms, then there are reasons to doubt of Evans' and Mercier's formulation of Consumerism. Having the manifest intention to participate in the relevant linguistic practice of using an expression with a certain meaning is arguably not sufficient for one's utterance of that expression to have that meaning. A condition concerning knowledge of stereotypes must also be fulfilled.

4. Discussion: non-intentional deference?

Is having an intention to defer a *necessary* condition for successful deference? Although the authors discussed so far seem to think so, others have argued that it is not. In their paper "Deferential Utterances" (2005) Stojanovic et al. introduce a series of useful distinctions between different kinds of semantic deference followed by a number of interesting comments. The authors make an attempt to mitigate between Andrew Woodfield's claim that deferring is an "intentional act, done by a person for a reason" (Woodfield 2000, 449-450) and François Recanati's (2000) claim that we implicitly defer for most terms that we use in utterances, without this being always an intentional act. Recanati postulates the existence of a "deferential operator", which is an unarticulated constituent that affects the contribution to truth-conditions of any word that we use and for which we are not experts, that is, lack perfect mastery. According to Recanati (2000, 282), this is usually deference that we are not aware of and which is not intentional. In order to do justice to the two

perspectives, Stojanovic et al. (2005) introduce a distinction between "default deference" and "deliberate deference". The former is the kind of deference that Recanati (2000) seems to have had in mind (leaving aside his claim about the existence of an unarticulated constituent, to which the authors do not subscribe), while the latter is the kind of deference that Woodfield (2000) had in mind. They write that "default deference usually goes unnoticed by speaker and hearer", while "[a] speaker who defers deliberately must intend to do so, and her intention must be recognized by her interlocutors." (Stojanovic et al. 2005, 4) Deliberate deference is characterized by the intention to use an expression in the way in which it is used in some dialect, sociolect or idiolect. A simple example of deliberate deference into a common language that they discuss is the following sentence: "Barthes described the book as "un choc historique" and "un repère nouveau et un départ pour l'écriture"." (Stojanovic et al. 2005, 20) The authors comment that the language-shift into French is deliberate, a fact that is exhibited by the contextual features employed to mark the shift, in this case, quotation (but also special intonation or metalinguistic comments in other cases). The authors show that deliberate deference might be to a sociolect (e.g., the way a certain word is used in a particular community) or idiolect (e.g., the peculiar way someone uses a certain word).

In what follows I leave aside what the authors call 'deliberate deference' and focus on what they call 'default deference'. In contrast to deliberate deference, default deference is ubiquitous, as it "is involved in every communicative act. When interpreting and evaluating an utterance, we must take into account a language parameter." (Stojanovic et al. 2005, 6) Usually, default deference is deference to one's own linguistic community, while deliberate deference usually involves a language-shift. But, as in the case of deliberate deference, not all default deference is deference to a language: it might be deference to a sociolect (e.g., the technical definition of 'walk' in the official regulation of race walking), or even someone's idiolect. What distinguishes deliberate from default deference is that in the case of the latter no special recourse is made to contextual features in order to make salient the linguistic parameter. In contrast, with deliberate deference the speaker "exploits certain contextual features in order to make salient the linguistic parameter" for the interpretation of her utterance, and "wants her exploitation of contextual resources to be recognized as part of her communicative intentions by the audience" (Stojanovic et al. 2005, 20).

The authors also claim that default deference usually involves no intention to defer, as opposed to deliberate deference, which does: "a speaker who is deferring deliberately must be aware of what she is doing" (Stojanovic et al. 2005, 6). Default deference usually passes unnoticed by both speaker and hearer. However, they add, the existence or inexistence of a conscious intention to defer is not criterial in making the distinction

between the two kinds of deference:

> this does not make the default/deliberate distinction collapse into the self-conscious/unconscious distinction, for in the case of default deference, too, the speaker may be perfectly aware of the fact that she is deferring by default. (2005, 6-7)

Instead, it is the appeal to contextual factors in conveying what language, sociolect or idiolect the speaker is deferring to that is criterial in making the distinction. Such appeals characterize deliberate deference and lack in cases of default deference.

Let us focus on the authors' remark that "[a] speaker who defers by default most often does not have the intention to defer." (Stojanovic et al. 2005, 4) Is it possible to defer semantically without a specific intention to do so? The claim that it is contradicts the standard view on what deference requires. Kripke, we have seen, writes that "[w]hen the name is 'passed from link to link', the receiver of the name must, I think, intend when he learns it to use it with the same reference as the man from whom he heard it." (Kripke 1980, 96) Others, such as Evans (1982) and Mercier (1995) take such an intention to be not only necessary, but also sufficient.

Contrary to Stojanovic et al. (2005), I think that deference is always intentional. The argument that I present in what follows is suggested by the above discussion of Putnam's approach. If successful deference requires knowledge of stereotypes, then we need to distinguish between *successful* and *failed* deference. Cases such as the ones discussed above in the section on stereotypes help to make this point. In Burge's (1979) example the speaker uses 'orangutan' to speaker-mean orange, but she uses it in ways that are sufficiently nonstandard so as to provide us with a clear example of failed deference. The reason why she uses it non-standardly is that she is confused about the meaning of the word, and associates the wrong stereotypes to the word. However, intuitively we would say that the speaker *intends* to use 'orangutan' as it is used in English. The problem is that he lacks the necessary knowledge concerning how the word is to be used in English. As a result, we cannot grant that the word in his mouth means orangutan. In order to characterize what is going on in such a case we need to talk about *intentions* to defer that do not achieve their aim. When deference is not successful we are left with is a failed attempt to defer. In general, in all cases in which the goal of an action is not achieved there is a need to distinguish between attempt and achievement.

In Burge's example, failure is due to the fact that the speaker associates the wrong stereotypes to the word. In other cases, failure of deference might not involve stereotypes at all. Suppose the word-form the speaker uses does not exist at all in the source language. Imagine a non-native speaker of a language who is led to believe that 'clorange' is the generic name of soda drink that tastes like orange but does not contain orange

juice. The speaker believes the word exists in English and uses it with the meaning she believes it has. But deference is, of course, unsuccessful. There is no question here of whether the speaker associates the right stereotypes with the word, given that 'clorange' is not a word, and so *a fortiori* it has no associated stereotypes. Even if we can retrieve a speaker meaning once we realize the communicative intention with which the speaker uses the word, this is a case in which the speaker forms an intention to defer but deference is unsuccessful.

A third kind of cases of failed deference is identified in Cappelen (2013a, 59; 2013b). The author proposes a strategy for identifying nonsensical uses of expressions generated by a failure of deference. He applies this strategy of diagnosis in order to argue that the use of the notion of 'intuition' in contemporary philosophy is confused and nonsensical. Cappelen's proposal is that there is evidence for a potentially meaningless use of an expression E if that use satisfies the following three conditions:

> (i) The speaker is a member of a number of distinct subcommunities in which E is used in significantly different ways; (ii) The speaker is unaware of (i); and (iii) The speaker defers to the use it has in 'the community' without any particular subcommunity in mind and with (broadly speaking) causal connection to a multiplicity of communities. (2013a, 39)

Independently of whether Cappelen's case for the meaninglessness of the philosophical use of 'intuition' is successful or not, Cappelen does identify a particular strategy for generating failures of deference. One example less controversial than that of a philosopher's use of 'intuition' could be the use of 'liberal' in some contexts of political debate. Consider a speaker who utters: 'When it comes to same-sex marriage I am a liberal.' Now suppose the speaker is insufficiently familiarized with the differences between the political terminologies on the two sides of the Atlantic, and, at the same time, has causal connections to both linguistic communities. She fails to realize that those who identify as 'liberals' in Europe tend to have more conservative views on the issue, while those who identify as 'liberals' in the U.S. tend to be more open. Even if we could retrieve a speaker meaning, it is doubtful that the sentence uttered has identifiable truth-conditions. The case is, rather, one in which the speaker fails to defer properly for the meaning of 'liberal'. The problem, as in the previous case, is not with the condition on stereotypes, but with the condition on the deferential intention, which is not correctly formed. However, intuitively we would say that the speaker does have an intention to defer, even if it is, in some sense, incorrect. This case is best described by saying that the speaker has the intention to mean by 'liberal' whatever experts in political theory 'in the community' mean by the word. The intention is incorrect because it falsely presupposes that there is a *unique* community and a unique use of the word in political contexts, when in fact, there are various.

The above cases show that we sometimes need to postulate an intention to defer when deference is unsuccessful. For parity of reasons, it is natural to think of deference as intentional also in those cases in which it is successful. This is exactly how Kripke (1980), Evans (1982), Mercier (1994), Soames (2002), among the authors quoted above, conceive of it. Of course, this is not to say that the speaker always forms a *conscious* intention to defer to a linguistic community. The actual process might often be readily characterized as unconscious, habitual, and automatic. It usually does not have the phenomenological features of conscious control. Nevertheless, it is still rational, and apt to receive a personal-level explanation in terms of attribution of propositional attitudes such as intentions. David Lewis (1975) makes this point with respect to action in general:

> An action may be rational, and may be explained by the agent's beliefs and desires, even though that action was done by habit, and the agent gave no thought to the beliefs or desires which were his reason for acting. A habit may be under the agent's rational control… (1975, 25-26)

To sum up, we have seen several cases that we can best make sense of by attributing to the speaker an unsuccessful intention to defer.[14] Endowing an expression-form with meaning through deference is not a feature of an utterance of that expression-form that is obtained *by default* (in contrast to, for instance, causing a movement of particles: one cannot utter a word without implicitly causing the particles in the air or in some other environment to move). Instead, it requires that certain conditions be satisfied. It often requires choosing the language, sociolect or idiolect to which one defers, and any choice is intentional. For similar reasons, the correct interpretation of an utterance of an ambiguous word depends on what intention it is reasonable to attribute to the speaker in using the word-form in that context. And the same applies to proper names: here the speaker is not only required to defer to the source language, but also to choose a particular naming-practice (i.e., the use of the name to refer to a particular individual) out of the various ones that the name-form has.

---

[14] An anonymous reviewer insists that the claim that deference can be both intentional and unconscious is not a *prima facie* intuitive one, so it requires some elucidation. However, I do not find the claim unintuitive at all. There are many examples of actions that are done without the full conscious of the presence of a clear formulated intention, but are best interpreted as performed with an intention. Walking home on a familiar route or grabbing your keys when leaving the house are such examples.

REFERENCES

Almog J., J. Perry and H. Wettstein, eds. 1989. *Themes From Kaplan.* Oxford: Oxford University Press.

Almog J. 1984. Semantical Anthropology. *Midwest Studies in Philosophy* 9 (1): 478-489.

Burge, T. 2007. *Foundations of Mind: Philosophical Essays Vol. 2.* Oxford: Oxford University Press.

———. 1989. Wherein is language social? In *Reflections on Chomsky*, ed. A. George, 175-191. Oxford: Blackwell. Repr. in: Burge 2007, 275-290.

———. 1986. Intellectual Norms and Foundations of Mind, *The Journal of Philosophy*, 83 (2): 697-720. Repr. in: Burge 2007, 254-275.

———. 1979. Individualism and the mental. *Midwest Studies in Philosophy* 4 (1): 73-122. Repr. in: Burge 2007, 100-150.

Cappelen, H. 2013a. Nonsense and illusions of thought. *Philosophical Perspectives*, 27 (1): 22-50.

————. 2013b. *Philosophy Without Intuitions*. Oxford: Oxford University Press.

Devitt, M. 1981. *Designation*. New York: Columbia University Press.

Donnellan, K. S. 1993. There is a word for that kind of thing: An investigation of two thought experiments. *Philosophical Perspectives* 7: 155-171.

Evans, G. 1982. *The Varieties of Reference*. Oxford: Oxford University Press.

Hale, B. and C. Wright, eds. 1997. *A Companion to the Philosophy of Language*. Oxford: Blackwell.

Gunderson, K., ed. 1975. *Minnesota Studies in the Philosophy of Science*. Minneapolis, MN: University of Minnesota Press.

Kaplan, D. 1989. Afterthoughts. In Almog, Perry and Wettstein (eds.), 565-614.

Kripke S. 1980. *Naming and Necessity*. Cambridge MA: Harvard University Press.

Lewis D. 1975. Languages and Language. In Gunderson (ed.), 3-35.

Pessin, A. and S. Goldberg, eds. 1996. *The Twin Earth Chronicles: Twenty Years of Reflection on Hilary Putnam's "the Meaning of 'Meaning'"*, eds. M. E. Sharpe.

Platts, M. 1997. *Ways of Meaning: An Introduction to a Philosophy of Language*. Cambridge MA: MIT Press.

Putnam, H. 1970. Is semantics possible? *Metaphilosophy* 1 (3): 187–201. Repr. in: Putnam, H. 1975. *Mind, Language and Reality: Philosophical Papers Vol.2*, 139-152. Cambridge MA: Cambridge University Press.

————. 1975. The Meaning of 'Meaning'. *Minnesota Studies in the Philosophy of Science* 7: 215–271. Repr. in: Putnam, H. 1975. *Mind, Language and Reality: Philosophical Papers Vol. 2,* 215-271. Cambridge MA: Cambridge University Press.

Stalnaker, R. 1997. "Reference and Necessity". In Hale and Wright, 534-554.

Sterelny, K. 1983. "Natural Kinds Terms". *Pacific Philosophical Quarterly* 64 (2): 100-125. Repr. in: A. Pessin and S. Goldberg (eds.), 1996. M. E. Sharpe. 98-114.

Stojanovic, I., De Brabanter, P., Villanueva Fernández, N., Nicolas, D. 2005. *Deferential Utterances*. <https://jeannicod.ccsd.cnrs.fr/ijn_00000575>.

Soames, S. 2002. *Beyond Rigidity: The Unfinished Semantic Agenda of 'Naming and Necessity'*. Oxford: Oxford University Press.

Wikforss, A. 2008. Semantic Externalism and Psychological Externalism. *Philosophy Compass*, 3(1): 158-181.

BRAINS IN VATS AND SEMANTIC EXTERNALISM: NEW HOPE FOR THE SKEPTIC*

STEFAN RINNER

Munich Center for Mathematical Philosophy, Germany

ABSTRACT

Different thought experiments have been offered to argue for the skeptical claim that sound empirical knowledge is impossible. One of these thought experiments assumes that we are eternal brains in a vat with systematically delusory experiences. In (Putnam 1981), Putnam responds to the skeptical challenge that contrary to our initial assumption we can know a priori, i.e. independent from experience, that we aren't eternal brains in a vat. Putnam argues that the skeptical hypothesis that we are eternal brains in a vat is inconsistent with the received view regarding reference and truth, semantic externalism, which says that a referential expression e refers to an object o if and only if e is appropriately causally related to o. There are different versions of Putnam's argument. In this paper, I will discuss the three main versions of the argument; i.e. a reconstruction of Putnam's original argument in (Putnam 1981), Brueckner's simple argument (Brueckner 2003; 2016, Section 3 and 4), and a reconstruction of Brueckner's disjunctive argument (Brueckner 2016, Section 4). It is generally assumed that Putnam's original argument does not show that the skeptical hypothesis that we are eternal brains in a vat is inconsistent with semantic externalism. In this paper, I will argue that the same is true of Brueckner's simple argument and of Brueckner's disjunctive argument. Although from this it won't follow that semantic externalism is consistent with the skeptical hypothesis, it will show that it is also not yet decided that it is not.

Keywords: *Brain-in-a-vat Scenario, Skeptical Challenge, Semantic Externalism, Hilary Putnam*

1. Introduction

Different thought experiments have been offered to argue for the skeptical claim that sound empirical knowledge is impossible. For example, it has been argued that since the experiences of a brain in a vat whose sensory impressions are generated by a supercomputer are qualitatively indistinguishable from the experiences of a normal perceiver, we cannot rule out that we are *eternal* brains in a vat[1] with systematically delusory experiences. Moreover, it seems that if we are eternal brains in a vat with systematically delusory experiences, our beliefs concerning the external world are false. Since fallible knowledge, knowledge despite unelimated possibilities of error, seems to be a contradiction in terms, the skeptic concludes that sound empirical knowledge is impossible. This is also known as the skeptical challenge or the skeptical argument.

In (Putnam 1981), Putnam responds to the skeptical challenge that the hypothesis that we are eternal brains in a vat is inconsistent with the received view regarding reference and truth; i.e. semantic externalism. According to semantic externalism, a referential expression *e* refers to an object/kind *o* if and only if *e* is appropriately causally related to *o*. This means that there is a causal relation of the right kind between *e* and *o*; e.g. a Kripkean chain of communication. Since we can know a priori that semantic externalism is true, Putnam's argument goes, we can know a priori that we aren't eternal brains in a vat.

There are different versions of Putnam's argument. In this paper, I will discuss the three main versions of the argument; i.e. a reconstruction of Putnam's original argument in (Putnam 1981), Brueckner's simple argument (Brueckner 2003; 2016, Section 3 and 4), and a reconstruction of Brueckner's disjunctive argument in (Brueckner 2016, Section 4). It is generally assumed that Putnam's original argument does not show that the skeptical argument is not sound; i.e. that one of its premises is false. In this paper, I will argue that the same is true of Brueckner's simple argument and of Brueckner's disjunctive argument. Although from this it won't follow that semantic externalism is consistent with the premises of the skeptical argument, it will show that it is also not yet decided that it is not.

---

[1] Eternal brains in a vat are brains that have been envatted for their entire life. The alternative to an eternal brain in a vat is a recently envatted brain in a vat. In this paper, I will concentrate on the eternal-envatment scenario. First of all, the recent-envatment scenario lacks the skeptical power of the eternal-envatment scenario, since as a recently envatted brain in a vat I still have true beliefs concerning the external world like my belief that I was born in Italy (Brueckner 2016, Section 4). Second of all, in this paper, I will concentrate on Putnam's argument against the skeptical challenge, and Putnam's argument is only an argument against the claim that we are eternal brains in a vat.

The paper is structured as follows. In section 1, I will present the skeptical argument in more detail. In section 2, I will discuss Putnam's original argument and Brueckner's objection to Putnam's argument that the argument is epistemically circular (Brueckner 1986). Following this (section 3), I will present Brueckner's simplification of Putnam's argument, and I will argue that neither Brueckner's simple argument nor Putnam's original argument is epistemically circular.

In section 4, I will argue that although neither Putnam's argument nor Brueckner's simple argument is epistemically circular, neither Putnam's argument nor Brueckner's simple argument shows that the skeptical hypothesis that we are eternal brains in a vat is inconsistent with semantic externalism. For example, following (Brueckner 2016, Section 4), I will argue that both Brueckner's simple argument and Putnam's argument contain a premise whose truth can only be known via experience. Since an argument against the skeptical claim that sound empirical knowledge is impossible should not contain premises whose truth can only be known via experience, it will follow that neither Putnam's argument nor Brueckner's simple argument shows that the skeptical hypothesis that we are eternal brains in a vat is inconsistent with semantic externalism. I will call this 'the problem of a posteriori truths'.

To avoid the problem of a posteriori truths, following the simple argument, Brueckner (2016, Section 4) presents a disjunctive argument. In section 5, I will try to reconstruct Brueckner's disjunctive argument. Following this (section 6), I will argue that although Brueckner's disjunctive argument avoids the problem of a posteriori truths, just like Putnam's argument and Brueckner's simple argument, it does not show that one of the premises of the skeptical argument is false. This will question the claim that semantic externalism is not consistent with the premises of the skeptical argument.

2. The Skeptical Challenge

The brain-in-a-vat thought experiment assumes that we are *eternal* brains in a vat whose sensory impressions are generated by a supercomputer; i.e. the supercomputer produces electrical impulses that stimulate the brain just like normal brains are stimulated as a result of perceiving external objects. Therefore, we can assume that the experiences of an eternal brain in a vat are qualitatively indistinguishable from the experiences of a normal perceiver. For example, the computer-generated experiences of trees are qualitatively indistinguishable from the experiences of trees of a normal perceiver. Following this, the skeptic argues as follows for the claim that sound empirical knowledge is impossible.

If the experiences of an eternal brain in a vat are qualitatively indistinguishable from the experiences of a normal perceiver, then,

according to the skeptic, we cannot rule out that we are eternal brains in a vat *with systematically delusory experiences*.

> (*A1*) I don't know that I am not an eternal brain in a vat with systematically delusory experiences.

For example, the computer program features would have us believe that we have hands, can walk etc. I will call an eternal brain in a vat with systematically delusory experiences a '*BIV*', and I will call the claim that I am an eternal brain in a vat with systematically delusory experiences 'the skeptical hypothesis'.

> (*SH*) I am an eternal brain in a vat with systematically delusory experiences.

If I am a *BIV*, the skeptic continues, then the propositions concerning the external world which I think I know to be true are false.

> (*A2*) For an arbitrary proposition *p* concerning the external world which I think I know to be true: If I am a *BIV*, then *p* is false.

Take, for example, my belief that I have hands. If based on the computer-generated experiences I come to believe the proposition that I have hands, then, according to the skeptic, I am sadly mistaken. Since, according to (*A1*), I cannot rule out that I am a *BIV*, it would follow that I cannot rule out that the propositions concerning the external world which I think I know to be true are false. Since fallible knowledge, knowledge despite uneliminated possibilities of error, seems to be a contradiction in terms, the skeptic concludes that sound empirical knowledge is impossible.

There are several possible responses to the above argument. For example, Lewis (1979) modifies the claim that knowledge implies that all possibilities of error have been ruled out by replacing (*A3*) with (*A3**).

> (*A3*) An agent *A* knows a proposition *p* if and only if *p* holds in every (epistemic) possibility left uneliminated by *A*'s evidence.

> (*A3**) An agent *A* knows a proposition *p* if and only if *p* holds in every (epistemic) possibility left uneliminated by *A*'s evidence – Psst! – except for those possibilities that we are properly ignoring.

Here an epistemic possibility with respect to *A*'s evidence is a (metaphysically) possible world *w* such that *A*'s evidence does not rule out that *w* is actual. Following this, Lewis notes that unless we deal with epistemology and skeptical scenarios we usually ignore the possibility that we are eternal brains in a vat. According to such a solution, 'know' is context-dependent. For example, with an utterance of 'I know that I have hands' I would say something true as long as my interlocutors and I ignore skeptical scenarios like the brain-in-a-vat scenario.

However, following (Putnam 1981), the standard response to the skeptical argument is that we can rule out a priori that we are *BIV*s. As said above, Putnam argues that the skeptical hypothesis that we are *BIV*s is inconsistent with the received view regarding reference and truth; i.e. semantic externalism. Since, according to Putnam, semantic externalism is a priori true, Putnam concludes that we can rule out a priori that we are *BIV*s. In the next section, I will examine Putnam's argument in more detail.

3. Putnam's Argument

The claim that a referential expression *e* refers to an object/kind *o* if and only if *e* is appropriately causally related to *o* goes back to Putnam (1975), Kripke (1980) and Donnellan (1970). For example, Kripke and Donnellan argue that the referent of an ordinary name like 'Napoleon' is not determined by the properties that the speaker associates with the name, but by a so-called chain of communication. Kripke extends this claim to natural kind terms like 'tiger', 'water' and 'brain'. Since, in this paper, I will examine the question *whether the skeptical hypothesis that I am a BIV is not consistent with semantic externalism*, in what follows I will simply assume that semantic externalism (*SE*) is true.

> (*SE*) My utterances of a referential expression *e* refer to an object/kind *o* if and only if they are appropriately causally related to *o*.

Following (*SE*), Putnam (1981) argues as follows for the claim that the skeptical hypothesis that I am a *BIV* is false. Assume that I am a *BIV*. Then my utterances of 'brain' are not appropriately causally related to brains, but to one of the following candidates (Brueckner 1986, cf. 2012, 11):

> (*i*) to the succession of experiences had by the *BIV* which are comparable to the experiences of brains of a normal perceiver,

> (*ii*) to the electrical impulses that stimulate the brain and thereby cause it to have experiences just like those described in (*i*), or

> (*iii*) to the computer program features that are causally responsible for the stimuli described in (*ii*).

Lets say that 'brain*' stands for one of the three candidates. Together with semantic externalism, it follows that my utterances of 'I am not a *BIV*' are true if and only if it is not the case that I am a *BIV**. Since, if I am a *BIV*, it is not the case that I am a *BIV**, it follows that as a *BIV* I say something true with my utterances of 'I am not a *BIV*'.

According to Putnam (1981, 7-8), the above argument has the following logical structure:

(1.) I am a *BIV*. (Assumption, Conditional Proof (CP))

(n.) I am not a *BIV*.

(n+1) If I am a *BIV*, then I am not a *BIV*. (1−(n+1), CP)

(∴) I am not a *BIV*. (from (n+1))

However, this presupposes that (*DW*) is true.

(*DW*) My utterances of 'I am not a *BIV*' are true if and only if I am not a *BIV*.

If (*DW*) is true, then we get the following argument for the claim that I am not a *BIV* (PA):

(1.) I am a *BIV*. (Assumption, CP)

(2.) If I am a *BIV*, then my utterances of 'I am not a *BIV*' are true if and only if I am not a *BIV**. (from (SE) and the Skeptical Scenario)

(3.) If I am a *BIV*, then I am not a *BIV**.

(4.) If I am a *BIV*, then my utterances of 'I am not a *BIV*' are true. (from 2 and 3)

(5.) My utterances of 'I am not a *BIV*' are true. (from 1 and 4)

(6.) My utterances of 'I am not a *BIV*' are true if and only if I am not a *BIV*. (from (DW))

(7.) I am not a *BIV*. (from 5 and 6)

(8.) If I am a *BIV*, then I am not a *BIV*. (1−7, CP)

(∴) I am not a *BIV*. (from 8)

PA is logically valid. However, Brueckner (1986, cf. 2012, 24-5) objects to *PA* that in order to know that (*DW*) is the correct statement of the truth-conditions of my utterances of 'I am not a *BIV*', and not (*DW**), I need to know that I am not a *BIV* (speaking English).

(*DW**) My utterances of 'I am not a *BIV*' are true if and only if I am not a *BIV**.

Therefore, following (Alston 1989), Brueckner (2016, Section 2) calls Putnam's argument *epistemically circular*; knowledge of one of its premises requires knowledge of its conclusion.

Before I will discuss Brueckner's objection to Putnam's argument in more detail, I will briefly present Brueckner's version of Putnam's argument. Following this, I will argue that neither Brueckner's argument nor Putnam's argument is epistemically circular.

4. The Simple Argument

In (Brueckner 2003) and (Brueckner 2016, Section 3), Brueckner proposes the following simplification (*SA*) of Putnam's argument:[2]

>   (1.) My word 'tree' refers to trees.
>
>   (2.) If I am a *BIV*, then it is not the case that my word 'tree' refers to trees.
>
>   (∴) I am not a *BIV*. (from 1 and 2)

For premise (2) we can argue again with semantic externalism:

>   (1.) If I am a *BIV*, then my word 'tree' is not causally related trees. (from the Skeptical Scenario)
>
>   (2.) If my word 'tree' is not causally related to trees, then my word 'tree' does not refer to trees. (from (*SE*))
>
>   (∴) If I am a *BIV*, then it is not the case that my word 'tree' refers to trees. (from 1 and 2)

Premise (1) follows from the description of the skeptical scenario. If I am a *BIV*, then my word 'tree' is not causally related to trees, but only to trees*. Premise (2), on the other hand, follows again from semantic externalism.

What about premise (1) of *SA*? Prima facie, premise (1) of *SA* leads to the same problem as premise (6) of *PA*. It seems that in order to know that my word 'tree' refers to trees, and not to trees*, I need to know that I am not a *BIV* (speaking English).

>   (*DR*) My word 'tree' refers to trees.
>
>   (*DR**) My word 'tree' refers to trees*.

However, Brueckner (2016, Section 4) notes that in order to know that (*DR*) is true I only need to know the meaning of 'refer' and the meaning of quotations marks. Then, I can know (a priori) that disquotation can be correctly applied to any referring term of my language, and that sentences like (*DR*) are true. Moreover, since disquotational principles like (*DR*) are metaphysically neutral, (*DR*) does not exclude that (*DR**) is true. Therefore, in order to know that (*DR*) is true, I don't need to know that I am not a *BIV*.

However, the same seems to be true of (*DW*). For example, following the above argument, an advocate of *PA* could argue that since we know the meaning of 'true' and the meaning of quotation marks, we can know (a

---

[2] The simple argument goes back to (Wright 1992). It is also Putnam's favourite version of his argument (Putnam 1994).

priori) that disquotation can be correctly applied to any truth-bearing sentence of our language, and that sentences like (*DW*) are true. Moreover, since, just like (*DR*), (*DW*) is metaphysically neutral, (*DW*) does not exclude that (*DW**) is true. It follows that in order to know that (*DW*) is true I don't need to know that I am not a *BIV*.

We see that contrary to Brueckner's assumption neither Brueckner's argument nor Putnam's argument seems to be epistemically circular. However, if I know the meaning of 'refer' and the meaning of quotation marks, then I only know that disquotation can be correctly applied to any *successfully referring* term of my language. Therefore, strictly speaking, a priori, I can only know that (1*) is true.

(1*) If my word 'tree' refers, then it refers to trees.

Following this, we have to reformulate Brueckner's simple argument as follows (Brueckner 2003; 2016, Section 4):

(1.*) If my word 'tree' refers, then it refers to trees. (Disquotation)

(2.*) If I am a *BIV*, then it is not the case that if my word 'tree' refers, then it refers to trees. (from (*SE*) and the Skeptical Scenario)

(∴) I am not a *BIV*. (from 1* and 2*)

A similar problem arises in connection with Putnam's original argument. If I know the meaning of 'true' and the meaning of quotation marks, then I only know that disquotation can be correctly applied to any truth-bearing sentence of my language. This means that strictly speaking, a priori, I can only know that (6*) is true.

(6*) If my utterances of 'I am not a *BIV*' are truth-bearing, then they are true if and only if I am not a *BIV*.

Following this, we have to reformulate Putnam's argument as follows:

(1.) I am a *BIV*. (Assumption, CP)

(2.) If I am a *BIV*, then my utterances of 'I am not a *BIV*' are true if and only if I am not a *BIV**. (from (*SE*) and the Skeptical Scenario)

(3.) If I am a *BIV*, then I am not a *BIV**.

(4.) If I am a *BIV*, then my utterances of 'I am not a *BIV*' are true. (from 2 and 3)

(5.) My utterances of 'I am not a *BIV*' are true. (from 1 and 4)

(6.*) If my utterances of 'I am not a *BIV*' are truth-bearing, then my utterances of 'I am not a *BIV*' are true if and only if I am not a *BIV*. (Disquotation)

(7) If my utterances of 'I am not a *BIV*' are true, then they are truth-bearing.

(8.) I am not a *BIV*. (from 5, 6* and 7)

(9.) If I am a *BIV*, then I am not a *BIV*. (1−8, CP)

(.:) I am not a *BIV*. (from 9)

We see that both Brueckner's simple argument and Putnam's argument are still logically valid if we replace premise (1) with premise (1*) and if we replace premise (6) with premise (6*) (respectively). Moreover, since both (*DR*) and (*DW*) are simply instances of Tarski's disquotational principles for reference and truth, neither Putnam's argument nor Brueckner's argument seems to be epistemically circular. Nevertheless, in the next section, we will see that neither Putnam's argument nor Brueckner's argument shows that I am not a *BIV*. For example, we will see that both Putnam's argument and Brueckner's argument contain a premise whose truth can only be known a posteriori; i.e via experience.

5. The Problem of A Posteriori Truths

As we have seen above, we have to reformulate Brueckner's simple argument as follows:

(1.*) If my word 'tree' refers, then it refers to trees. (Disquotation)

(2.*) If I am a *BIV*, then it is not the case that if my word 'tree' refers, then it refers to trees. (from (*SE*) and the Skeptical Scenario)

(.:) I am not a *BIV*. (from 1* and 2*)

Premise (2*) is logically equivalent to (*T*):

(*T*) If I am a *BIV*, then my word 'tree' refers, and it is not the case that my word 'tree' refers to trees.

This leads to the question, how an advocate of semantic externalism could argue for premise (2*). For example, the following won't do.

(1.) If I am a *BIV*, then my word 'tree' is not causally related trees. (from the Skeptical Scenario)

(2.) If my word 'tree' is not causally related to trees, then my word 'tree' refers, and it is not the case that my word 'tree' refers to trees.

(. :) If I am a *BIV*, then my word 'tree' refers, and it is not the case that my word 'tree' refers to trees. (from 1 and 2)

From the fact that my word 'tree' is not causally related to trees, it does not follow together with semantic externalism that my word 'tree' refers. Therefore, premise (2) does not follow from semantic externalism.

An advocate of Brueckner's simple argument seems to be left with the following option:

(1.) If I am a *BIV*, then my word 'tree' is only appropriately causally related to trees*. (from the Skeptical Scenario)

(2.) If my word 'tree' is only appropriately causally related to trees*, then it is not appropriately causally related to trees.

(3.) If my word 'tree' is appropriately causally related to trees*, then my word 'tree' refers. (from *SE*)

(4.) If my word 'tree' is not appropriately causally related to trees, then it is not the case that my word 'tree' refers to trees. (from *SE*)

(. :) If I am a *BIV*, then my word 'tree' refers, and it is not the case that my word 'tree' refers to trees. (from 1, 2, 3 and 4)

Both premise (3) and premise (4) follow from semantic externalism. However, premise (2) is problematic.

As Brueckner (2016, Section 4) rightly points out, premise (2) presupposes that (*G1*) is true.

(*G1*) Trees ≠ trees*.

Prima facie, this seems to be unproblematic, since it seems to be obvious that trees* are not trees. However, Brueckner notes that we can only know via experience that trees are not computer program features. Therefore, we can only know via experience that trees* are not trees. Since an argument against the skeptical claim that sound empirical knowledge is impossible should not contain premises whose truth can only be known via experience, it follows that Brueckner's simple argument does not show that I am not a *BIV*.

A similar problem also arises in connection with *PA*. As we have seen above, *PA* contains the following premise:

(3.) If I am a *BIV*, then I am not a *BIV**.

This presupposes that (*G2*) is true.

(*G2*) *BIV*s ≠ *BIV*s*.

However, since we can only know via experience that *BIV*s are not computer program features, also (*G2*) is a posteriori true if true. I will call this 'the problem of a posteriori truths'.

To avoid the problem of a posteriori truths, following the simple argument, Brueckner (2016, Section 4) presents a disjunctive argument. Therefore, next, I will try to reconstruct Brueckner's disjunctive argument, and I will show how it avoids the problem of a posteriori truths.

6. Brueckner's Disjunctive Argument

In (Brueckner 2016, Section 4), Brueckner addresses the problem of a posteriori truths by drawing our attention to (*A2*).

> (*A2*) For an arbitrary proposition p concerning the external world which I think I know to be true: If I am a *BIV*, then *p* is false.

For example, Brueckner argues that if (*G1*) is true, then (*A2**) is true.

> (*A2**) There is a proposition *p* concerning the external world which I think I know to be true such that if I am a *BIV*, then *p* is true.

It would follow that (*G1*) is not consistent with (*A2*).

Brueckner's argument runs as follows. Assume that trees are trees*. Together with (*DG*), it would follow that (*DG**) is true.

> (*DG*) The proposition that I am in the presence of trees is true if and only if I am in the presence of trees.

> (*DG**) The proposition that I am in the presence of trees is true if and only if I am in the presence of trees*.

Moreover, if (*DG**) is true, then the proposition that I am in the presence of trees is true if I am a *BIV*, since, as a *BIV*, I am in the presence of trees*. It would follow that in order to know the proposition that I am in the presence of trees I don't have to rule out that I am a *BIV*.

Following this, Brueckner (2016, Section 4) proposes a disjunctive argument which can be reconstructed as follows:

> (1.) Either trees = trees*, or trees ≠ trees*.

> (2) Trees ≠ trees*. (Assumption, Disjunctive Proof (*DP*))

> (2.1) If my word 'tree' refers, then it refers to trees. (Disquotation)

> (2.2) If I am a *BIV*, then my word 'tree' is only appropriately causally related to trees*. (from the Skeptical Scenario)

> (2.3) If my word 'tree' is only appropriately causally related to trees*, then it is not appropriately causally related to trees. (from 2)

(2.4) If my word 'tree' is appropriately causally related to trees*, then my word 'tree' refers. (from *SE*)

(2.5) If my word 'tree' is not appropriately causally related to trees, then it is not the case that my word 'tree' refers to trees. (from *SE*)

(2.6) If I am a *BIV*, then my word 'tree' refers, and it is not the case that my word 'tree' refers to trees. (from 2.2, 2.3, 2.4 and 2.5)

(2.7) I am not a *BIV*. (from 2.1 and 2.6)

(3.)   Trees = trees*. (Assumption, *DP*)

(3.1) The proposition that I am in the presence of trees is true if and only if I am in the presence of trees.

(3.2) The proposition that I am in the presence of trees is true if and only if I am in the presence of trees*. (from 3 and 3.1)

(3.3) If I am a *BIV*, then I am in the presence of trees*.

(3.4) If I am a *BIV*, then the proposition that I am in the presence of trees is true. (from 3.2 and 3.3)

(3.5) I believe the proposition that I am in the presence of trees. (Introspection)

(3.6) The proposition that I am in the presence of trees is a proposition concerning the external world.

(3.7) There is a proposition p concerning the external world which I think I know to be true such that if I am a *BIV*, then *p* is true. (from 3.4, 3.5 and 3.6)

(. :) I am not a *BIV*, or there is a proposition *p* concerning the external world which I think I know to be true such that if I am a *BIV*, then *p* is true. (from 2.7 and 3.7, *DP*)

Premise (1) is a priori true, since we can know a priori that either trees = trees*, or trees ≠ trees*. With the above argument, Brueckner infers from this that I am not a *BIV*, or there is a proposition *p* concerning the external world which I think I know to be true such that if I am a *BIV*, then *p* is true. According to Brueckner, the skeptic loses in each case.

Prima facie, the questionable premises of the above reconstruction of Brueckner's disjunctive argument are (3.1) and (3.3). For example, the skeptic could object that (3.1) presupposes that the proposition that I am in the presence of trees is about trees, and not about trees*. Therefore, the objection continues, (3.1) presupposes that I am not a *BIV*. However, just like (*DW*) and (*DR*), (3.1) is metaphysically neutral. This means that (3.1) does not exclude that trees = trees*, and, therefore, (3.1) does not presuppose that I am not a *BIV*.

For premise (3.3), on the other hand, we can argue as follows:

(1) I have *treeish* experiences. (Introspection)

(2) If I am a *BIV* and I have *treeish* experiences, then I am in the presence of trees*. (from the Skeptical Scenario)

(∴) If I am a *BIV*, then I am in the presence of trees*.(from 1 and 2)

Premise (2) follows from the description of the skeptical scenario. If I am a *BIV*, then my *treeish* experiences are generated by computer program features. Since these computer program features are trees*, it follows that as a *BIV* I can only have *treeish* experiences if I am in the presence of trees*.

We see that by drawing our attention to the dialectical situation between skeptic and anti-skeptic Brueckner can avoid the problem of a posteriori truths. Nevertheless, next, we will see that Brueckner's disjunctive argument is still problematic in the light of the skeptical challenge presented in section 1.

7. The Skeptical Challenge Again

In section 1, we saw that the skeptic can argue as follows for the claim that sound empirical knowledge is impossible:

(*A1*) I don't know that I am not a *BIV*.

(*A2*) For an arbitrary proposition p concerning the external world which I think I know to be true: If I am a *BIV*, then *p* is false.

(*A3*) An agent *A* knows a proposition *p* if and only if *p* holds in every (epistemic) possibility left uneliminated by *A*'s evidence.

(∴) For an arbitrary proposition *p* concerning the external world which I think I know to be true: I don't know *p*.

Therefore, if we want to avoid the skeptic's conclusion that sound empirical knowledge is impossible, we have to show that one of the premises of the above argument is false. As we have seen, with Putnam's argument and with Brueckner's simple argument advocates of semantic externalism tried to show that contrary to the skeptic's assumption I can know that I am not a *BIV*. However, since both arguments contain a premise whose truth can only be known a posteriori, i.e. via experience, neither Putnam's argument nor Brueckner's simple argument can be used to argue against the skeptical claim that sound empirical knowledge is impossible.

What about Brueckner's disjunctive argument? Does Brueckner's disjunctive argument show that one of the premises of the skeptical argument is false? I will argue that it does not. Brueckner's disjunctive argument only shows that I am not a *BIV*, or that there is a proposition *p* concerning the external world which I think I know to be true such that *p* is true if I am a *BIV*. From this it does not follow that I know that I am not a *BIV*, or there is a proposition *p* concerning the external world which I think I know to be true such that *p* is true if I am a *BIV*. Therefore, Brueckner's disjunctive argument does not show that (*A1*) or (*A2*) is false. Moreover, the argument neither shows that (*A2*) is false, nor that (*A1*) is false. For example, following Brueckner's disjunctive argument, I only know (a priori) that I am not a *BIV*, or that there is a proposition *p* concerning the external world which I think I know to be true such that *p* is true if I am a *BIV*. Hence, following Brueckner's disjunctive argument, I still cannot rule out (a priori) that I am not a *BIV*.

An advocate of Brueckner's disjunctive argument could respond that Brueckner's disjunctive argument shows that if I am a *BIV*, then there is a proposition *p* concerning the external world which I think I know to be true such that *p* is true if I am a *BIV*. In other words, the argument shows that if (*A2*) is true, then I am not a *BIV*. Following this, an advocate of Brueckner's disjunctive argument could argue as follows against the skeptical argument.

I know a priori that either (*A2*) is true, or that (*A2*) is false. Assume that (*A2*) is false. Then, the skeptical argument is not sound. Next, assume that (*A2*) is true. Then, it follows together with the conclusion of Brueckner's disjunctive argument that I am not a *BIV*. Therefore, an advocate of Brueckner's disjunctive argument may conclude that Brueckner's disjunctive argument shows that either (*A2*) is false, or I can rule out that I am a *BIV*. Either way, it would follow that the skeptical argument is not sound.

However, if we look at the argument in more detail, we see that it does not show that the skeptical argument is not sound, but that it only shows that either the skeptical argument is not sound, or I am not a *BIV*.

(1) Either (*A2*) is true, or (*A2*) is false.

(2) (*A2*) is false. (Assumption, *DP*)

(2.1) If (*A2*) is false, the skeptical argument is not sound.

(2.2) The skeptical argument is not sound. (from 2 and 2.1)

(3) (*A2*) is true. (Assumption, *DP*)

(3.1) If (*A2*) is true, then I am not a *BIV*. (Brueckner's Disjunctive Argument)

(3.2) I am not a *BIV*. (from 3 and 3.1)

(. :) Either the skeptical argument is not sound, or I am not a *BIV*.
   (from 2.2 and 3.2, *DP*)

From this it does not follow that either the skeptical argument is not sound, or *I know* that I am not a *BIV*. This was to be expected, since also Brueckner's disjunctive argument only shows that (*A2*) is false, or I am not a *BIV*. As said above, from this it does not follow that (*A2*) is false, or I know that I am not a *BIV*.

Let us recap: Brueckner's disjunctive argument neither shows that (*A1*) is false, nor that (*A2*) is false, nor that (*A1*) or (*A2*) is false. It follows that just like Putnam's argument and Brueckner's simple argument Brueckner's disjunctive argument does not show that one of the premises of the skeptical argument is false. Therefore, also Brueckner's disjunctive argument does not avoid the skeptic's conclusion that sound empirical knowledge is impossible.

8. Conclusion

Although neither Brueckner's simple argument nor Putnam's argument seems to be epistemically circular, neither Brueckner's simple argument nor Putnam's argument shows that the skeptical hypothesis that I am a *BIV* is inconsistent with semantic externalism. For example, both Brueckner's simple argument and Putnam's argument contain a premise whose truth can only be known via experience. Since an argument against the skeptical claim that sound empirical knowledge is impossible should not contain premises whose truth can only be known via experience, it follows that neither Brueckner's simple argument nor Putnam's argument shows that (*A1*) is false.

Following this, Brueckner argues for the weaker claim that if semantic externalism is true, then I don't have to rule out that I am a *BIV* in order to possess sound empirical knowledge, or I am not a *BIV*. According to Brueckner, the skeptic loses in each case. However, as we have seen above, just like Putnam's argument and Brueckner's simple argument, Brueckner's disjunctive argument does not show that one of the premises of the skeptical argument is false. Although from this it does not follow that semantic externalism is consistent with the premises of the skeptical argument, it shows that it is also not yet decided that it is not.

REFERENCES

Alston, W. 1989. Epistemic Circularity. In *Epistemic Justification: Essays in the Theory of Knowledge*, W. Alston. Ithaca: Cornell University Press.

Brueckner, A. 1986. Brains in a Vat. *Journal of Philosophy* 83: 148-167. Reprinted in (Brueckner A. and Ebbs G. 2012), 9-27.

_____. 2003. Trees, Computer Program Features, and Skeptical Hypotheses. In *The Skeptics: Contemporary Essays*, ed. S. Luper. Burlington: Ashgate.

_____. 2016. Skepticism and Content Externalism. *The Stanford Encyclopedia of Philosophy* Edited by Edward N. Zalta. Accessed March 6, 2017. https://plato.stanford.edu/archives/win2016/entries/skepticism-content-externalism/.

Brueckner, A. and Ebbs G. 2012. *Debating Self-Knowledge*. Cambridge: Cambridge University Press.

Donnellan, K. S. 1970. Proper Names and Identifying Descriptions. *Synthese* 21: 335- 358.

Kripke, S. 1980. *Naming and Necessity*. Oxford: Blackwell.
Lewis, D. 1979. Attitudes De Dicto and De Se. *The Philosophical Review* 88: 513-543.

Putnam, H. 1975. The Meaning of 'Meaning'. In *Language, Mind, and Knowledge*, ed. K. Gunderson, 131-93, Minneapolis, Minnesota: University of Minnesota Press.

_____. 1981. Brains in a Vat. In *Reason, Truth, and History*, H. Putnam, 1-21. Cambridge: Cambridge University Press.

_____. 1994. Crispin Wright on the Brain-in-a-vat Argument. In *Reading Putnam*, ed. P. Clark, B. Hale, 283-288. Cambridge MA: Blackwell.

Wright, C. 1992. On Putnam's Proof That We Are Not Brains-in-a-Vat. *Proceedings of the Aristotelian Society* 92: 67-94.

DREAMS IN A VAT*

DANILO ŠUSTER
University of Maribor, Slovenia

ABSTRACT

Putnam's semantic argument against the BIV hypothesis and Sosa's argument against dream skepticism based on the imagination model of dreaming share some important structural features. In both cases the skeptical option is supposed to be excluded because preconditions of its intelligibility are not fulfilled (affirmation and belief in the dream scenario, thought and reference in the BIV scenario). Putnam's reasoning is usually interpreted differently, as a classic case of deception, but this feature is not essential. I propose to interpret BIV's utterances as cases of reference failure best captured by truth-value gaps. Both anti-skeptical strategies are then vulnerable to the same type of objections (how do we know what state we are in or how do we know what kind of language do we speak).

Keywords: *Putnam, Sosa, brain in a vat, dream argument, disquotation, negation*

"To see a vat in your dreams, foretells anguish and suffering from the hands of cruel persons, into which you have unwittingly fallen."[1]

---

[1] *Psychologist World*, https://www.psychologistworld.com/dreams/dictionary/vat (accessed March 30, 2017).

1. Introduction

Putnam's famous thought experiment of brains in vats (BIVs, for short) is usually considered a contemporary version of Descartes's skeptical argument of the Evil Genius.[2] Recall: "... some evil spirit, supremely powerful and cunning, has devoted all his efforts to deceiving me. ... What truth then is left? Perhaps this alone, that nothing is certain." (Descartes 2008, 16). According to the hypothesis advanced by Putnam's skeptic, the universe, by accident, just happens to consist of automatic machinery tending a vat full of brains. In the BIV world everyone is raised as brains in vats, and their perceptual input is qualitatively just like ours. How do we know that we are not in this miserable predicament? The matrix of the skeptical argument is familiar: In the *bad* situation (dream, BIV world ...) it is possible to have the best experiential evidence that *p*, yet *p* is false (where *p* is a proposition about the external world). But, given the description of the scenario, the possibility of a bad situation cannot be excluded. So we do not know that *p*.

Putnam argues from some plausible assumptions about the nature of reference to the conclusion that it is not possible that all sentient creatures are brains in a vat. If we were brains in a vat in this way, we could not say or think that we were, the story "cannot possibly be true, because it is, in a certain way, self-refuting" (Putnam 1981, 7). Putnam's argument is based on an analysis of the truth conditions for the sentences uttered (or thought) by a BIV. These conditions depend on the assignments of references which one would make in evaluating the truth value of BIV's utterances. According to semantic externalism when S uses a referring term, she refers to whatever typically causes her uses of that term. So 'tree' refers to trees-in-the-image in vat-English, or something related (computer program features), and 'vat' refers to vats-in-the-image in vat-English, or something related (Putnam 1981, 15):

> It follows that if [the brains'] 'possible world' is really the actual one, and we are really the brains in a vat, then what we now mean by 'we are brains in a vat' is that *we are brains in a vat in the image* or something of that kind (if we mean anything at all). But part of the hypothesis that we are brains in a vat is that we aren't brains in a vat in the image (i.e. what we are 'hallucinating' isn't that we are brains in a vat). So, if we are brains in a vat, then the sentence 'we

---

are brains in a vat' says something false (if it says anything). In short if we are brains in a vat then 'we are brains in a vat' is false. So it is (necessarily) false.

Putnam also remarks that his "procedure has a close relation to what Kant called a 'transcendental' investigation; for it is an investigation, of the preconditions of reference and hence of thought — preconditions built in to the nature of our minds themselves" (Putnam 1981, 16).

What is the role of global *deception*, which seems to be inherited from the Cartesian tradition? How to understand the possibility that the "envatted" words might not mean anything at all? And what, exactly, is the anti-skeptical potential of Putnam's *transcendental* procedure? I will explore these questions by comparing Putnam's anti-BIV strategy with Sosa's (and late Wittgenstein's) reply to dream skepticism, where the threat is that we are unable to distinguish waking life from the corresponding dream. Both anti-skeptical strategies are based on the allegedly conceptual "impossibility" of a bad scenario (dream or BIV) for *us*, but both are deficient as a final antidote to radical skepticism, for roughly the same reasons. Or so I will try to argue. I do not claim particular originality in my critical assessment. Still, although the parenthetical possibility of "not meaning anything at all" with one's utterances when in the BIV scenario has been noticed, it has not been sufficiently explored. Especially not in connection with Sosa's "transcendental" strategy against dream scepticism (so far as I know).[3]

2. Dream skepticism

Let me start with a little piece of fiction: *The Testimony of a Woodcutter Questioned by a High Police Commissioner*.[4]

The old man was found dead on a bench in a grove. His death was considered suspicious and the woodcutter strolling in the vicinity is questioned by the police commissioner.

The woodcutter

How did the old man die? This is a strange story. After a long walk in a grove the old man got tired and sat on the bench to rest himself. He fell asleep and dreamt that he was an armored medieval knight fighting with another knight to earn the hand of a beautiful princess. In the very moment when the two of them, on

---

[3] Putnam (1981) is cited by 4974 according to *Google Scholar* (2017) and the number is growing almost on a daily basis.

[4] Some will perhaps recognize – intentional – similarities with "In a Bamboo Grove", a short story by Ryunosuke Akutagawa (1915), filmed as *Rashomon*, by Akira Kurosawa (1950).

horses, clashed with their spears, it started to rain outside. I was walking nearby and being sure the man would not like to get wet, I poked him gently with the umbrella, just to warn him. But the poor old man, immersed in his dream, thought that he was stabbed by a spear, this *belief* provoked a stroke and he died.

The police commissioner

You liar! You just made everything up. Suppose the story is true – then nobody could truly tell this story – the poor man died. This story is untellable. You are a liar and the prime suspect!

The woodcutter

But why don't you check my testimony with the *Testimony of the Dead Man's Spirit Told through a Medium?*[5] True, *I* might not have been in the exact position to retell the events, but it does not follow that the story is impossible.

The police commissioner (very smart, acquainted with Wittgenstein 1969 and Sosa 2007)

Absolutely no need to engage in dubious spiritual practices. Consider this (cf. Sosa 2007, 7-8):

Dreaming does not involve forming beliefs, but merely consists in imagining or simulating experiences. Dreaming resembles imagining in that, when one imagines, one does not thereby acquire beliefs, but only certain propositionally contentful states that are known as "make-beliefs". … We are guided by our imagination but have no beliefs about what we are experiencing: we don't take what we are experiencing to stand for something in reality. We do not affirm anything.

And also:

''My exposition relies heavily on distinguishing between two expressions: 'in my dream' and 'while I dream.' From the fact that in my dream something happens it does not follow that it happens while I dream. From the fact that in my dream I am chased by a lion it does not follow that while I dream I am chased'' (Sosa 2007, 4).

So, you see, there is a profound distinction between what goes on ''in the old man's dream'' (allegedly hit by the spear) and what goes on ''while he dreams'' (supposedly poked by the umbrella). It is just a conceptual confusion to think that any kind of epistemic "transaction" between the two domains is possible. His so called "testimony" could not express any beliefs at all and his dream mental states could not cause any real action. His "report" would

---

[5] Akutagawa, R. 2006. *Rashōmon and Seventeen Other Stories. Penguin Classics, p. 14.*

be just a piece of fiction, irrelevant as a description of what went on "in reality" and so of no value, if not outright nonsensical since no *genuine* statements would be made. You should rather confess!

The woodcutter (smart too!)

Your procedure has a close relation to what Kant called a "'transcendental' investigation" (Putnam 1981, 16) and not a real police investigation. But "transcendental deduction" is not any super-duper deduction, the term actually comes from German legal vocabulary – an argument intended to yield a justification on the court. And in this context – what is more convincing, my story or your ruminations about the conceptual preconditions of having *real* thoughts? Is this not enough for a reasonable doubt? Come to think of it, how do *you* know that you have real beliefs and not some kind of imaginations and quasi-beliefs right now?

We are now in the deep waters of skepticism and epistemology students will know how to continue the dialogue. The commissioner is using *modus ponens*: the story is conceptually inconsistent so excluded *apriori*, but the woodcutter returns with *modus tollens*: the story is epistemically possible, so there is no conceptual inconsistency. My philosophical sympathies are with the woodcutter – he is guilty, probably, but we are talking conceptual possibilities now. I doubt not only that conceptual investigations exclude metaphysical possibilities (this much is granted by Putnam – according to him the BIV scenario is *physically* possible), I doubt that they have the power to exclude epistemic possibilities and doubts raised by the skeptic. How could a mere reflection on our concepts (rather than on proper evidential considerations) give us some concrete information about the external world?

The skeptical threat of *dreams* is familiar from Descartes – when dreaming, a subject has misleading sensations, which typically lead to false beliefs. According to this, *hallucination* model of dreaming, when subjects dream, they undergo perceptual experiences of the same kind that they do while waking and form real beliefs about what is happening on their basis. But there is an alternative, *imagination* model of dreaming. When subjects dream, the experiences they suffer are different in kind from those involved in perception. They are exercises of the imagination, including sensory imagery and propositional imagination (Ichikawa 2016, 150). I do not want to take stands (although arguments in favor of the imagination model could be given: structural similarity between dreams and fictions; brain areas particularly implicated in dreaming seem to be the same as those involved in imagery, as opposed to those involved in perceptual experience, etc., cf. Ichikawa 2009), the issue is still controversial. I am mainly interested in epistemological consequences of accepting the imagination model as developed by Sosa (and earlier by Wittgenstein).

According to Sosa, to dream is to imagine something and not to hallucinate or falsely believe it. Even if we are unable to distinguish waking life from the corresponding dream, there can be no *deception* (false beliefs) in our dreams, because there are *no* beliefs at all. While dreaming, one does not form false beliefs, nor even consciously affirm anything, "in dreaming we do not really believe; we only make-believe" (Sosa 2007, 8). And according to Witttgenstein in dreaming we do not really assert, we only quasi-assert:

> The argument "I may be dreaming" is senseless for this reason: if I am dreaming, this remark is being dreamed as well – and indeed it is also being dreamed that these words have any meaning. [OC 383; Wittgenstein 1969, 387]

> I cannot seriously suppose that I am at this moment dreaming. Someone who, dreaming, says "I am dreaming," even if he speaks audibly in doing so, is no more right than if he said in his dream "it is raining", while it was in fact raining. Even if his dream were actually connected with the noise of the rain. [OC 676, Wittgenstein: 1969, 670]

The notion of "quasi" statements is sometimes used by the commentators (Hamilton 2014, 235):

> Someone who, while dreaming, utters the words "I may be dreaming" has made no genuine statement. Only when false does the utterance of "I may be dreaming" constitute a genuine statement; dream scepticism, like scepticism about the meanings of one's words, is self-defeating. Despite appearances, therefore, it is not a "genuine statement"; it is not false, but nonsensical. Its very expression raises doubts about whether the speaker understands what they are saying.

It is a vexed issue how to interpret Wittgenstein's cryptic remarks so my main source for this anti-skeptical strategy will be Sosa, who is much more explicit. Affirmations of wakefulness are automatically justified – the claim that one is not just dreaming must, like the *cogito*, be right if affirmed (Sosa 2007, 16). If we are awake, we affirm truly. But if we happen to be asleep and dreaming, we only have a belief *in* our dreams (not *while* we dream) and do not affirm anything (Sosa 2007, 17). So we are "automatically rationally committed to supposing" we are not dreaming, whenever we reflect upon the possibility that we might be (Sosa 2007, 20).

How does this differ from: we are automatically rationally committed to supposing we are *not* brains in a vat whenever we reflect upon the possibility that we might be? Or, as Putnam says about the BIV hypothesis: "If we can consider whether it is true or false, then it is not true... Hence it is not true" (Putnam 1981, 8) and again " 'We are brains in a vat' is necessarily false" (Putnam 1981, 15). Also, BIV mental states

are in certain respects dream-like states, being causally isolated from the environment they lack the very preconditions for being representational (according to semantic externalism). Putnam's semantic argument against the BIV hypothesis and Sosa's anti-skeptical argument against dream skepticism based on the imagination model of dreaming share some important structural ("transcendental") features. The rejection of the hallucination model seems to be the basis of Sosa's reply to dream skepticism in the same way as Putnam's externalism is the basis of his reply to BIV skepticism: no *false* beliefs because *no* real beliefs (thoughts) at all. And both seem to be vulnerable to the same type of objections raised, informally, by the woodcutter in the story.

The analog of "I am awake" in the dream scenario is the sentence "I am not a BIV" in Putnam's scenario. Consider the following condensed version of Putnam's reasoning (DeRose 2000, 124):

> If I am a BIV, then by, "I am not a BIV," I mean that I am not a BIV-in-the-image (or some closely related true thing), which is in that case true. On the other hand, if I am not a BIV, then by "I am not a BIV," I mean that I am not a BIV, which is in that case true. Thus, whether I am a BIV or whether I am not, my use of "I am not a BIV" is true. Either way, it's true; so, it's true: I'm not a BIV.

But wait – this reasoning is based on the assumption that as a BIV I have thoughts with *genuine* truth conditions ("I am not a BIV" is true). In the standard BIV scenario the opposite thought "I am brain in a vat" is *false* – we are being *deceived* and fed falsities about our poor situation (as if we are walking in the sunshine or something, we do not have images of being envatted). But the corresponding claim that one is just dreaming is *not* false but pragmatically incoherent according to Sosa (2007,16). Does this not break the analogy? Is *deception* essential for Putnam's anti-skeptical line of reasoning?

3. Deception and disquotation

It is part of the traditional skeptical challenge that a world fitting the skeptic's description would appear to its inhabitants just as *our* world appears to us. There is an implicit assumption that if the BIV possibility were actual, *all* of our beliefs would be false. Even the belief expressed by the sentence "I am a BIV." Putnam's thought experiment is usually interpreted along the lines which emphasize the role of *deception* (Brueckner 1986, 151, among others):

> The BIV's utterance [of "I am a BIV"] would be true iff he had sense impressions as of being a BIV. But by Putnam's hypothesis, a BIV never has such sense impressions. A BIV has only sense impressions as of being a normal, embodied human being moving through a richly varied world of physical objects. Thus a BIV's

utterance of 'I am a BIV' would never be true …

But if I am not a BIV, then my utterance of 'I am a BIV' is obviously false. We thus get a standard, dilemma version of the *Disjunctive Argument* (DA for short), formulated by Brueckner (first in Brueckner 1986, 154, below is the version from Brueckner 2010, 137).

(1) Either I am a BIV or I am a non-BIV.

(2) If I am a BIV, then my utterances of 'I am a BIV' are true iff I have sense-impressions as of being a BIV.

(3) If I am a BIV, then I do not have sense-impressions as of being a BIV (instead, I have sense-impressions as of being a normal, embodied human).

(4) If I am a BIV, then my utterances of 'I am a BIV' are false. [(2), (3)]

(5) If I am a non-BIV, then my utterances of 'I am not a BIV' are true iff I am a BIV.

(6) If I am a non-BIV, then my utterances of 'I am a BIV' are false. [(5)]

(7) My utterances of 'I am a BIV' are false. [(1),(4),(6)]

(8) My utterances of 'I am not a BIV' are true. [(7)]

The vat-English truth conditions of 'I am a BIV' are not satisfied because of *deception* (I am not fed experiences representing me to be a disembodied BIV). And the normal, English truth conditions of 'I am a BIV' are not satisfied for obvious reasons. So I am not a BIV. Hmm, this looks very quick.

There are two components in this standard version: (i) semantic externalism – in order for our word 'vat' etc. to refer to a particular kind of thing, it is necessary for our uses of the term to be causally connected – in an appropriate way – with things of that kind; (ii) *deception* – what we are 'hallucinating' isn't that we are brains in a vat.

The semantic core of the argument is usually interpreted as analogous to the distinction between our "water" (H2O) and Twin Earth "water" (XYZ). The Earthian's sentence "Water is clear" expresses a different thought than the corresponding Twin Earthian thought expressed by the same sentence. In the same way our sentence "We are brains in vat" expresses a different thought than the corresponding sentence uttered by the envatted brains. Their utterances have non-standard truth conditions – computer states that causally affect their brains so as to produce corresponding experiences or something like that.

But recall Putnam's initial analogy – an ant is crawling on a patch of sand and as it crawls, it traces a line in the sand which ends up looking like a caricature of Winston Churchill (Putnam 1981, 1).

Does the ant mispresent? Or represent Churchill*? The Putnamian intuition is that the caricature does not refer to or represent Churchill, because the presuppositions of successful reference are not fulfilled. The main problem with BIV mental states is not deception, but lack of proper connection. This suggests that the envatted utterances of "I am a BIV" are in certain respects like the famous (S) "The present king of France is bald." According to Strawson a speaker does not succeed in making a truth-evaluable claim by uttering (S). The logic and formal semantics of a language which contains singular terms without denotations is best captured by introducing truth value gaps – (S) is neither true nor false. And the same diagnosis would then apply to BIV's utterances.

The idea is hinted by Putnam – "the sentence 'we are brains-in-a-vat' says something false (if it says anything)." If it does not say anything then it is neither true nor false. Can we work out the anti-skeptical argument on the assumption that "We are not brains in a vat" is not false, rather, the preconditions for its having a truth value are not fulfilled? Various parts of the original argument will now be affected. Suppose I am a BIV, then the gist of the reasoning seems to be:

> My utterances of 'I am a BIV' are not true.

So,

> My utterances of 'I am a BIV' are false.

So,

> I am not a BIV.

The first inference is based on standard understanding of negation (if not true then false) and the final step on the principle of disquotation:

(Disq)   My utterances of 'I am not a BIV' are true iff I am not a BIV.

Both are now problematic. We can agree that "The present king of France is bald" is not true. But it does not follow that it is false – it is neither true nor false. Negation behaves differently, once truth value gaps are admitted as the third logical value. The negation of a truth value gap results in a truth value gap. Consider another area, where this idea usually finds its domicile. Since future is yet to be, some will say that it is not true (today) that I will be at home tomorrow at noon. But it is not false either, it is neither true nor false (according to Łukasiewicz, for instance). So we cannot conclude that I will not be at home tomorrow at noon.

Consequently we can no longer affirm (4) in the above DA version of the argument. And instead of (7) we now have:

> (7') My utterances of 'I am a BIV' are not true.

Classically, (8) still follows from (7'), but not in the logic of truth value gaps. My utterances might not be true because preconditions for being true or false are not fulfilled. Even more – what we really want is the

conclusion that I am not a BIV, not the conclusion that my sentence 'I am not a BIV' is true. Brueckner (2010, 137) already noticed that we must add a disquotation principle (Disq) in order to get the desired anti-skeptical result. Perhaps (4) – a BIV's sentences being false, which is based on (2) and (3) in DA above is then *no* longer needed at all. Once we have disquotation, deception is no longer a necessary ingredient of the anti-skeptical argument.

What happens if we drop deception from the BIV scenario? Suppose a supercomputer is running a program that affects my brain in such a way as to produce experiences representing me to be a disembodied BIV (premise 3 of DA is false). Consider the case of dreams – the analog would be a *lucid* dreamer, who "believes" her dream to be a dream. Well, according to Sosa lucid dreaming is a kind of daydreaming, one is still protected from the dream skeptic. If you lucidly dream that you face a fire you will only believe that you face a fire *in* the dream, not *while* you dream (Sosa 2007, 19-20, fn. 18). So, dreaming that you dream does not give you any beliefs about the *reality* of your situation. The same is true, I think, about the BIV scenario:

> 'But surely a community of brains-in-a-vat could work through just these thoughts, and so convince themselves quite spuriously that they were not brains-in-a-vat?' No, they could not. They might work through these words, and soundly convince themselves of something. But only creatures which are not brains-in-a-vat can have these thoughts (Wright 1992, 85).

Even if you, as a BIV, have sense-impressions as of being a BIV, you still cannot have real thoughts about the reality of your situation. Why not? Because your utterances do not disquote and your thoughts do not represent. It was soon recognized, even by Putnam himself, that *disquotation* plays the central role in the semantic version of the anti-skeptical argument (Putnam 1992, 404, fn 29):

> Here is the simplest form I know of the Brain in a Vat Argument (this form is due to Crispin Wright, based on a suggestion from me):
>
> In Vat English (the language spoken by the Brains in a Vat) "vat" does not refer to vats. [From the description of the Brain in a Vat world and the causal constraints on reference];
>
> in my language "vat" refers to vats [Disquotation applied to my own language].
>
> Therefore my language is not Vat English-i.e., I am not a Brain in a Vat.

All the work is done by the *semantic* ingredients: (i) reference to common objects like vats, and their physical properties is only possible if one has information carrying causal interactions with those objects; (ii) the disquotation scheme for reference and the predicate 'true'.

According to Putnam "one cannot refer to certain kinds of things, e.g., trees, if one has no causal interaction at all with them" (1981, 16). By hypothesis BIVs have no causal interaction with trees, brains and the like, so their terms do not refer to these things. In vat-English 'brain in a vat' does not refer to (real) brains in a vat, so disquotation fails. It is usually assumed that they have non-standard truth conditions – computer states that causally affect my brain so as to produce corresponding experiences (designated as 'BIV*'). So:

> My utterances of 'I am a BIV' are true iff I am a BIV*.

The vat-English truth conditions of my utterances are non-disquotational. But why assume *any* truth-conditions at all? A BIV's utterances might lack truth conditions altogether. This much is recognized by Brueckner who gives the following stripped-down semantic argument:[6]

(I) If I am a BIV, then my utterances do not have disquotational truth conditions.

(II) My utterances of sentences have disquotational truth conditions and express disquotational contents.

(III) I am not a BIV. [(I),(II)]

Two premises only, (I) looks OK, but how is (II) defended?

> The second premise seems as good as gold: of course my sentences have disquotational truth conditions and express disquotational contents. My utterances of 'A rabbit is present' are true iff a rabbit is present, and they express my belief that a rabbit is present (Brueckner 2010, 138-139).

Not much of an argument. In some version Brueckner analyses the following reasoning (Brueckner 2016a, 24, I have modified the numbering of premises):

> (A*) If I am a BIV, then it is not the case that if my word 'tree' refers, then it refers to trees.
>
> (B*) If my word 'tree' refers, then it refers to trees.
>
> (C) I am not a BIV.

(B*) is licensed by my knowledge of the semantics of my own language (Brueckner 2016b):

> But I do know certain things about my own language (whatever it is and wherever I am speaking it). By virtue of knowing the meaning of 'refers' and the meaning of quotation marks, I know that disquotation can be correctly applied to any successfully

---

[6] Slightly changed. Brueckner credits Loar for the remark that a BIV's utterances might lack truth conditions altogether. Still, in the main text he follows Putnam in assuming that a BIV's terms refer to computer states (BIV*) and states premise (I) as "If I am a BIV, then my utterances of sentences have non-disquotational truth conditions and express non-disquotational contents" (2010, 138).

referring term of my language, in the way that (B*) indicates for my word 'tree'. This is a priori knowledge of semantic features of my own language (whatever it is — English or vat-English). I know (A*) in virtue of my a priori, philosophical knowledge of the theory of semantic externalism and of how it applies to the case of the BIV. Knowing (A*) and (B*), I can then knowledgeably deduce that I am not a BIV.

Surely, I know the language I am speaking and I am justified in believing the disquotation principles in virtue of understanding my own language? But what if my singular terms lack reference? Do I not beg the question by assuming that my utterances disquote? The disquotation scheme for sentences is just the Tarski's schema:

(T) "P" is true if and only if P

According to the disquotational principle for the predicate 'true,' a true biconditional results whatever sentence is substituted for the variable 'P' in this scheme. If truth-value gaps are admitted, then this principle is no longer valid. Van Fraassen (1966) in his classical development of supervaluation semantics denies this principle in the cases of reference failure for singular terms (empty names). His reasoning is usually applied in the logic of vagueness, but the idea is clear. Consider (Williamson 1994, 196)

The phrase 'this dagger' may fail to single anything out when used by someone under a hallucination. Arguably, utterances such as 'This dagger is sharp' in which the phrase is used fail to say anything in this context, and so are neither true nor false. That includes complex utterances; even the biconditional ' "This dagger is sharp" is true if and only if this dagger is sharp ' fails to say anything, for it uses the phrase 'this dagger' on its right-hand side.

If we adopt for vat-English a semantic treatment of names which lack reference, then a BIV's utterances do *not* have the disquotational property – this much we knew that already. But it seems to me that disquotation cannot be automatically "licensed by my knowledge of the semantics of my own language" and it cannot be true that "disquotation is, on both the standard and skeptical hypotheses, a valid step within either English or vat-English" (Christensen 1993, 305). It is not valid in vat-English, when the BIV utterances are interpreted as neither true nor false. Moreover, it is not uncontroversially valid even in plain vernacular English containing empty names (and perhaps vague expressions). Our knowledge of semantic features of our own language cannot be *apriori* or obvious. Contrary to Christensen and Brueckner the plausibility of applying disquotation to *ourselves* really does seem to rest on first rejecting the brain-in-vat hypothesis and assuming that our terms refer and our thoughts represent.

4. Quasi-beliefs and quasi-thoughts

Forming beliefs (according to Sosa) is incompatible with dreaming, every time we actually form a perceptual belief, we are not dreaming. The preconditions of having a belief and truly affirming are not fulfilled when dreaming. In the same vein – thinking that I am a BIV is incompatible with the BIV scenario, since the preconditions of forming such a thought are not fulfilled according to Putnamian conception of meaning and intentional content of thoughts. Every time we actually make utterances with genuinely referring expressions, we are (at the very least) not in a BIV scenario. If the imagination model of dreaming is correct, then the belief that one is awake has a cogito-like status—necessarily, if one believes that one is awake, then one is awake (Sosa 2007, 20): "We can just as well affirm <I think, therefore I am awake> as <I think, therefore I am>." We might now add <I disquote, therefore I am not a brain in a vat>. It is impossible to *affirm* falsely: "I am awake" and it is impossible to *utter* falsely "I am not a brain in a vat." If I were dreaming, I would not have that belief, only an *imagining* with that content, a quasi-belief, perhaps. And if I were a BIV, I would not have that thought, only a quasi-thought, perhaps.

Small comfort, though. I can be in a state subjectively indistinguishable from one in which I judge that I am awake, even though I am asleep and so not judging anything at all. So we do not really affirm:

> But we do engage in another activity that is in some ways similar to affirmation: we come to imagine. I shall call this activity 'quasi-affirmation'. Quasi-affirmation is not affirmation, but it is in many ways similar to affirmation, just as imagination is in many ways similar to belief. From an internal point of view, for the dreamer quasi-affirmation is importantly like affirmation and indistinguishable from it (Ichikawa 2008, 523).

Instead of worrying that my belief is false, I now have to worry whether my mental state is a belief and not a quasi-belief. I avoid the risk of believing falsely, but I do not avoid epistemic risks in general. The risk of quasi-belief, internally indistinguishable from real belief, is equally severe. And so, it seems to me, is the risk of "quasi-thought" expressed in vat-English internally indistinguishable from real thought expressed in normal English. Early Brueckner was aware of this deficiencies (1986, 164):

> If I do not know whether I am speaking vat-English or English, then I do not know which proposition my utterance of 'I am a BIV' expresses … . I cannot apply [disquotation] to my own utterances of 'I am [in a vat]' as a step toward the conclusion that I know that I am not [in a vat].

He later claimed that we have *apriori* semantic knowledge of our own

language which is enough to justify the disquotation. How could that be if we cannot tell the difference between thought and quasi-thought, when the principle of disquotation fails even for some of the terms in our own language? Analogously, Ichikawa (2008, 253) complains:

> It does not follow from the fact that I know no affirmation of *p* will be a mistake that it is rational for me to affirm *p*. If, for all I know, the mental act I am to engage in will be a false quasi-affirming, then knowledge that I shall never affirm falsely is insufficient.

There is no threat of our having false perceptual beliefs when dreaming but this is insufficient as a reply to dream skeptic in the same way as my knowledge that I shall never think falsely that I am a BIV seems to be insufficient as a reply to BIV skeptic.

But can we really have any sensible doubts about disquotation in our own language, e.g.: (2) "My word 'brain' refers to brains"? Several authors have argued that I cannot even *entertain* the skeptical hypothesis unless I have thoughts about BIVs and unless I can refer to brains and vats.[7] Here is Button (2013, 125-126):

> … even to understand or talk about the BIV scenario at all, we need to rely on disquotation. Otherwise, the BIV scenario does not confront us with the worry that we are brains in vats. In short, premise (2) is required by the BIV sceptic herself. ... the falsity of (2) is genuinely unrepresentable.

The issue is complex, but let me notice, first, that the target of skepticism will now really be *different*. Usually we are confronted with the skeptical possibility ("Bad") of having massively false beliefs. In Bad (dreams, BIV) our beliefs are *insensitive*. I do not know that I am not a BIV, for if I were, I would falsely believe not to be. According to Putnam and semantic externalism about thought content I know that "I am not a BIV" since the opposite belief, expressed as "I am a BIV," (as uttered by a BIV) fails to meet necessary condition for being a real belief at all, its content is not representational (alternatively, the sentence does not disquote). The new skeptical challenge will now be more *general*, for instance:

> If I accept the argument, I must conclude that a brain in a vat can't think truly that it is a brain in a vat, even though others can think this about it. What follows? Only that I cannot express my skepticism by saying "Perhaps I am a brain in a vat." Instead I must say "Perhaps I can't even think the truth about what I am, because I lack the necessary concepts and my circumstances make it impossible for me to acquire them!" If this doesn't qualify as skepticism, I don't know what does (Nagel 1986, 73).

There are, for instance, contexts in which disquotation fails, but, for all

---

[7] Thanks to anonymous referee for drawing my attention to this problem.

we know, we might be in such a situation. Dream scenarios on *both* models are an example – hallucination was mentioned above (Williamson 194, 196), and some influential theoreticians of imagination (e.g. Walton 1990) argue that names in the context of imagination do not refer.

The skeptic will thus draw attention to intelligible cases which are incompatible with the disquotation principles and unproblematic semantic knowledge. The anti-skeptic might reply: *represent* them (describe them in meaningful linguistic terms)! But something peculiar is going on in this dialectics. The *riposto* of the anti-skeptic reminds of Berkeley's Master argument, sometimes simplistically put as: show me an unobserved tree! Now, Berkeley's argument seems to conflate the represent*ation* (what we conceive with) and the represent*ed* (what we conceive of—the content of our thought).[8] The fact that we cannot *really* entertain (represent) the sceptical scenario does not entail that the scenario represent*ed* is impossible. A similar point is sometimes made in terms of the first person / third person distinction – a brain in a vat can't think truly (of itself) that it is a brain in a vat, even though others can think this about it:

> The hypothesis that I'm a brain-in-a-vat is unthinkable (thinking it requires the use of symbols tokenings of which are causally linked to actual items in ways in which no tokenings by the brains-in-a-vat in his scenario can be) (Wright 1992, 86).

We might interpret this with a help of a comparison – I can never truly (and loudly) say "I am silent now", but the others can report about the fact of my silence. But the situation with the BIV scenario is different, it seems to me. In the same way as there are genuine similarities between the dream state and a waking belief there might be genuine similarities between states with representational (referential) and nonrepresentational (non-referential) mental contents. In the vat I cannot think "I am a brain in a vat" since I cannot think about real world brains and real world vats. But, as Folina (2016, 172) rightly notices, it does not follow that I cannot have thoughts that are epistemically identical to the BIV thought (or nearly so). Or, to be more cautious, quasi-thoughts.

Consider the analogous question – can we, in our dreams, entertain the hypothesis of dream scepticism? In a sense, *no* – the preconditions of our having any beliefs are not fulfilled (according to the imagination model). Still, if there are genuine similarities between the dream state and a waking belief we might be engaged in "quasi-thinking". Now the objection can be made that the BIV scenario is significantly different since "even to understand the BIV scenario at all, we need to rely on disquotation." But BIVs cannot be like ants, for instance, they have to be

---

[8] Downing, L. "George Berkeley", *The Stanford Encyclopedia of Philosophy* (Spring 2013 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2013/entries/berkeley/>.

relevantly similar to us – capable of engaging in cognitive mental activities. And one cannot just stipulate that engaging in these activities presupposes disquotation – we can imagine or hallucinate that "tree" refers to tree. Perhaps we just "quasi-understand" and "quasi-entertain" and not really understand or entertain in these scenarios, where quasi-understanding is similar to understanding in the same way that imagination is similar to belief.

5. Conclusion

Both, the full anti-skeptical argument which includes deception and the stripped down version, which rests only upon the claim that the referents and contents in the BIV scenario differ from normal referents and contents, are question-begging. This has been noticed before (Brueckner 1986 and many other commentators). How do I know that I speak English and not Vat-English? Do I have thoughts or quasi-thoughts? How does my negation behave? Do my terms disquote? Until we establish answers to these questions the core version of the semantical argument fails as a fully satisfactory reply to the skeptical challenge posed by the BIV scenario. But wait – what *kind* of skepticism is this? Ichikawa, in his discussion of dream scepticism (2016, 159), is aware of this question:

> The central question becomes, what is required of skeptical scenario? It is clearly not enough that the belief in question is false; the imagination, dreaming, and hallucination subject must be in some sense similar to the subject's actual state. But whether belief is necessary, or sufficient, or neither is an outstanding question at this stage in inquiry.

The sceptic will add the BIV subject to the list and argue, against Putnam, that the doubts raised above are enough to cool down our epistemic hyper-ambitions. True, the skeptical challenge has now become more general[9] and consequently more radical. Putnam (1994, 284) offers the following protection against certain radical types of skepticism:

> One sort of skeptic — a very uninteresting sort — may raise a skeptical doubt only so that, no matter what premises one may rely on in answering the doubt, he or she can respond, "and how do you know that?" Obviously, this sort of skepticism — call it infinitely regressive skepticism is "unanswerable," but equally obviously the existence of infinitely regressive skepticism shows only that justification must end somewhere. My argument was obviously not meant to refute infinitely regressive skepticism.

---

[9] Or less specific, »I might not be able to say, utter, know, or imagine the precise conditions that yield my situation of ignorance,« (Folina 2016, 172).

"Infinitely regressive skepticism" – is this not the very essence of skepticism since the times of Phyrro? Maybe not very attractive nor particularly worthy of serious consideration, but this alone does not make it refutable. As an antidote for skepticism Putnam's reasoning remains unconvincing (most of the vast literature has been critical). Still, the argument remains intriguing and I have far from explored all of the challenges it poses. As a philosopher I am also tempted to use a reflection on our use of language (our concepts, ideas …) and thereby acquire some concrete information about the (external) world. But how can that be?

REFERENCES

Brueckner, A. 2016a. Putnam on Brains in a Vat. In Goldberg (ed.), 19-26.

_____. 2016b. Skepticism and Content Externalism. *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition). Edited by Edward N. Zalta. https://plato.stanford.edu/archives/win2016/ entries/skepticism-content-externalism

_____. 2010. *Essays on Skepticism*. Oxford: Oxford University Press.

_____. 1986. Brains in a Vat. *The Journal of Philosophy* 83: 148-167. Repr. in Brueckner 2010, 115-132.

Button, T. 2013. *The Limits of Realism*. Oxford. Oxford University Press.

Christensen, D. 1993. Skeptical Problems, Semantical Solutions. *Philosophy and Phenomenological Research* 53: 301-321.

Clark, P. and B. Hale, eds. 1994. *Reading Putnam*. Oxford: Blackwell.

Descartes, R. 2008. *Meditations on First Philosophy*. Translated by Moriarty, M. New York: Oxford University Press.

DeRose, K. 2000. How Can We Know that We're Not Brains in Vats?, *The Southern Journal of Philosophy* 38, Spindel Conference Supplement: 121-148.

Folina, J. 2016. Realism, Skepticism, and the Brain in a Vat. In Goldberg (ed.), 155-173.

Goldberg, S. C., ed. 2016. *The Brain in a Vat*. Cambridge: Cambridge University Press.

Hamilton, A. 2014. *Wittgenstein and On Certainty*. London and New York: Routledge.

Ichikawa, J. 2008. Scepticism and the imagination model of dreaming. *The Philosophical Quarterly* 58(232): 519–527.

_____. 2016. Imagination, Dreaming, and Hallucination. In Kind (ed.), 149-162.

_____. 2009. Dreaming and Imagination. *Mind & Language* 24(1): 103-121.

Kind, A., ed. 2016. *The Routledge Handbook of Philosophy of Imagination.* London and New York: Routledge.

Nagel, T. 1986. *The View From Nowhere*. New York: Oxford University Press.

Putnam, H. 1994. Comments and Replies. In Clark and Hale (eds.), 242-295.

_____. 1992. Reply to David Anderson. *Philosophical Topics* 20 (1): 347-408.

_____. 1981. *Reason, Truth and History*. New York: Cambridge University Press.

Sosa, E. 2007. *A Virtue Epistemology: Apt Belief and Reflective Knowledge*. Oxford: Oxford University Press.

van Fraassen, B. C. 1966. Singular Terms, Truth-Value Gaps, and Free Logic. *Journal of Philosophy* 63: 481–95.

Walton, K.L. 1990. *Mimesis as Make-Believe*. Cambridge, MA: Harvard University Press.

Williamson, T. 1994. *Vagueness*. London and New York: Routledge.

Wittgenstein, L. 1969. *On Certainty*. Eds. G. E. M. Anscombe and G. H. von Wright. Oxford: Basil Blackwell [OC]

Wright, C. 1992. On Putnam's Proof that We Are Not Brains in a Vat. *Proceedings of the Aristotelian Society* 92: 67-94.

IN DEFENSE OF THE TWIN EARTH–THE STAR WARS CONTINUE*

NENAD MIŠČEVIĆ

University of Maribor, Slovenia, Central European University, Hungary

ABSTRACT

The paper discusses the meta-philosophy of thought-experiments, in particular its neglected diachronic aspect, focusing on Putnam's work as the paradigm case, and on the trail(s) that developed out of the Twin Earth thought-experiment. Putnam's experiment is discussed from a perspective that combines metaphilosophy and actual history of analytic philosophy. Peter Unger has branded the whole debate around it as empty and fruitless. A meta-philosophical analysis shows him to be wrong. The experiment as originally proposed already appeals to a broad variation of examples and intuitive induction; the variation continues in other works addressing the issue, and produces interesting results. The second aspect is the search for reflective equilibrium, lasting till the present day. The internal logic of these processes is discussed, in order to show that the accusation for emptiness turns against Unger himself. In general, debates around thought-experiments, the already famous and also around new ones, make a large part of contemporary analytic philosophy. The way to understand a large part of this, and of debates surrounding it, is to link it to the internal understanding of a typical thought-experiment; stages of a particular experiment get discussed, developed and changed in the history of a particular trail produced by it. This is an important way in which a philosophical tradition is born, and we need to combine synchronic and diachronic approaches in order to understand it.

Keywords: *thought-experiments, Twin Earth, meta-philosophy*

1. Introduction

Putnam is one of the most successful thought-experimenters in twentieth-century philosophy; he has produced two great thought experiments, Twin Earth one and the Brain-in-a-vat one, and many related smaller ones.The two great ones have marked analytic philosophy of our time, from philosophy of language through philosophy of mind to philosophy of science and general metaphysics. Here we shall look at his Twin Earth thought experiment (Twin Earth TE for short), anticipated by Putnam in his 1970 paper „Is semantics possible?" and formulated in 1975 in „The Meaning of 'Meaning'", and at a recent criticism of it, due to Peter Unger (2014). We shall briefly defend Putnam from the criticism, and concentrate on Unger's complaint about endless "spates of papers" generated by TEs in recent analytic philosophy. We shall argue that the prolonged discussions of fundamental thought experiments (TEs) TEs in various (sub-) disciplines of philosophy define big chunks of the history of analytic philosophy (not to mention three centuries of debates over Social contract TEs and Evil Demon TE). The meta-philosophy and epistemology of TEs has to take this phenomenon into account, thus connecting history of philosophy with meta-philosophical theorizing. We shall argue for this suggestion in connection with Twin Earth TE, and the ensuing debate.

Let me just mention that recently the publication of such discussions of TE is becoming systematic. Andrew Pessin and Sanford Goldberg have collected important papers on Twin Earth TE in their *The Twin earth chronicles* from 1996 (see References). Two years ago Cambridge University Press has started a series, Classic Philosophical Arguments, pretty much concentrated on TEs, like *The Brain in a vat* (Goldberg 2016), *The Prisoner's dilemma* (Peterson 2015) and *The Original position* (Hinton 2015). The meta-philosophy (or epistemology) of TEs has to join on the theoretical side, connecting history of philosophy with meta-philosophical theorizing.

Here is then the preview. Section II focuses on Twin Earth thought experiment, starting with a short reminder, and continuing with a brief proposal of how to distinguish stages in thought-experimenting, in particular within the Twin Earth example(s). Section III asses to Unger's criticism (from his *Empty Ideas,* (Unger 2014)) and then turns to the defense of Twin Earth TE. Section IV continues the discussion with Unger by turning to the trails-traditions that have sprung out Twin Earth TE, and connecting the trails to various aspects of the experiment itself. Thus the short-term and long-term patterns of Putnam's thought-experimental reasoning and its continuation are brought together. Section VI attempts to generalize the morals of the defense beyond Twin Earth cases, first mentioning Putnam's other famous TE, the Brain-in-the-Vat, and then moving very briefly to other examples central in the history of

philosophy. It then summarizes our answer to one of the central questions of the paper: why go historical in the meta-philosophy of TEs, rather than stay with the usual style of topical discussion?

2. Reminder: The Twin Earth TE and its stages

Let me start with a general characterization of thought experiments. So, what is a TE? It is an investigative procedure „in the armchair", which normally involves

0. the formulation of experimental design pointing to

   i) the determination of the goal(s),in particular the thesis/ theory to be tested, and

   ii) the construction of a (typically) counterfactual scenario to be considered

1. the presentation of the scenario thus constructed to the experimental subject (either the author of the scenario herself, or an interlocutor),[1]

2. understanding done by the experimental subject

3. the (typically imaginative) contemplation of the scenario and some piece of reasoning,

4. the decision („intuition") concerning the thesis/theory to be tested, and then the variations and generalizations from the result.

Once this result is achieved, it can be, and often is compared with results of other thought experiments in the vicinity. We shall see that the characterization captures the TE we are interested in.

In his "The Meaning of 'Meaning'" Putnam famously introduced his Twin Earth TE as a „science-fiction example" (!):

For the purpose of the following science-fiction examples, we shall suppose that somewhere in the galaxy there is a planet we shall call 'Twin Earth'. Twin Earth is much like Earth; in fact, people on Twin Earth even speak English. In fact, apart from the differences we shall specify in our science-fiction examples, the reader may suppose that Twin Earth is exactly like Earth.

Twin Earth is very much like Earth; in fact, people on Twin Earth even speak English. In fact, apart from the differences we shall specify in our science-fiction examples, the reader may suppose that Twin Earth is exactly like Earth.

---

[1] In a philosophy seminar the design of the scenario is usually given through a text formulated in a paper. This is what is colloquially then called „TE".

He may even suppose that he has a Doppelgänger - an identical copy – on Twin Earth, if he wishes, although my stories will not depend on this. (Putnam 1975, 223)

The next move can be described as separating chemical composition of the liquid known as "water" from its stereotype.

I shall suppose that the oceans and lakes and seas of Twin Earth contain XYZ and *not water*, that it rains XYZ on Twin Earth and not water, etc. (Ibid.)

Let me note that I have asked Putnam in conversation whether we can use Kant's idea of "separating" in "experiments with our mind" of elements that normally go together to characterize what he is doing in Twin Earth scenario, and he very emphatically agreed.

Next comes the question for the reader: when a Twin Earthling uses the word "water" does he refer to water? So, after the stage of presentation and question raising, we presumably have reader's understanding. The reader is expected to imagine the situation, to build a model, so to speak. The next comes the reader's intuitive answer. The expected answer is negative: the Twin Earthling refers to what is in fact XYZ, and it is not water!

Let me mention some obvious problems with the performance of the TE at these early stages. First, consider the construction of a (typically) counterfactual scenario to be considered, and its presentation at the stage one. The proponent, say Putnam, is testing people's intuitions, and hopefully, the non-specialists will be included, since he wants ordinary intuitions from non-experts (as well as from his colleagues, the experts). So, the story has to be *relatively short and not too technical*; the best would be if it were not technical at all. This is the first sore point: any expert will be able to come with some counterexample to the short story. Note the parallel with "real" experiments: typically a single experiment concerns just one variable, and taken in isolation proves very little. What is being done is rich controlled variation; real laboratory is in this respect very similar to the laboratory of the mind.

One feature of experiments closely connected to learning is their variability: scientists perform them over and over again with modifications which may be systematic and intentional or intuitive and exploratory. Variation is one of the key factors in the success of experiment. Frequency of variation in Faraday's experiments is apparent from his laboratory notebooks: these records suggest a lot of unrecorded and unpremeditated variation. (James 1989, 67)

The construction and the presentation culminate with a question. We might assume that the main question in the original version of the TE concerns reference: does the word "water" in the Twin Earth language refer to water or not?

Next, the understanding. It is happening within the conversation, or at least the reading. Strong pragmatic considerations might intervene. First, the motivational ones. The recipient might wonder why she is being asked a particular question, what are the expectations, and so on. Next comes the understanding of the content received; no non-technical, ordinary piece of discourse is free of potential ambiguity, and we know these days that ambiguities often do show their teeth.

Unfortunately, philosophers sometimes suggest more than is needed for mere understanding. Putnam does it. And here is the simple illustration, in the passage we just quoted. "I shall suppose that the oceans and lakes and seas of Twin Earth contain XYZ and not water…", he writes at the very beginning of his presentation of Twin Earth. But does this not assume that XYZ is not water? Does this not dictate the intuition that the reader is suppose to contribute himself? What if the reader has the opposite feeling: if it looks like water, tastes like water, and so on, than it is water; I don't care for chemistry! Too bad, the TE was supposed to test intuitions, not to suggest them!

I shall assume that the recipient next builds her model of the situation; let it be the stage Three of the initial TE. This stage is not controllable by the proponent any more, and the model-building depends a lot on specific skills, memories and interest of the recipient. But Putnam does suggest the answer, although he should not be doing this. Let me assume that the recipient goes through some unconscious reasoning; we shall not need it here. So much about the problems.

The stage Four is then the rise of intuition concerning the thesis/theory to be tested.: "It seems to me, very obviously, that the Twin Earthling does not refer to water".

Normally, the recipient is expected to generalize: is what is valid for water also valid for gold, silver, and so on? The process has been called 'intuitive induction' by Roderick Chisholm (1977, *Theory of Knowledge,* ch. 4, section ' Intuitive induction'), who borrowed the term from W.D. Ross. Anyway, the intuitive induction ends with a general judgment; in this case "what is valid for water should be valid for gold, silver, and so on." I would add some obvious and immediate consequences, for instance: it is the deep structure of the matter that ultimately determines reference for typical mass terms.

Again, this stage, our number Five, is rife with problems. Assume that the intuition is valid for gold; what about "wine" or "brandy"? Is it really the chemical composition that counts, rather than drinkability, taste and the effects on the consumer?

What normally follows is theorizing. The new generalization (and its immediate consequences) should confront the rest of our theoretical commitments, in this case ranging from the ones in chemistry to the ones in semantics. And it should explain the intuitions discovered in the course

of TEs. Let me borrow the famous term from Rawls, "reflective equilibrium" and call this, sixth stage, the search for reflective equilibrium. Putnam did find one, and a quite radical one, for that matter, namely a restructuring of the whole of semantics. This is the final stage, number seven, the reflective equilibrium itself. We shall return to it later.

So much about the first (version of) Twin Earth TE. We noted the existence of challenging points at each stage. If you are pessimistic, you can call them "sensitive points" of a TE, if you are optimistic, you can describe them as "productive points", since they obviously invite questions and discussion.

Why "productive"? Consider how the experimenting proceeds further, in Putnam's text. After coming to the conclusion, one might ask a slightly different question, introducing history. Perhaps our intuitions concerning water are really dictated by Lavoisier's discovery of its chemical composition. We need to introduce a more dramatic variation.

So, we arrive at Version Two: this time by introducing history of chemistry. Stages One and Two are re-done: Imagine $Oscar_1$ and $Oscar_2$ living in 1750. Did each understand the term 'water' differently from the other? Again, the intuition seems to suggest that they did: the first referred to what is in fact H_2O, the second to what is in fact XYZ. So, $Oscar_1$ and $Oscar_2$ understood the term 'water' differently in 1750 although they were in the same psychological state, and although, given the state of science at the time, it would have taken their scientific communities about fifty years to discover that they understood the term 'water' differently. We perform some intuitive induction and end up with the general judgment, valid also for gold, and the like.

But what about the kinds of stuff that are both present around, perceptually indistinguishable, and strongly associated with some familiar product; or at least one of them is. This suggests further variation.

Version Three: "we will now suppose that molybdenum is as common on Twin Earth as aluminum is on Earth, and that aluminum is as rare on Twin Earth as molybdenum is on Earth. In particular, we shall assume that ' aluminum' pots and pans are made of molybdenum on Twin Earth. Finally, we shall assume that the words 'aluminum' and 'molybdenum' are switched on Twin Earth: ' aluminum' is the name of molybdenum and ' molybdenum' is the name of aluminum.". (Putnam 1975, 225)

Version Four: the final question concerns non-exotic, not rare specimens, and moves from science fiction to ordinary examples and from mass terms to kind-terms in botanics and the like; this is the elm-beech example (Putnam 1975, 226).

We arrive now to the general conclusion: meanings can vary with psychological states remaining constant explanation: meaning depends on

external causal connections.

> We claim that it is possible for two speakers to be in exactly the same psychological state (in the narrow sense), even though the extension of the term A in the idiolect of the one is different from the extension of the term A in the idiolect of the other. (Putnam 1975, 222)

How is this possible? We need explanation, and it will appeal to causal theory and division of linguistic labor. Further steps lead to the final reflective equilibrium, strongly revisionary of traditional Fregean semantics (for a historical overview see Floyd 2005, in References). Here are then the stages.

STAGES	EVENTS
Stage 0 - design	
Stage 1 - the question	The scenario presented and the question: To what does the Twin Earth speaker refer with her word "water"?
Stage 2 - understanding	The reader understand the scenario and the question.
Stage 3 - tentative conscious production: tries to imagine the arrangement and does it to one's satisfaction.	Imagining the other planet.
Stage 3a - possible non-conscious elaboration	Possible problems: what is my body made of there? Is such a scenario possible at all?
Stage 4 - Intuition: immediate spontaneous answer	*The Twin Earth speaker refers to XYZ, not to water.*
Stage 5 - Intuitive induction: varying and generalizing	Variations: First, other materials. Next: History of chemistry, Aluminum-molybdenum, Elm-beech
Stage 6 - general belief	Extensions can differ, although the speakers are in the same psychological state.
Stage 6a - explanation	"Meanings are not in the head", causal theory, division of linguistic labor
Stage 7 - search for reflective equilibrium	Discussion of description theory in semantics, criticism of it
Stage 7a - reflective equilibrium achieved	Keep the general belief and the explanation, and restructure semantics.

This concludes the short reminder of the original TE(s). Very soon after the original presentation, several philosophers, and Tyler Burge in particular in 1979 paper "Individualism and the mental", famously noted that the story about linguistic meaning can be extended to mental content as well. Burge developed it into an externalist theory of mental content. Again, we have a variation, but rather dramatic one, primarily at the level of immediate conclusions from Twin Earth and similar TEs, and perhaps even deeper, at the level of initial question. The reader is presented with variations Oscar$_1$ and Oscar$_2$ scenario, and now, in the 'laboratory of her mind she has to test a bit different assumption: is the content of Oscar$_1$'s mental state when thinking of the relevant metal the same as Oscar$_2$'s or not? Burge famously answered in the negative. Here is Putnam's comment on Burge from 1996: "…I have come to believe that he is right" and "…I agree with his paper" (Goldberg 1996, xxi).

Once the new question was around, a new round of reactions followed, represented by ten famous paper reproduced in (Goldberg 1996) mostly from the eighties, when the action was at its peak. Searle (1983) insisted that the difference between contents is internal, due to the nature of intentionality, and reflexive functioning of the indexical(s) allegedly involved ("this metal"). Fodor (1991) proposed two kinds of content, and argued for the primacy of the internal, "narrow", one. Dennett came up with the idea of a "notional content" (see his contribution in Goldberg 1996) and Jackson (in Goldberg 1996) proposed a stable, compromise solution, that later developed into "two-dimensionalism". The difference in the question asked, or at least the difference in the weight assigned to the mental in contrast to the linguistic, resulted in differences in intuitions, then in generalizations arrived at, and then projected deep into the area of explanation and reflective equilibrium (see Goldberg 1996) for a nice collection of historically crucial papers). So, hopefully, the attention to stages might help us understand the further history of the Twin Earth TE.

But why care about such matters? Why go historical in the meta-philosophy of TEs, rather than stay with the usual style of topical discussion? We shall say more about the process after we take a look at the recent criticism of Putnam and of methodology of analytic philosophy in general.

3. Empty thoughts: Peter Unger's critism of Putnam and analytic methodology

Let me start by quoting Peter Unger's interview in which he nicely summarizes his view of contemporary philosophy:

> Philosophers easily get the idea that somehow or other, just by considering things about the world that they already know, they can write up deep stories which are true, or pretty nearly true, about how it is with the world. By that I especially mean the world of things that includes themselves, and everything that's spatio-temporally related to them, or anything that has a causal effect on anything else, and so on. They think they can tell a deep story about how it is that all of this stuff really hangs together, that's much deeper, more enlightening and more comprehensive than anything that any scientist can do.

> And so philosophers proceed to write up these stories, and they're under the impression that they're saying something new and interesting about how it is about the world, when in fact this is all an illusion. To say new and interesting things about the world — and that's very hard, things of any generality I mean, or even anything interesting — you really have to engage with a lot of science. And very few philosophers do any of that, at least in any relevant way. (Boey and Unger 2014)

All this, Unger thinks, applies to Putnam, to whom several sections, and the whole chapter three of his book *Empty Ideas: A Critique of Analytic Philosophy*, titled Earth, Twin Earth and History, are dedicated.

His first criticism is simple: there is too much material, just spates of articles. He said he entered "twin earth", coming up with 1,941 items, comprising 1,571 articles and 261 reviews, and the remainder in other categories. Is it good or bad? Depends on the quality of the articles. And here comes his second criticism: "...a fair lot of much of this remarkably empty literature" (Ibid.)

And Unger comes, surprisingly, with an alternative proposal, a new Twin Earth thought experiment. Here is just the barest sketch. He invites us to imagine someone like Lavoisier, „on the verge of discovering that earthly water is composed of hydrogen and oxygen." At the same time, the twin Lavoisier is on the verge of discovering the chemical composition of twin water. „In all their real mental powers, our two Joseph Antoines will be precisely similar, however we sensibly construe "mental powers". not only as regards what behavior each is apt to produce (in any encountered environment) but also as concerns what experiencing each is apt to enjoy, each of our two chemists is, at our start, precisely similar to the other." (2014: 64). "…/N/ow, suppose each ... of them „ to switch places with the other, almost instantaneously. with this switch, twinchemist will

be here on earth, in 1750". Of course, each proceeds with his discovery, only that it is now the twinchemist who discovers the composition of water. So, what is the big deal, Unger asks. "The capacity to „think of water" seems not to be needed for the job." Having beliefs about water appears irrelevant to being able to discover water's composition. Our belief-states have full general propensities, they react in the same way to water and twin water, Unger claims. Therefore, Putnam's line is irrelevant!!! Unger than offers a very complicated and sophisticated TE with elementary particles, to introduce what he sees as the relevant alternative scenario. We shall not reproduce it here.

Let us now pass to a brief defense of Twin Earth. First, Unger's own methodology suggests the indispensability of TEs: even if you see analytic philosophy as a bunch of empty ideas, you will have to propose TEs of your own. This is a deep inconsistency in what Unger is doing. Alternatively, if he had looked at alternatives outside of analytic philosophy and had turned to recent continental authors, he would, with his criteria of clarity, probably see it as passing from contradiction to contradiction; not a very attractive option.[2]

Unger sees the biggest part of analytic philosophy as being empty of concrete content. and proposes an alternative relying on extremely sophisticated scientific knowledge as the background for analytic philosopher. Emptiness is for him tied to the endless production of tons of empty material. So, let me briefly address his objection, and then concentrate upon the first one.

---

[2] Think of famous continental thinkers, writing in the second half of 20th century. Heidegger suggests that philosophy should become poetic. Here is a typical quote on human condition:

> But what is it that touches us directly out of the widest orbit? What is it that remains blocked off , withdrawn from us by ourselves in our ordinary willing to objectify the world? It is the other draft: Death. Death is what touches mortals in their nature, and so sets them on their way to the other side of life, and so into the whole of the pure draft. Death thus gathers into the whole of what is already posited, into the positum of the whole draft. As this gathering of positing, death is the laying-down, the Law, just as the mountain chain is the gathering of the mountains into the whole of its cabin. (Heidegger 1971, 123).

Somebody like Unger, who finds Putnam not rigorous enough, would be probably shocked by the very idea that this claim of Heidegger is a piece of philosophy. Similarly with Derrida. Here is how Derrida expresses his view that a non-figurative treatment of metaphor is impossible:

> I am obliged to speak of [metaphor] *more metaphorico*, to it in its own manner. I cannot treat it (entraiter) without dealing with it (sans traiter avec elle) ... I do not succeed in producing a treatise (une traite) on metaphor which is not treated with (traite avec) metaphor which suddenly appears intractable (intraitable). (Derrida 1998, 102–3).

For more examples and a systematic discussion see my https://www.academia.edu/1828410/PHILOSOPHIZING_WITHOUT_ARGUMENT

The sources of potential philosophical knowledge are not mysterious as Unger sees them. Remember: "Philosophers easily get the idea that somehow or other, just by considering things about the world that they already know, they can write up deep stories which are true, or pretty nearly true, about how it is with the world." This simply doesn't fit the Twin Earth TE**.** The semantic part is about us. We might be unaware of the structure of our semantic habits – the reflection on cases can make us aware of it. As far as the non-human world is concerned, the info in the theory of Twin Earth TE comes from science, chemistry etc. and is then brought together with our understanding of meanings.

This is, it seems to me, quite a general pattern with famous TEs in descriptive-explanatory theoretical philosophy. For instance, Jackson's Black and White Mary makes sense only on the background of quite developed neurological picture of human mind, and questions one central ambition of it. The same with Putnam's Brain-in-a-vat TE. One element that made it interesting was his replacing of the Cartesian thinker, ego, or whatever, with brain, and his connecting this to the up-to-date debates concerning Turing machines and Skolem-Lowenheim theorem (in chapter one of *Reason, Truth and History;* see the paper by Danilo Šuster in this volume). I think Tim Williamson is right about the abuse of the epithet "empty" by Unger:

> Unger's use of the term "empty" is just an advertising trick. It's like a competitor who defines "empty" as "containing nothing but brand X fruit juice" and then puts up posters warning that cartons of brand X fruit juice are empty. To read Empty Ideas, one must get through the equivalent of numerous elaborate descriptions of cartons of brand X fruit juice of various types, each concluding that the carton was empty, and for contrast some elaborate descriptions of cartons of brand Y fruit juice of various other types, each concluding that the carton was full. The reader's task is made no easier by Unger's loquacious, attention-seeking prose. (Williamson 2015, 22-23).

We should then turn to meta-philosophy and history of philosophy. What about Unger's first criticism? Does the spate of articles make any sense? Obviously it does for Unger, for he tries to add to it another version of Twin Earth TE. At the same time, it goes on his nerves. Here is a therapy: let us try to understand where the spates are coming from.

4. Stages and traditions - from the synchronic to the diachronic

Unger is obviously right about TEs generating spates of articles, full of comments, criticisms and alternative proposals. Where do these spates come from? How does a spate begin? In order to answer the questions we have to dig deeper into the matter, and ultimately, we need some meta-philosophy explaining the recent history of analytic philosophy. First, we can note that spates come in trails. Unger himself notes the following:

> „Just a few years later and largely thanks to their contemplating Putnam's Twin Earth scenarios, several other philosophers did endorse this further idea, concerning who thinks what, the two most timely being, perhaps, Tyler Burge and Colin McGinn" (Unger 2014 ,77- 60).

But the spate of articles that come out of Putnam's and Burge's efforts constitutes a trail, the externalist one. We can even call it a tradition. There is the contrary trail-tradition, represented by more internalistically oriented authors, from Searle to Jackson.

The first thing to note is that the trails start as reactions to particular features of the original Twin Earth TE, and the second, and a surprising one, that reactions target particular stages of the TE, particular sore-fruitful spots which one can use as starting points from one's own proposals. Some authors, perhaps most famously Burge, widen the range of examples (from water to arthritis, and from true beliefs to systematically false ones), thus targeting the initial design of the TE and changing a bit the question asked, others stay with original example, but, at the stage of intuition-generation suggest a different intuition: some content of Twin Earthling's thought is identical to the content of Earthling's thought. The new intuition demands new explanation, and commands a different process of intuitive induction.

In other words, *there is a connection between the synchronic structure of stages and the diachronic process of generation of trails, even traditions*. Let us take a closer look at the externalist and internalist trails.

Start with Tyler Burge. Here is his question, stage 1: Does the ignorant person who believes that he has arthritis in his thighs refer to arthritis? Yes, the intuition suggests (at stages 4-5). The same for sofa and contract. We can see his initial proposal as a variation of the crucial example, replacing water with items like sofa, arthritis or contract. This changes the TE already in its initial stages. Now, intuitive induction, our stage 6: the contents of the thought of ignorant persons all depend on the public meaning of the terms involved. The very content of thought is determined externally; a strong externalism about content, not just about linguistic meaning, follows. Of course, a lot of work is to be done at the next stages to relate the view to other accepted ideas in philosophy of mind. Burge and his followers will be adamant about the externalist intuition, its

externalist explanation, and the general externalist stance; all else can be changed, but not these results. Some kind of reflective equilibrium follows (of course, it can be challenged by the internalist, who might point out that on this account we don't know contents of our thoughts, as Boghossian famously did-see References).

Incredibly enough, the trail generated by the externalist proposal will be crossing another trail, the Demon TE that comes from a Cartesian tradition, and the fans of the trail will try to subvert the results from the Cartesian one. Consider a brain-in-a-vat: does it have false thoughts about the external world, a Cartesian thinker will ask. Well, does it have thoughts about the external world at all? The externalist line suggests a negative answer. But, if it has no thoughts about the external world, it does not have any false thoughts about the external world; skepticism is inconsistent. No wonder, there is a recent book bringing together the papers from the crossing, Sanford C. Goldberg's 2016 collection *The Brain in a Vat*. A discussion of this crossing of the two traditions would demand a separate paper. However, mere pointing out the phenomenon of crossing is all we can do here.

To conclude, the history of semantic externalism can be reconstructed to a large extent as history of the Twin Earth TE(s) debate. Attention to the internal structure of the TE(s), especially to the sore-fruitful spots at each stage of the TE that prompt questioning and the debate make see the discussion and the 'spate of articles' produced as rational response to the crucial issue involved. The elementary reconstruction of the TE(s) in terms of stages, and of problems connected with the performance of the experiment combine with diachronic rational reconstruction of the history of externalism as a whole and make us understand the important trail of contemporary analytic semantics and theory of mental content in rational and philosophically relevant terms.

Let me now pass to the opposite, internalist trail, extremely popular, with a lot of followers. First, a moderate proposal. Searle insisted that the difference between contents is internal, due to the nature of intentionality, and reflexive functioning of the indexical(s) allegedly involved ("this metal") (Searle 1983, 206). Next, consider a more radical proposal, involving a deeper analysis of earthling vs. Twin earthling thoughts. There is a kind of mental content, narrow content, that is the same across Earth-Twin Earth contrast. Just consider causal powers of the thoughts: „being a water thinker is the same causal power as being a twater thinker, only instantiated in a person with a different causal history." (Fodor 1991, 5-26, reprinted in Pessin and Goldberg (eds.) 1996, 275 ff). His further conclusion is that the narrow content is the right content for psychology (and psychological generalizations), the view known as methodological solipsism The final reflective equilibrium reached is completely opposite from Burge's and later Putnam's, and it has had a lot of followers.

So, we may look at the longer term history of different stages within the history of a TE or its variants. It starts with initial experiment, including variation, general belief-conclusion, explanation, and early reflective equilibrium, and can for example proceed to a new question. Is the linguistic meaning determined by external factors is supplemented with a more dramatic question: isn't the mental state itself then determined by external factors? There will be a new example (arthritis in the thigh), or a new reading of the old one (or one can accept two kinds of content, like Fodor, and Jackson). The new process of 'intuitive induction' will lead to a new conclusion, say that one extended to mental content that now varies with the surrounding. Then we need a new explanation: causal theory is widened to the mental content, and the division of linguistic labor has internal psychological consequences. This is a new reflective equilibrium, different from the earlier one(s). With some luck a new trail is created, which can become a real new tradition. And this can be repeated many times, passing, for instance to analysis of ethical discourse, and proposing Moral Twin Earth TE (see T. Horgan and M. Timmons 1992). But, there is space for other kinds of variation, for instance, including information about scientific treatment of kinds (chemistry, biology, etc.), what Unger would describe as Substantial Scientiphicalism. Or, one can proceed to psychological questioning as has been done in experimental philosophy (see Haukioja 2015 for some discussion of Putnam).

What is crucial of us is the link between short-term and long-term pattern of (thought-) experimental reasoning. This has not been noticed in the literature. Let me summarize what has been said in this section in the following table:

STAGES	VARIATION AND NEW TRAILS	EXAMPLES
Stage 0 - design	Accommodating data from stages 6, 7, 8	Redesigning the TE Tyler Burge: externalist proposal Jackson: varying which world is actual Fodor: methodological solipsism
Stage 1 - the question	New question: reformulation or complete transformation	Widening: From language to the mind or From mind to ethics: moral Twin Earth

Stage 2 - understanding	(adding explanatory material)	
Stage 3 - tentative conscious production: tries to imagine the arrangement and does it to one's satisfaction. Stage 4 - possible non-conscious elaboration Stage 5 - Intuition: immediate spontaneous answer	Questioning intuition, suggesting a different judgment	Jackson: a very different intuition XYZ is water in one sense, is not water in another Fodor: We and Twin Earthlings share one content (narrow), but not the other (wide)
Stage 6 - Intuitive induction: varying and generalizing	Conservative: Widening the range Revolutionary: Problematizing the original examples ????	Jackson and Fodor both widen the range enormously, so as to cover all Putnam's cases (and perhaps more)
Stage 7 - general belief 7a explanation	A new general belief G: Conservative extension Revision ??? A new explanation	A new general belief G*: Jackson: Two meanings Fodor: 2 kinds of content methodological solipsism
Stage 8 - search for reflective equilibrium	Traditionalist: revise G and keep the inherited beliefs revolutionary: keep G and change the inherited beliefs	
Stage 8a - reflective equilibrium achieved	The new theory: conservative revolutionary	

But why do philosophers go through all these efforts. One answer is that TEs are indispensable. Philosophy does not use laboratory to test its theories; the only experiments available here are those in thought. TEs play in philosophy the crucial role that laboratory experiments play in science. Philosophers are vitally interested in *connections between our spontaneous understanding of important items, like meaning and content of our thoughts, and the results of science*. In order to answer the question about the relation between, say, cognitive science-cum-neurology and our feeling of having contentful thought, we need to confront the two, and we cannot do it within science alone. We need the bridge, and a TE is a perfect candidate. TEs are our laboratory, and philosophers return to their experiments, as scientists do to theirs.

We now understand why one should go historical in the meta-philosophy of TEs, rather than stay exclusively with the usual style of topical discussion. The spates of articles, lamented by Unger, are trails, promising to become traditions, and mark the present-day analytic philosophy, as their ancestors marked the early twentieth century efforts.

5. Trails-traditions: towards a meta-philosophy of analytic efforts

It is time to generalize, and draw the morals from our discussion of Putnam's TEs. We obviously need a historically informed meta-philosophy of analytic efforts, in this case ones connected to thought-experimenting, and combining the awareness of the inner structure of a TE with insights into historical, diachronic process of generation of trails leading from various components of a TE to further and further developments. Some famous examples, much older than the analytic tradition, like Plato's TE of ideal state, and Descartes's Evil Demon TE present the same trail-forming tendency.

Here are the elements we were able to identify. First, the indispensability of thought-experimenting. There is often no other way to address central philosophical questions. So, there is no wonder that challenges TEs are revisited time and again

Second, performing a TE involves asking a question and then letting the subject decide; like in laboratory TEs, the designer of the experiment is not the absolute master. There are performance challenges, issuing in hot spots: the reader responds differently from the expectations of the designer. So, the designer has to go back to the drawing board, and vary the initial proposed scenario. Or, a colleague with different background opinions might find the intended answer incredible. Then she has to propose an alternative scenario that will vindicate her intuitions. This produces endless variation in scenarios and questions and production of new intuitions and immediate generalizations.

Next, the history sets in, with long term development of these generalizations, attempts to explain them, and new candidates for reflective equilibrium, from very conservative to very radical ones. Trails are being born, that, if there is luck enough, turn into solid tradition.

Note that laboratory experiments are discussed and varied in the same way as TEs. Take Edison's discovery that direct current can do marvelous things. But then comes Tesla: "Yes, boss, but alternating current can do many more, and much more marvelous ones." What followed was "War of Currents" in the 1880s over whose electrical system would power the world — Tesla's alternating-current (AC) system or Edison's rival direct-current (DC) electric power.

So with laboratory experiments we have the same pattern: first, the initial experiment, followed by small, routine variations, then, in case of really important ones, a spectacular variation. Scientists discuss the prospects of both, and propose further variations, if needed. Remember the quote from F. James who stresses the variability of experiments, and compare his diagnosis to what happened with Twin Earth TE.

Back to the diachronic developments in a long-term life of a TE. Let me make a first step towards generalizing the morals of it and offer a new proposal for the understanding of the history of debates around important TEs that make up a significant part of analytic philosophy. Let me again mention two to three famous TEs: Evil demon-Brain in a vat TE and the Social contract. These have engendered long term traditions, each lasting around three hundred years, if we count modern philosophy only, and set aside the Greeks and Saint Augustine. If we could bring together short-term and long-term stages, this would make the historical process more intelligible. For instance, take the relatively recent variations of the two mentioned thought-experimental oldies. The Evil Demon has been innovatively transformed into (or replaced with) the Brain-in-a vat (see Goldberg,2016), and Social contract has generated the Original position proposal (see Hinton, 2015) and its relatives, due to Scanlon and Habermas.

The patterns seems recognizable. The tradition starts with an initial TE, and proceeds through three possible kinds of reactions. First, a new variant of the TE or a new, but related TE is proposed. In the Social Contract tradition, one can change the characterization of the parties involved (males only, females as well, what about mentally challenged persons, and so on). The dramatic example of the latter is Brain-in-a-vat: replace the thinking person from the original Cartesian scenario with a brain wired to a control panel.

Or, there can be a non-thought experimental reaction to the initial TE: think of the numerous objections to Descartes First Meditation, collected in *Meditations with Objections and replies*, or of the "negative program" within experimental philosophy, or, concerning Original position and its

consequences, the criticisms of B. Barry (1973).

Finally, we can have both, criticisms plus new TE, the way we just saw in Unger's reaction to Putnam. Once we have the big picture, we understand the strange reaction from Unger: on the one hand criticizing the spate of papers discussing Twin Earth TE as empty, meaningless, and over-complicated, and on the other, adding to the spate, by proposing further, related, but much more complicated TEs. On the one hand, a philosopher is tempted to a negative reaction when one notices the sheer length of a famous TE trail. On the other, the only legitimate way to criticize is to add to the trail!

The history is in some respect analogous to short term reflection and discussion of such a TE, say in a seminar or a conference. Stages of understanding and debating normally structure the short term life of the TE. Longer lasting attempts (with variations, often fundamental) are part of the further, historical story. So, we can integrate the view of history of (analytic) philosophy with meta-philosophy of TEs, and then, we shall find Unger-like disappointments and criticisms less persuasive, and less worrisome. Twin water has become part of the philosophical ecology, and has a right to stay with us. It is part of our life as philosophers. In the case of science, understanding the history of science, we need the understanding of the longer history, of the laboratory experimental tradition tied to some important initial experiment. The same holds for TEs in philosophy and science: we need the understanding of the longer history of each given TE in the history of philosophy We want to know what is the relation between such historical changes, and the usual micro-variations of a given TE. Once we have a worked out the answer we shall be able to integrate the meta-philosophy of TEs with information from history of philosophy. Here, we have the first steps. And Putnam's work remains as the paradigmatic example of depth and fruitfulness of thought-experimenting in philosophy. The Star Wars continue.

REFERENCES

Barry B. 1973. *The Liberal Theory of Justice: A Critical Examination of the Principal Doctrines in A Theory of Justice by John Rawls*. London: Oxford University Press.

Boey, G and Unger, P. 2014. *Philosophy is a bunch of empty ideas: interview with Peter Unger*, Monday, June 16, available at (http://www.(.com/3quarksdaily/2014/06/philosophy-is-a-bunch-of-empty-ideas-interview-with-peter-unger.html)

Boghossian, P. 1998. What the Externalist Can Know A Priori. In Wright, Smith, and Macdonald (eds.), 271–284.

Burge, T. 1996. Other bodies. Repr. in: Pessin and Goldberg, 142-160.

Derrida, J. 1998. The Retrait of Metaphor. In Wolfreys (ed.).

Descartes, R. 2006. *Meditations, Objections, and Replies*, edited and translated by Ariew, R. and Cress, D., Indianapolis: Hackett Publishing Company, Inc.

Floyd, J. 2005. Putnam's "The Meaning of 'Meaning': Externalism in Historical Context" in *Hilary Putnam*, ed. Y. Ben-Menahem, 17-52. Cambridge: Cambridge University Press.

Fodor, J. A. 1991. A modal argument for narrow content, *The Journal of Philosophy*, Vol. 88, No. 1 (Jan.), 5-26.

Goldberg, S. C., ed. 2016. *The Brain in a Vat*. Cambridge: Cambridge University Press.

Haukioja, J., ed. 2015. *Advances in Experimental Philosophy of Language*. London: Bloomsbury Publishing.

Heidegger, M. 1971. *Poetry, language, thought*. Harper & Row.

Hinton, T., ed. 2015. *The Original Position*. Cambridge: Cambridge University Press.

Jackson, F. and Pettit, Ph. 1996. Functionalism and Broad Content. Repr. Pessin and Goldberg, 219-237.

Miščević, N. *Philosophizing_Without_Argument*. Available at: https://www.academia.edu/1828410/PHILOSOPHIZING_WITHOUT_ARGUMENT

Pessin, A. and S. Goldberg, eds. 1996. *The Twin earth chronicles: twenty years of reflection on Hilary Putnam's " The meaning of'meaning"*. New York: Routledge and M. E. Sharpe.

Peterson, M., ed. 2015. *The Prisoner's Dilemma*. Cambridge: Cambridge University Press.

Putnam, H. 1975. The Meaning of 'Meaning'. First published in K. Gunderson (ed.), *Language, Mind and Knowledge*, Minnesota, Studies in the Philosophy of Science, University of Minnesota Press. Repr. in *Mind, Language and Reality, Philosophical Papers*, Volume 2, Cambridge: Cambridge University Press.

_____. 1981. *Reason, truth and history*. Cambridge: Cambridge University Press.

Unger, P. 2014. *Empty Ideas: A Critique of Analytic Philosophy*. Oxford: Oxford University Press.

Searle, J. 1983. *Intentionality.* Cambridge: Cambridge University Press.

Williamson, T. 2015. Interview in *The Times Literary Supplement*, January 16, 22-23.

Wolfreys, J., ed. 1998. *The Derrida Reader: Writing Performances*. Edinburgh University Press.

Wright C, B. Smith and M. Macdonald, eds. 1998. *Knowing Our Own Minds*, Oxford: Clarendon Press.

PUTNAM'S CONCEPTION OF TRUTH

PUTNAMOVA KONCEPCIJA ISTINE

MASSIMO DELL'UTRI
University of Sassari, Italy

ABSTRACT

After stressing how the attempt to provide a plausible account of the connection between language and the world was one of Putnam's constant preoccupations, this article describes the four stages his thinking about the concepts of truth and reality went through. Particular attention is paid to the kinds of problems that made him abandon each stage to enter the next. The analysis highlights how all the stages but one express a general non-epistemic stance towards truth and reality—the right stance, according to Putnam, in order to develop full-blooded realism. Since the last stage combines a version of direct realism with a pluralist conception of truth, the article proceeds by focusing on Putnam's alethic pluralism, carefully distinguishing it from alethic deflationism. Finally a suggestion is made as to where Putnam's alethic pluralism may be placed within the constellation of current pluralist positions about truth.

Keywords: Truth, alethic pluralism, alethic deflationism, realism, Hilary Putnam

SAŽETAK

Nakon što naglašava kako je pokušaj iznošenja plauzibilne teorije o povezanosti između jezika i svijeta bio jedna od Putnamovih stalnih preokupacija, ovaj članak opisuje četiri faze koje je njegovo razmišljanje o konceptima istine i stvarnosti prošlo. Posebna pažnja usmjerena je na vrste problema koji su ga natjerali da napusti svaku od faza u korist iduće. Analiza naglašava kako sve faze osim jedne izražavaju općeniti neepistemički stav naspram istine i stvarnosti – ispravan stav, prema Putamu – kako bi se razvio punokrvni (*full-blooded*) realizam. Budući da zadnja faza kombinira verziju direktnog realizma s pluralističkom koncepcijom istine, članak nastavlja tako da se usredotočuje na Putnamov aletski (*alethic*) pluralizam, pažljivo ga razlučujući od aletskog deflacionizma. Na kraju, predloženo je gdje bi Putnamov aletski pluralizam mogao biti smješten unutar konstelacije trenutnih pluralističkih teorija istine.

Ključne riječi: istina, aletski pluralizam, aletski deflacionizam, Hilary Putnam

CONCEPTUAL RELATIVITY MEETS REALISM IN METAPHYSICS

KONCEPTUALNA RELATIVNOST SUSREĆE REALIZAM U METAFIZICI

TOMÁŠ MARVAN

Institute of Philosophy, Czech Academy of Sciences, Czech Republic

ABSTRACT

The paper adresses the relationship between ontological realism and Putnam's thesis of conceptual relativity. The paper divides into three parts. The first part aims to reconstruct the notion of conceptual relativity, focusing on Putnam's example involving mereological principles of individuation of objects. The second part points to some major shortcomings of the mereological example of conceptual relativity and then moves to a different version of conceptual relativity, which targets objects posited by mature scientific theories. I claim that the mereological and the scientific version of conceptual relativity are different in important respects and that two main types of conceptual relativity therefore need to be distinguished. In the third part, I show that conceptual relativity is not in tension with realism. More specifically, conceptual relativity is not in tension with "realism in metaphysics" that Putnam adopted in the last decade before his death.

Keywords: Hilary Putnam, conceptual relativity, realism, optional languages

SAŽETAK

Rad se odnosi na vezu između ontološkog realizma i Putnamove teze konceptualne relativnosti. Rad je podijeljen na tri dijela. Prvi dio cilja na rekonstrukciju ideje konceptualne relativnosti, usredotočujući se na Putnamov primjer koji uključuje mereološke principe individuacije objekata. Drugi dio ukazuje na neke od glavnih nedostataka mereološkog primjera konceptualne relativnosti, a zatim prelazi na drugačiju verziju konceptualne relativnosti koja je usmjerena na objekte uspostavljene zrelim znanstvenim teorijama. Tvrdim da su mereološka i znanstvena verzija konceptualne relativnosti različite u važnim aspektima i da stoga dva glavna tipa konceptualne relativnosti treba razlikovati. U trećem dijelu prikazujem da se konceptualna relativnost ne kosi s realizmom. Točnije, konceptualna relativnost se ne kosi s „realizmom u metafizici" koji Putnam preuzima u zadnjem desetljeću prije svoje smrti.

Ključne riječi: Hilary Putnam, konceptualna relativnost, realizam, neobavezni (*optional*) jezici

ARGUING ABOUT REALISM: ADJUDICATING THE PUTNAM-DEVITT DISPUTE

RASPRAVLJANJE O REALIZMU: SUĐENJE U DEBATI PUTNAMA I DEVITTA

JADE FLETCHER

University of Leeds, United Kingdom

ABSTRACT

In this paper I want to adjudicate the dispute between those philosophers who do and those who do not think that the philosophy of language can illuminate metaphysical questions. To this end, I take the debate between Devitt and Putnam as a case study and diagnose what I take to be illuminating about their disagreement over metaphysical realism. I argue that both Putnam and Devitt are incorrect in their assessment of the significance of the model theoretic argument for realism. That, whilst Devitt is entitled to claim that truth does not have anything to do with realism, Putnam's challenge can still gain traction and seriously call into question our ability to engage in realist metaphysics. I argue that even if a completely semantically neutral conception of realism can be successfully articulated, doing so has the potential to bankrupt the methodology of metaphysical realism. Having taken this debate as a case study, I then offer some brief remarks on how to understand the philosopher who claims that realist metaphysicians should care about discussions of metasemantics and truth. Whilst I want to be cautious about generalising on the basis on this case alone, I think there are important lessons to be learned about the way in which considerations to do with language can shed light on the concerns of metametaphysics.

Keywords: Metaphysics, Metametaphysics, Reference, Indeterminacy, Truth, Hilary Putnam, Michael Devitt

SAŽETAK

U ovom radu želim rasuđivati o raspravama između onih filozofa koji misle i onih koji ne misle da filozofija jezika može rasvijetliti metafizička pitanja. S ovim ciljem na umu, uzela sam debatu između Devitta i Putnama za predmet istraživanja i dijagnosticirala ono što smatram prosvijetljujućim u njihovom neslaganju oko metafizičkog realizma.

Argumentiram da su obojica, Puntam i Devitt, nekorektni u svojoj procjeni važnosti teorijskog modela argumenta za realizam. Odnosno, dok Devitt ima pravo tvrditi da istina nema nikakve veze s realizmom, Putnamov izazov može još uvijek postići popularnost i ozbiljno dovesti u pitanje našu sposobnost da se bavimo realističkom metafizikom. Branim tezu da čak i ako potpuno semantički neutralna koncepcija realizma može biti uspješno izražena, izražavanje iste polučuje potencijal da se metodologija metafizičkog realizma potkopa. Nakon što sam uzela ovu debatu kao predmet istraživanja, nudim nekoliko kratkih opaski o tome kako razumjeti filozofa koji tvrdi da bi realistički metafizičari trebali usmjeriti pažnju na diskusije o metasemantici i istini. Iako želim biti oprezna u vezi s generaliziranjem na temelju samo ovog slučaja, smatram da postoje važne lekcije koje se mogu naučiti o načinu na koji razmatranja što imaju veze s jezikom mogu bacati novo svjetlo na pitanja metametafizike.

Ključne riječi: metafizika, metametafizika, značenje, neodređenost, istina, Hilary Putnam, Michael Devitt

DEFERENCE AND STEREOTYPES

PRIKLANJANJE I STEREOTIPI

ANDREI MOLDOVAN
University of Salamanca, Spain

ABSTRACT

In this paper I discuss Hilary Putnam's view of the conditions that need to be fulfilled for a speaker to successfully defer to a linguistic community for the meaning of a word she uses. In the first part of the paper I defend Putnam's claim that knowledge of what he calls "stereotypes" is a requirement on linguistic competence. In the second part of the paper I look at two consequences that this thesis has. One of them concerns the choice between two competing formulations of consumerist semantics. The other concerns the notion of deference, and in particular the question whether deference can be non-intentional. Although the standard view is that deference is intentional, it has also been argued (Stojanovic et al. 2005) that most common forms of deference are not. I argue that deference is best understood as intentional, given the possibility of failures of deference. Cases in which the requirement that the speaker know the stereotypes associated with a particular word is not fulfilled are examples of unsuccessful attempts to defer.

Keywords: deference, Putnam, stereotypes, intention, default deference

SAŽETAK

U ovom radu raspravljam o gledištu Hilary Putnama o uvjetima koji trebaju biti ispunjeni da bi se govornik uspješno priklonio lingvističkoj zajednici u vezi sa značenjem riječi koje koristi. U prvom dijelu rada branim Putnamovu tvrdnju da je znanje o onome što on naziva „stereotipi" obavezno za lingvistički kompetenciju. U drugom dijelu rada nudim pregled dviju posljedica koje ta teza ima. Jedna od njih tiče se izbora između dvije konkurentne formulacije konzumerističke semantike. Druga se tiče ideje priklanjanja, točnije pitanja može li priklanjanje biti nenamjerno. Iako je standardni stav da priklanjanje jest namjerno, također se argumentira (Stojanovic i dr., 2005) da najčešće vrste priklanjanja to nisu. Tvrdim da se priklanjanje najbolje može razumjeti kao namjerno, s obzirom na mogućnost neuspjeha priklanjanja. Slučajevi u kojima obaveza da govornik zna stereotipe povezane s određenom riječju nije ispunjena, primjeri su neuspješnih pokušaja priklanjanja.

Ključne riječi: priklanjanje (*deference*), Putnam, stereotipi, namjera, zadano priklanjanje

BRAINS IN VATS AND SEMANTIC EXTERNALISM: NEW HOPE FOR THE SKEPTIC

MOZGOVI U POSUDAMA I SEMANTIČKI EKSTERNALIZAM: NOVA NADA ZA SKEPTIKA

STEFAN RINNER
Munich Center for Mathematical Philosophy, Germany

ABSTRACT

Different thought experiments have been offered to argue for the skeptical claim that sound empirical knowledge is impossible. One of these thought experiments assumes that we are eternal brains in a vat with systematically delusory experiences. In (Putnam 1981), Putnam responds to the skeptical challenge that contrary to our initial assumption we can know a priori, i.e. independent from experience, that we aren't eternal brains in a vat. Putnam argues that the skeptical hypothesis that we are eternal brains in a vat is inconsistent with the received view regarding reference and truth, semantic externalism, which says that a referential expression e refers to an object o if and only if e is appropriately causally related to o. There are different versions of Putnam's argument. In this paper, I will discuss the three main versions of the argument; i.e. a reconstruction of Putnam's original argument in (Putnam 1981), Brueckner's simple argument (Brueckner 2003; 2016, Section 3 and 4),

and a reconstruction of Brueckner's disjunctive argument (Brueckner 2016, Section 4). It is generally assumed that Putnam's original argument does not show that the skeptical hypothesis that we are eternal brains in a vat is inconsistent with semantic externalism. In this paper, I will argue that the same is true of Brueckner's simple argument and of Brueckner's disjunctive argument. Although from this it won't follow that semantic externalism is consistent with the skeptical hypothesis, it will show that it is also not yet decided that it is not.

Keywords: Brain-in-a-vat Scenario, Skeptical Challenge, Semantic Externalism, Hilary Putnam

SAŽETAK

Različiti misaoni eksperimenti bili su ponuđeni kako bi se argumentiralo za skeptičku tvrdnju da je zdravo empirijsko znanje nemoguće. Jedan od ovih misaonih eksperimenata pretpostavlja da smo vječni mozgovi u posudama s neprekidnim obmanjujućim iskustvima. Putnam (u Putnam, 1981.) odgovara na skeptički izazov da, suprotno našoj prvobitnoj pretpostavci, možemo znati *a priori*, to jest neovisno o iskustvu, da nismo vječni mozgovi u posudama. Putnam argumentira da skeptička hipoteza o tome da smo vječni mozovi u posudama nije konzistentna s usvojenim gledištem koje se tiče značenja i istine, semantičkog eksternalizma, a koje kaže da referencijalni izraz *e* referira na objekt *o* akko je *e* odgovarajuće kauzalno povezano s *o*. Postoje različite verzije Putnamovog argumenta. U ovom radu, prodiskutirat ću tri glavne verzije argumenta, odnosno rekonstrukciju Putnamovog izvornog argumenta (u Putnam, 1981.), Bruecknerov jednostavni argument (Brueckner, 2003; 2016, dijelovi 3 i 4), i rekonstrukciju Bruecknerovog disjunktivnog argumenta (Brueckner 2016, dio 4). Općenito se pretpostavlja da Putnamov izvorni argument ne pokazuje kako je skeptička hipoteza da smo vječni mozgovi u posudama nekonzistentna sa semantičkim eksternalizmom. U ovom radu, argumentirat ću da isto vrijedi i za Bruecknerov jednostavni argument, kao i za Bruecknerov disjunktivni argument. Iako iz ovoga neće slijediti da je semantički eksternalizam konzistentan sa skeptičkom hipotezom, pokazat će se kako nije još odlučeno ni da nije.

Ključne riječi: scenarij mozga u posudi, skeptički izazov, semantički eksternalizam, Hilary Putnam

DREAMS IN A VAT

SNOVI U POSUDI

DANILO ŠUSTER
University of Maribor, Slovenia

ABSTRACT

Putnam's semantic argument against the BIV hypothesis and Sosa's argument against dream skepticism based on the imagination model of dreaming share some important structural features. In both cases the skeptical option is supposed to be excluded because preconditions of its intelligibility are not fulfilled (affirmation and belief in the dream scenario, thought and reference in the BIV scenario). Putnam's reasoning is usually interpreted differently, as a classic case of deception, but this feature is not essential. I propose to interpret BIV's utterances as cases of reference failure best captured by truth-value gaps. Both anti-skeptical strategies are then vulnerable to the same type of objections (how do we know what state we are in or how do we know what kind of language do we speak).

Keywords: Putnam, Sosa, brain in a vat, dream argument, disquotation, negation

SAŽETAK

Putnamov semantički argument protiv MUS hipoteze i Sosin argument protiv skepticizma u vezi sa snovima temeljen na imaginativnom modelu sanjanja, dijele neke važne strukturalne značajke. U oba slučaja skeptička pozicija je isključena zbog neispunjavanja preduvjeta njene razumljivosti (potvrđivanje i vjerovanje u scenariju sna, misao i značenje u MUS scenariju). Putnamovo razmatranje je obično interpretirano drugačije – kao klasični slučaj obmane – no ova značajka nije ključna. Predlažem da se MUS-iskazi interpretiraju kao slučajevi propalog referiranja najbolje izraženog preko praznina unutar istinosnih vrijednosti. Obje antiskeptičke strategije su tada ranjive na isti tip prigovora (Kako znamo u kojem smo stanju? ili: Kako znamo koju vrstu jezika govorimo?).

Ključne riječi: Putnam, Sosa, mozak u posudi (MUP), argument sna, diskvotacija, negacija

IN DEFENSE OF THE TWIN EARTH–THE STAR WARS CONTINUE

U OBRANU ZEMLJE-BLIZANKE–RATOVI ZVIJEZDA SE NASTAVLJAJU

NENAD MIŠČEVIĆ
University of Maribor, Slovenia, Central European University, Hungary

ABSTRACT

The paper discusses the meta-philosophy of thought-experiments, in particular its neglected diachronic aspect, focusing on Putnam's work as the paradigm case, and on the trail(s) that developed out of the Twin Earth thought-experiment. Putnam's experiment is discussed from a perspective that combines metaphilosophy and actual history of analytic philosophy. Peter Unger has branded the whole debate around it as empty and fruitless. A meta-philosophical analysis shows him to be wrong. The experiment as originally proposed already appeals to a broad variation of examples and intuitive induction; the variation continues in other works addressing the issue, and produces interesting results. The second aspect is the search for reflective equilibrium, lasting till the present day. The internal logic of these processes is discussed, in order to show that the accusation for emptiness turns against Unger himself. In general, debates around thought-experiments, the already famous and also around new ones, make a large part of contemporary analytic philosophy. The way to understand a large part of this, and of debates surrounding it, is to link it to the internal understanding of a typical thought-experiment; stages of a particular experiment get discussed, developed and changed in the history of a particular trail produced by it. This is an important way in which a philosophical tradition is born, and we need to combine synchronic and diachronic approaches in order to understand it.

Keywords: thought-experiments, Twin Earth, meta-philosophy

SAŽETAK

Rad diskutira o metafilozofiji misaonih eksperimenata, točnije, o njenom zanemarenom dijakronijskom aspektu, fokusirajući se na Putnamov rad kao primjerni slučaj i na tragove koji su se razvili iz misaonih eksperimenata o Zemlji-blizanki. Putnamov eksperiment prodiskutiran je iz perspektive koja kombinira metafilozofiju i aktualnu povijest analitičke filozofije. Peter Unger proglasio je cijelu debatu oko navedenog kao ispraznu i jalovu. Metafilozofska analiza pokazuje da je u krivu.

Eksperiment, kako je izvorno predložen, već privlači široke varijacije primjera i intuitivne indukcije; varijacije se nastavljaju u drugim radovima koji se odnose na problem i polučuju zanimljive rezultate. Drugi aspekt je potraga za reflektivnim ekvilibrijem koja traje do današnjih dana. Unutrašnja logika ovih procesa prodiskutirana je ne bi li se ukazalo da se optužba za ispraznost okreće protiv samog Ungera. Općenito, debate oko misaonih eksperimenata, one već slavne i one oko novih, čine velik dio suvremene analitičke filozofije. Način na koji razumjeti velik dio ove teme i debate koje se vrte oko nje, jest da se poveže ovu temu s unutarnjim razumijevanjem tipičnog misaonog eksperimenta; stadiji određenog eksperimenta raspravljeni su, razvijeni i promijenjeni unutar povijesti određenog traga koji su ostavili. Ovo je važan način na koji je filozofska tradicija rođena i trebamo kombinirati sinkronijske i dijakronijske pristupe ne bi li ga razumjeli.

Ključne riječi: misaoni eksperimenti, Zemlja-blizanka, metafilozofija

INSTRUCTIONS FOR AUTHORS

Submission

Submitted manuscripts ought to be:

- unpublished, either completely or in their essential content, in English or other languages, and not under consideration for publication elsewhere. If it emerges that an accepted submission does not satisfy the previous conditions, the Editors reserve the right to reject it. If, due to the authors' failure to inform the Editors, already published materials will appear in EuJAP, the Editors will report the authors' unethical behaviour in the next issue. In any case, the Editors and the publisher will not be held legally responsible should there be any claims for compensation following from copyright infringements by the authors

- sent exclusively electronically to the Editors (eujap@ffri.hr) (or to the Guest editors in the case of a special issue) in a Word compatible format

- prepared for blind refereeing: authors` names and their institutional affiliations should not appear on the manuscript. Moreover, "identifiers" in MS Word Properties should be removed

- accompanied by a separate file containing the title of the manuscript, a short abstract (not exceeding 250 words), keywords, academic affiliation and full address for correspondence including e-mail address

- in American or British English

- no longer than 8000 words, including footnotes and references

Style

Accepted manuscripts should:

- follow the guidelines of the most recent Chicago Manual of Style

- contain footnotes and no endnotes

- contain references in accordance with the author-date Chicago style, here illustrated for the main common types of publications (T = in text citation, R = reference list entry)

Book
T: (Nozick 1981, 203)
R: Nozick, R. 1981. *Philosophical Explanations*. Cambridge: Harvard University Press.

Chapter or other part of a book
T: (Fumerton 2006, 77-9)
R: Fumerton, R. 2006. The Epistemic Role of Testimony: Internalist and Externalist Perspectives. In *The Epistemology of Testimony*, ed. J. Lackey and E. Sosa, 77-92. Oxford: Oxford University Press.

Edited collections
T: (Lackey and Sosa 2006)
R: Lackey, J. and E. Sosa, eds. 2006. *The Epistemology of Testimony*. Oxford: Oxford University Press.

Article in a print journal
T: (Broome 1999, 414-9)
R: Broome, J. 1999. Normative requirements. *Ratio* 12: 398-419.

Electronic books or journals
T: (Skorupski 2010)
R: Skorupski, J. 2010. Sentimentalism: Its Scope and Limits. *Ethical Theory and Moral Practice* 13: 125-136. http://www.jstor.org/stable/40602550

Website content
T: (Brandon 2008)
R: Brandon, R. 2008. Natural Selection. *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. Accessed September 26, 2013. http://plato.stanford.edu/archives/fall2010/entries/natural-selection

Forthcoming
For all types of publications followed should be the above guideline style with exception of placing 'forthcoming' instead of date of publication. For example, in case of a book:
T: (Recanati forthcoming)
R: Recanati, F. forthcoming. *Mental Files*. Oxford: Oxford University Press.

Unpublished material
T: (Gödel 1951)
R: Gödel, K. 1951. *Some basic theorems on the foundations of mathematics and their philosophical implications*. Unpublished manuscript, last modified August 3, 1951.

Final proofreading

Authors are responsible for correcting proofs.

Permission to republish

For permission to republish material in print or electronic format please contact the Editors (eujap@ffri.hr).

Subscriptions

A subscription comprises two issues. All prices include postage.

Annual subscription:
International:
individuals € 10
institutions € 15
Croatia:
individuals 40 kn
institutions 70 kn

Bank: Zagrebačka banka d.d. Zagreb
SWIFT: ZABAHR 2X
IBAN: HR9123600001101536455
Only for subscriptions from Croatia,
please add: "poziv na broj": 0015-03368491

European Journal of Analytic Philosophy is published twice per year.

The articles published in the European Journal of Analytic Philosophy are indexed in The Philosopher's Index and included in ERIH PLUS.