A Priori truth in the natural world : a non-referentialist response to Benacerraf's dilemma

Novák, Z.

CHAPTER 5

Referentialist Responses to Benacerraf's Dilemma II: Anti-Realist Construals of Truth

Introduction

Among those who reject the deflationist idea that a proper theory of truth is neutral to the metaphysical concern of what there is and the epistemological concern of what we can know, some query the adequacy of realism about truth in the semantics of at least some discourses on explanatory considerations. According to these critics, the realist construal of certain knowable truths excludes the possibility of a reasonable explanation of how we can acquire knowledge, or even develop ideas, of the intended circumstances.[1]

By adopting an anti-realist construal of truth conditions and intended referents, the advocate of standard referentialism can provide a *prima facie* acceptable response to Benacerraf's epistemological challenge in the semantics of our paradigm *a priori* discourses. As we have seen in chapter 3, in its most generic form, the epistemological argument queries the adequacy of a referentialist *and* realist construal of truth in the semantics of discourses in which we are supposed to acquire knowledge or reliable beliefs about causally inert domains. Similarly to her deflationist colleagues, by abandoning the realist tenet, an anti-realist may subscribe to the standard referentialist construal of truth without falling prey to Benacerraf's argument. If the intended referents of our claims about causally inert domains are

[1] A detailed articulation of what is meant by realism and anti-realism in this work can be found in the third section of chapter 1.

construed in an anti-realist way, then by adopting a referentialist construal of the truth conditions of these claims one does not commit oneself to the idea that our knowledge of these domains is knowledge of the extra-mental existence of some causally inert (e.g. platonic) entities.

In the previous chapter, I argued that the endorsement of a realist construal of the truth conditions of a given claim is a precondition for a suitable explanation of the objectivity of the truth value of this representation. In accordance with the methodological principles put forward in chapter 2, I assume that this result is sufficient to cancel out the envisaged advantages of anti-realism about truth, and to establish the inadequacy of an anti-realist form of referentialism in the semantics of our paradigm *a priori* discourses as well. Nonetheless, in the light of the same principles, we must recognise also that a thorough defence of realism about truth requires more than the exposition of the explanatory advantages of this conception. In particular, it requires the demonstration of the failure of the opponents' arguments against this doctrine too.

In this chapter, I shall examine two major groups of arguments that are usually taken as the most influential anti-realist challenges to realism about truth: in section 1, I shall discuss Michael Dummett's acquisition and manifestation arguments against semantic realism, and in section 2, Hilary Putnam's internal realist argumentation against metaphysical realism and the correspondence theory of truth. The structure of discussion in the two sections will be parallel: first, I shall provide a brief reconstruction of the arguments; second, I shall consider whether they attack realism in the relevant sense of the term and whether the semantical doctrines adopted by Dummett and Putnam could help the advocates of referentialism escape Benacerraf's original or modified and generalised challenge in the semantics of discourses about causally inert domains; and finally, I shall examine the presented arguments, identify the most plausible considerations that may support the adoption of their crucial premises, and explain why I think that the reasons identified are

wrong and thus unable to justify the intended anti-realist conclusions.

1. Dummett: The Acquisition and the Manifestation Argument

There are two influential anti-realist considerations, propounded by Michael Dummett and further developed by his supporters, such as Crispin Wright, Bob Hale and Neil Tennant, to the effect that, contrary to what realists suggest, our actual understanding of truth cannot be verification-transcendent in character. They are best known as the acquisition and the manifestation challenge to semantic realism.[2] The first purports to show "that no one could

[2] Dummett (1978), Dummett (1991), Dummett (1993), Wright (1993), Hale (1997) and Tennant (1997). Beyond these major challenges, authors working on this Dummett-inspired wing of the anti-realist tradition have advanced a number of other cases against their realist opponents. The two most influential among them are Wright's "argument from normativity" and "argument from rule-following". The first attempts to show that semantic realism "offends against the essential normativity of meaning, whereby meaning has to be determined by constraints by which one can aim to regulate one's linguistic practice" Wright (1993), 26). The second purports to demonstrate that in view of Wittgenstein's discussion of rule-following in the *Philosophical Investigations* the realist idea of the objectivity of meaning, on which our case for realism about truth in chapter 4 was also based, is untenable. A complete defence of the realist position advocated in this work would require a detailed discussion of these more recent arguments as well. In order to abridge this section, however, let me get over them by making a brief note on why I think they fail to achieve the dialectical purpose they were designed to achieve. The main problem with Wright's argument from normativity is that it assumes the "essential normativity" of meaning and truth. In contrast to this assumption, I believe that meaning and truth are not essentially normative. In other terms, I think that the contrast between correct (regular) and incorrect (irregular) use could exist even if we did not appraise truth or the correct application of representations in general. In this regard, I agree with Horwich, who denies what Gibbard, Brandom and Lance & Hawthorne assume, namely that the evaluative import of a meaning-property is enough to make that property *constitutionally* evaluative (i.e. to be explicated in terms of what one ought and ought not to say). Horwich (2005), 12-13, 81-82, Gibbard (1994), Brandom (1994), and Lance and Hawthorne (1997). Wright's second argument, which

actually form an understanding of a statement if to do so required grasping transcendent truth-conditions", while the second "that no one who had somehow achieved such grasp could give sufficient reason to another to suppose that he had, no matter how extensive the survey of his linguistic behaviour".[3] Before examining the cogency of these challenges, let me spell out the

attacks the realist idea of the objectivity of meaning and truth on broadly Wittgensteinean considerations, is more interesting and harder to refute. Nevertheless, a hint of the right direction can be easily given here. The crucial assumption of this argument is that Wittgenstein's considerations against the *objectivity* of meaning should not be taken, *pace* Kripke, as considerations against the *existence* of meaning. Rather, they are best regarded as, correctly, pointing to an alternative, anti-realist and naturalist construal, according to which our judgements about what counts as following a rule are "ceaselessly determined by features of our sub-rational natures". Wright (1993), 28. In other terms, Wright's suggestion is that the contrast between correct and incorrect applications is created and maintained by our natural propensities to react and judge in particular ways, so it cannot be characterised as a judgement-independent feature to be detected in the world. Now, I fully agree with Wright, and Nelson Goodman for that matter, that our natural propensities to react and judge in particular ways are essential for the functioning of our classificatory and recognitional capacities, and thereby for the existence of stable semantic properties in the world. A condition in the mind-independent world qualifies as the correct declarative use condition of a certain representation partly because its obtaining triggers a certain judgement in us. Thus, the epistemic fact that we have certain propensities to judge is, indeed, constitutive of the semantic contents of our representations. This circumstance, however, does not undermine the objectivity of truth or correct use. Once the semantic content of our representations is fixed along the suggested anti-realist lines, the obtaining or absence of the conditions specified by these contents will be entirely independent of what we ever believe or know about this issue. The epistemisation of content would destroy the idea of objective semantic values only if there were no difference between content-fixing stipulations and content-asserting judgements. A content-fixing stipulation, however, is never false. If truth-apt at all, it is true because we take it to be so. So, if an anti-realist intends to maintain the idea of false or non-trivially true beliefs, then she must concede that genuine judgements (as opposed to mere stipulations) have fixed semantic content, and thus determinate truth conditions, which either obtain or not, independently of our actual opinions of this circumstance. In other terms, she must cease to be an anti-realist about truth after all. The core idea behind this response to Wright's second argument will appear in the discussion of Dummett's notion of semantic realism in the main text below as well.

[3] Wright (1993), 26.

most important assumptions that they rely on. According to one possible reconstruction, both cases start with the adoption of two presumably uncontentious premises:

1. We can understand effectively not decidable declarative sentences.
2. Understanding a declarative sentence is a matter of knowing its truth conditions.[4]

From this, they conclude:

3. We have knowledge of the truth conditions of effectively not decidable declarative sentences.

After this, they invoke the crucial realist premise:

4. Truth can be a recognition transcendent property: effectively not decidable declarative sentences have recognition transcendent truth conditions, which may obtain even if we have no effective method to establish this, implying that the sentences in question have also objective, determinate truth value.[5]

[4] Sometimes this premise is taken as a specific tenet of semantic realism. This is certainly legitimate, if truth conditions (as opposed to, say, assertability conditions) are understood in a realist way. In his earlier writings, Dummett often depicted his anti-realist as an opponent of the received truth conditional semantics. Dummett (1973), 225, 226-227, 230-231. As we observed in chapter 4, however, the idea that understanding a sentence is to know under which conditions it is true does not imply anything about the metaphysical or epistemological status of these conditions, and thus can be endorsed by semantic anti-realists and deflationists as well. For anti-realist recognitions of this point, see Dummett (1973), 232, Dummett (1991), 317-318, and Wright (1993), 18.

[5] As Dummett rightly emphasises, the real issue between realists and anti-realists is not whether a correct semantics is two-valued or many-valued, but rather whether these values are determined by the obtaining of some recognition transcendent or some verifiable conditions. Dummett (1991), 304-305. On the basis of her understanding (i.e. her knowledge of truth conditions), in the case of

From (3) and (4), then, they derive the realist conclusion:

5. We have knowledge of recognition transcendent truth conditions.

Having reached this point, the challenges diverge. Each invokes two further (presumably incontestable) premises, which, taken together, seem to be incompatible with (5). In the case of the acquisition challenge, the two premises are:

6. All knowledge of truth conditions is acquired.
7. We cannot acquire knowledge of recognition transcendent truth conditions.

In the case of the manifestation challenge:

8. All knowledge of truth conditions must be manifest in the use of the relevant declarative sentences.
9. Knowledge of recognition transcendent truth conditions cannot be manifest in the use of the relevant declarative sentences.

Supposing that the latter pairs of premises are correct, the challenges arrive at the conclusion that (5) must be false (i.e. that we cannot have knowledge of recognition transcendent truth conditions). Finally, supposing that (1) and (2) are also incontestable, from the falsity of (5) the challenges derive that (4) must be false too (i.e. that truth cannot be a recognition transcendent property after all).

A realist may respond to these challenges in several ways. The responses can be divided into two major categories. First, one may categorically refuse at least one of the two premises invoked in the second half of the challenges (i.e. (6) or (7), and (8) or (9), respectively), and maintain that (5), and with it (4), is still true. Alternatively, one may argue that there is a sense in which those pairs of premises are, indeed, true, but this sense requires an understanding of the term 'recognition transcendent' under which (4) is not a correct representation of the realist position.[6] Some of the logically available responses falling into these two categories are fairly implausible, and as such can be easily set aside. For instance, hardly any realist would deny the adequacy of premise (6). The acceptability of premise (8) may be more contentious, but there is certainly a sense in which no realist would deny that our ideas of truth conditions must be manifest in the way we use our truth-apt representations. So, the *prima facie* interesting responses that a realist may give to the above challenges are either the categorical refusal of premise (7) and premise (9), or the acceptance of them only under a certain interpretation of the term 'recognition transcendent' on which premise (4) does not represent well her semantical position.

Now, the best way to start the assessment of Dummett's challenges is to address the pending interpretative question, and clarify how the (closely related) key terms of 'effectively not decidable' and 'recognition transcendent' are to be understood in a faithful reconstruction of the position advocated by Dummett's opponents. With the resulting interpretation in mind, we can then examine the plausibility of Dummett's crucial assumptions, namely that we cannot acquire and make manifest in use knowledge of such *unrecognisably* obtaining truth conditions.

In order to find out more about the intended notion of recognition transcendence, let me briefly review Dummett's

an effectively not decidable declarative sentence, a realist believes that she cannot establish whether or not these conditions obtain, while an anti-realist supposes that she can establish that the conditions (actually) do not obtain. It must be noted, however, that Dummett also separates a narrow understanding of the term 'realism', which is associated with the principle of bivalence. In this sense realism belongs, with other finitely many-valued accounts, to a broader category of "objectivist" theories, which are in turn contrasted with the representatives of anti-realism. Dummett (1991), 326.

[6] For a concise review of the grounds on which the adequacy of Dummett's construal of realism may be doubted see Hale (1997), 283-288.

major paradigms of the disputed realist position.[7] His prototypical examples include realism about the physical (external) world, realism about the abstract universe of mathematics, realism about the mind, realism about the theoretical entities of science, realism about moral facts, and realism about the past and the future. One way to get closer to Dummett's idea of recognition transcendence is to recall those conditions whose obtaining or absence his paradigm anti-realists would regard as verifiable, and thus suitable for being taken as conditions of truth. In the case of discourses about the physical world, these conditions obtain in sense experience. In the case of mathematics, they are typically conceived as inferential relations with certain sets of axioms. In the case of mental state attributions, they are features of overt behaviour. In the case of scientific claims about theoretical entities, they are observable public events explained or predicted by the theory at hand. In the case of moral discourses, they are conditions obtaining in the evaluating subject. Finally, in the case of claims about the past, they are data that can be recalled from memory or some currently available public records, while in the case of claims about the future, they are currently observable tendencies in the world.[8]

The common feature of these conditions, according to Dummett, is that we have an effective method to establish whether or not they obtain.[9] We can check on our experience, we can execute a proof, we can observe a person's overt behaviour or the public events explained or predicted by our scientific theories, we can detect conditions obtaining in ourselves while evaluating things, we can retrieve data from our memory or our history books, and we can recognise current tendencies in the world. On Dummett's view, our judgements about the domains listed above are, in fact, governed by the obtaining or absence of these verifiable conditions, rather than the obtaining of those epistemically remote ones that these judgements purport to be about. Therefore, what we can learn about the correct declarative use or truth conditions of our representations, and what can be manifest from the resulting knowledge in our use of these representations is all to be specified in terms of these potentially observable conditions. The main problem with semantic realism, according to this picture, is that it construes the truth conditions of a sentence in terms of what it purports to be about, thus implying that it is determinately either true or false, even if we have no effective method to establish whether or not the conditions of its truth actually obtain. As Dummett never failed to emphasise, it is due to this commitment that a common characteristic of realist doctrines is an insistence on the principle of bivalence or determinate semantic values. In contrast, one of the most important concomitants of the suggested anti-realist construals is that if we have no effective means to establish the truth value of a sentence, then we have no reason to suppose that it has a determinate truth value either.

Now, the first important thing to be observed concerning Dummett's characterisation of the dispute between realists and anti-realists is that it is slightly at odds with the one I put forward earlier in this work. On the construal I proposed, the dispute concerns the metaphysical status of thinkable individuals, properties, facts and domains, and can be characterised as being about the question whether or not these thinkable entities exist independently of our actual thoughts and knowledge of this circumstance. When applied in semantics, the contrasted doctrines were taken to concern the metaphysical status of semantically relevant entities, such as truth conditions or subject matters. Realism about truth, for instance, was characterised as the view that the truth conditions of our representations obtain independently of our actual thoughts or knowledge of this circumstance, while the opposite doctrine was described as the view that some epistemic facts involving certain truth conditions are constitutive of the obtaining of these conditions.

On Dummett's construal, what an anti-realist argues for is not the epistemisation of the *obtaining* or *absence* of truth conditions, but the epistemisation of *semantic contents* (i.e. the

[7] Dummett (1991), 4-8.
[8] Dummett (1991), 322.
[9] Dummett (1991), 317.

judgement-independently obtaining or absent truth conditions themselves). In other terms, what Dummett's anti-realist assumes is not that the truth value of our beliefs depends on anyone's actual knowledge or opinion of this issue, but instead that the conditions whose obtaining or absence determines those values are always verifiable (i.e. such that we have an effective, though fallible, method to determine whether or not they actually obtain). This is why most anti-realist truth conditions cited above can qualify as real according to the understanding adopted earlier in this work: the fact that we can establish whether or not they obtain is constitutive of their qualifying as conditions of truth, but not necessarily of their actual obtaining or absence.[10] Dummett's verifiability constraint, accordingly, need not conceptually prevent anyone, not even an entire linguistic community, at any single moment, from committing an epistemic mistake while trying to establish the truth value of a certain declarative sentence. Semantic anti-realism, in this sense, does not exclude the explanation of the objectivity of truth and correct use along the realist lines suggested in chapter 4, and thus need not prove to be inadequate in view of the second adequacy condition set for a theory of truth in chapter 2.

A second important observation concerning Dummett's characterisation of the debate between realists and anti-realists is that in the semantics of discourses about the specified recognition transcendent domains the conception that Dummett's anti-realists advocate about truth is clearly non-referentialist in character, in so far as it holds that the truth

[10] In those cases in which the suggested anti-realist conditions are supposed to be mental in the narrow sense of the term, as in the case of phenomenalism and constructivism, the epistemisation of content involves the epistemisation of truth and falsity as well. This, however, is certainly not a general characteristic of anti-realism in Dummett's sense of the term. In fact, one may even argue that an advocate of the acquisition and manifestation challenges cannot consistently subscribe to a doctrine according to which the truth conditions of some declarative sentences obtain within a private domain. For such a theory would render these conditions no more acquirable and manifest in use than, on Dummett's view, their verification-transcendent counterparts.

conditions of these sentences are not to be construed in terms of the relevant (recognition transcendent) subject matters. In consequence of this non-referentialist commitment, we may conclude, Dummett's anti-realism is not a suitable doctrine for a referentialist to endorse in response to the amended and generalised form of Benacerraf's challenge in the semantics of discourses in which we are supposed to acquire knowledge or reliable beliefs about causally inert domains.

In view of these two observations, one might think that Dummett's anti-realist position is not even so far from the non-referentialist account advocated in this work in relation to the paradigms of *a priori* truth. In fact, both of us believe that there are good explanatory considerations against the universal adequacy of the standard referentialist construal of truth, and also that these considerations are broadly epistemological in character. It must be also noted, however, that the target class of Dummett's challenges may be much larger than that of my broadly Benacerrafian challenge, and our problems with these classes are not entirely the same either. What Dummett queries is whether we can acquire and manifest in use ideas of conditions whose obtaining or absence cannot always be effectively verified. My question, in contrast, is whether we may have any reason to suppose, as referentialists do, that our knowledge about causally inert domains is knowledge of the (thick) obtaining of conditions within these domains. Dummett worries about the communicability of knowledge of meanings that are construed in a "realist" way, while I worry merely about the acquirability of knowledge of the obtaining of causally inert referential conditions.

Having said this, of course, we may still wonder whether Dummett's considerations provide us with good reasons for replacing the standard referentialist construal of meaning and truth in the semantics of one or another of the above problematic discourses with his "anti-realist" alternative.[11] The first question

[11] Despite the observed difference between my and Dummett's construal of semantic realism, I will keep referring to the position he argues against by the

that we set out to clarify was how the (closely related) key terms of 'effectively not decidable' and 'recognition transcendent' were to be understood in a faithful reconstruction of the position advocated by Dummett's opponents.

Well, in view of the above examples and the subsequent characterisation of the debate, what Dummett's opponents apparently wish to maintain is that we can form ideas of conditions whose obtaining or absence we may not be able to establish with our actual epistemic methods and capacities. An advocate of this view may deny that any of these conditions is beyond the reach of all conceivable epistemic activity (i.e. that there is anything in the world that resists discovery even in principle). It may well be that we have no idea how to determine the truth value of a belief about, say, a remote past event or an abstract mathematical fact, but this is merely a consequence of our contingent epistemic predicament (i.e. the actual lack of any means to gain evidence of that event or that fact), and does not imply that the beliefs in question do not have determinate truth values. If we had other epistemic capacities or available methods to acquire the missing evidence, then we could establish the truth value of these beliefs just as much as we happen to do that now in the case of some others about the very same domains. Accordingly, the view against which Dummett must be taken to argue is not that we have knowledge of truth conditions that are recognition transcendent in the radical sense of the term, but merely that:

> (5*) We have knowledge of truth conditions whose obtaining or absence we cannot recognise by means of our actual methods and epistemic capacities.[12]

Now, Dummett's reasoning against the advocates of (5*) would be clearly fallacious if the term 'recognition transcendent' received a more radical reading in the subsequent premises of his arguments. So, in order to ensure the validity of his reasoning, the notion of recognition transcendence must be understood along the same lines in premise (7) as well as in premise (9). This leaves us with the following formulations:

> (7*) We cannot acquire knowledge of truth conditions whose obtaining or absence we cannot recognise by means of our actual methods and epistemic capacities.

> (9*) Knowledge of truth conditions whose obtaining or absence we cannot recognise by means of our actual methods and epistemic capacities cannot be manifest in the use of the relevant declarative sentences.

Arguably, the assessment of Dummett's acquisition and manifestation challenges to the semantical conception expressed by (5*) depends largely on the cogency of premises (7*) and (9*), respectively. So, the question that I should briefly examine in the remaining part of this section is whether it is plausible to suppose that there is no way to acquire or make manifest in use ideas of conditions whose obtaining or absence we cannot recognise by means of our actual methods and epistemic capacities.

There are several realist attempts to answer Dummett's challenges by querying the correctness of premise (7*) and premise (9*). Most of these answers were sufficiently discussed

term 'realism' or 'semantic realism' (without using quotation marks) in the remaining part of this section.

[12] Fitch (1963) advances a simple case to demonstrate that if there are truths that we actually do not know, then there must be truths that are unknowable even in principle. In particular, he argues that if *p* is a truth that is never known then it is unknowable *that* p *is a truth that is never known*. If the argument is correct, then a realist has in fact no reason to replace (5) with (5*) in the characterisation of her semantical position. According to the argument, if she endorses (5*), then she must, on pain of inconsistency, also endorse (5). For an analysis of the argument see Williamson (2000), 270-275. (Thanks to András Simonyi for reminding me of this point.)

and responded to by present-day advocates of Dummett's semantic anti-realism.[13] The simplest realist point, however, which confronts the proponents of (7*) and (9*) with the fact that our actual practice is realist in character, has never been, in my view, convincingly dealt with by Dummett or his followers.

It is not as if any of them denied that the anti-realist conception of meaning and truth defended by Dummett implies a revisionist approach to our received linguistic practice. For Dummett, the most important feature of this practice that is usually associated with a realist understanding is our propensity to reason in accordance with the laws of classical logic. Beyond this feature, however, there are a number of other characteristics that a realist could regard as a manifestation of our ideas of verification-transcendent truth conditions. First, we do not generally licence the identification of truth with provability or supportability by evidence. For instance, we do not think that any collection of observable historical records or aspects of another person's overt behaviour would qualify as conditions of truth, respectively, for our beliefs about past events or for those about the person's mental states. Second, we construe truth as a stable and absolute property. Once a truthbearer acquires its truth value, it never loses it, and the value in question does not come in degrees. If Dummett's anti-realist were right, then a belief about a currently observed event could gradually cease to be true, as its object can fade away without leaving a trace about its existence for posterity. Similarly, if Dummett's anti-realist were right, then the truth value of an empirical belief could hardly be absolute, since empirical verification always comes in degrees. Third, the simple fact that we understand effectively undecidable sentences, and maintain that our understanding consists in grasping truth conditions seems already a clear manifestation of our recognition transcendent understanding of truth.[14] Finally, in so far as our concept of truth satisfies the equivalence or disquotation schema, while our concepts of genuine epistemic properties (e.g. actual provability, effective decidability, rational acceptability or warranted assertability) do not, truth cannot be consistently identified with any of these epistemic features at all.[15]

Dummett does not overlook the difficulty that the existence of our received linguistic practice may present for an anti-realist construal of meaning and truth. For instance, he devotes long passages to explaining how someone who adopts a use theory of meaning can still be a revisionist *vis-à-vis* our actual linguistic practice.[16] His ideas on this issue may become clear from the following quotations:

> An existing practice in the use of a certain fragment of language is capable of being subjected to criticism if it is impossible to systematise it, that is, to frame a model whereby each sentence carries a determinate content which can, in turn, be explained in terms of the use of that sentence. What makes it possible that such a practice may prove to be incoherent and therefore in need of revision is that there are different aspects to the use of a sentence; if the whole practice is to be capable of systematisation in the present sense, there must be a certain harmony between these different aspects.[17]

The aspects that Dummett has in mind are:

> …the conventions governing the occasions on which the utterance is appropriately made and those governing both the responses of the hearer and what the speaker commits himself to by making the

[13] For a short summary and critical discussion of these realist answers see Hale (1997), 276-283.

[14] McDowell (1981), 322. The same insight is at the heart of Alexander Miller's defence of semantic realism against the manifestation argument in Miller (2002).

[15] Wright (1992).

[16] Dummett (1973), 218, Dummett (1991), 341-342.

[17] Dummett (1973), 220.

> utterance: schematically, […] the *conditions for* the utterance and the *consequence of* it.[18]

Notice, however, that Dummett's realist opponents need not deny the compatibility of a use theory of meaning, on the one hand, and a revisionist stance toward our actual linguistic practice, on the other. They may admit that if this practice could not be systematised in the sense specified above, then it could not be taken as a sign of a coherent realist understanding of truth either, and it should in some way be revised after all. Dummett's problem with this practice, however, is not that it resists systematisation. It is rather that a realist semantics, which is supposed to provide this systematisation, raises serious epistemological questions that its advocates cannot hope to answer in a suitable way. In particular, it undermines the explanation of how we could acquire and make manifest in use our knowledge of these conditions. What the opponents of this argumentation may not understand is why the highlighted characteristics of our actual linguistic practice should not be taken as a manifestation, and an empirical ground for believing in the existence, of our realist ideas of truth conditions.[19]

The question seems certainly well-addressed. In Dummett's view, the specified characteristics do not imply the existence of a realist understanding. It is true that such an understanding would induce a practice with these characteristics. This is why by

[18] Dummett (1973), 221.

[19] The dialectical situation seems to be similar to that observed in section 3 of chapter 1, where we examined the implications of Blackburn's quasi-realist considerations. In both cases, arguments are put forward to the effect that realists are not in a position to manifest, or make sense of, their metaphysical or semantical commitments. Still, in both cases, the arguments are meant to provide realists with reasons for giving up their, allegedly non-manifestable or nonsensical, commitments. The moral, in both cases, seems to be the same: one may argue against a certain metaphysical or semantical position by demonstrating the incoherence of the cognitive or linguistic practice that is taken to be the manifestation of the position under consideration, but no such argument can start from the assumption that the view under attack cannot be acquired from or made manifest in communication.

observing the latter one may tend to believe in the existence of the former. But, as Dummett emphasises, our actual linguistic practice can be the result of blind training as well:

> Imagine, for example, that we had been subjected, since childhood, to a training in applying to counterfactual conditionals the laws of classical logic, construing the negation of a counterfactual as the opposite counterfactual. We should then be under a strong compulsion to do what we are often tempted to do now, namely, to suppose that any counterfactual must be determinately true or false independently of our knowledge, as when we wonder what would have happened if we had made some important decision in our lives otherwise than we did, in a frame of mind in which we submit to the illusion that there must be some definite answer, whether or not we can know it. But the fact that we reasoned in accordance with these classical laws would not show that we really had a realist notion of truth for counterfactual conditionals.[20]

So, the crucial question that Dummett must answer before advancing his two arguments presented earlier against semantic realism is why he thinks that the highlighted features of our actual linguistic practice are the result of blind training or some other coincidence, rather than the consequence of a realist construal of the circumstances whose obtaining is endorsed by the declarative use of semantically contentful representations. Why is it not enough, for instance, in order to manifest our realist understanding of the claim 'Caesar sneezed 15 times on his 19th birthday' to maintain that no presently observable fact, on the basis of which we would declaratively apply this representation, would guarantee its truth, since what this property consists in in

[20] Dummett (1991), 342.

the current case is the obtaining of some conditions in the remote past, not in the observable present?

Dummett's response to this question seems to be based on two tenets. First, although he acknowledges that in particular cases our knowledge of the truth conditions of a certain truth-apt representation can be made manifest by using some others whose meaning has been previously made manifest and thus accessible to the hearer, nevertheless he also emphasises that this method cannot be the only way to display this sort of knowledge, since our understanding exceeds our knowledge of synonymities.[21] At some point we must be able to manifest our knowledge of truth conditions in an implicit manner as well. Second, he believes that the only way in which we can implicitly manifest our knowledge of the truth conditions of a certain representation is by applying it when (and of course only when) the conditions in question recognisably obtain.[22] Now, suppose that Dummett's opponents are right, and we have knowledge of recognition transcendent truth conditions. Suppose that we are asked to make manifest this knowledge in our linguistic practice. First we may try to specify the content of this knowledge by using other sentences that are supposed to be used exactly when the conditions to be specified obtain. Upon request, we can continue this explicit specification until we run out of further *explicantes*. Since the sentences of the resulting sequence are all meant to specify the same conditions, the truth conditions of the ultimate *explicantes* must still be recognition transcendent in character. Of course, in order to satisfy the original requirement, we must finally be able to make manifest our knowledge of these conditions in an implicit way. And this is where we seem to get into trouble. For, if Dummett's second tenet is correct, then our knowledge of these conditions cannot be made manifest by the use of the ultimate *explicantes*. The only way we could show what we mean by these representations would be to use them when the relevant conditions recognisably obtain. But if the conditions are, indeed,

[21] Dummett (1973), 224.
[22] Dummett (1973), 224-225.

recognition transcendent in character, then we can hardly apply this method to display our knowledge of them. If this is so, however, then Dummett's premise (9*) is right: our alleged knowledge of truth conditions whose obtaining or absence we cannot recognise by means of our actual methods and epistemic capacities cannot be manifest in the use of the relevant declarative sentences.[23]

One may object that even if Dummett were right in maintaining that ideas of recognition transcendent truth conditions cannot be made manifest in linguistic practice, the earlier highlighted characteristics of our actual practice can still show that our understanding is not anti-realist in character. To this challenge, however, we already know Dummett's reply: since all knowledge of truth conditions must be manifest in linguistic practice, and no knowledge of the above realistically construed truth conditions can be made manifest in this manner, therefore anything that contradicts the anti-realist semantics in our actual practice must be eliminated from there once and for all.

With the above reconstruction of Dummett's considerations in support of the key premises of his manifestation and acquisition argument, I hope to have managed to present his case against semantic realism in its strongest possible form. Dummett's two tenets about the available methods of displaying semantical knowledge in linguistic practice seem certainly plausible, they seem to provide a clear case in favour of premises (7*) and (9*), and the adoption of these premises seems sufficient for deriving the anti-realist conclusion that our understanding cannot be realist in character. Is there anything else that a realist may object to this argumentation? What could be the intuitive ground for resisting Dummett's anti-realist conclusions?

[23] By reference to the correctness of premise (9*), then, one can make a strong case in support of premise (7*) too. Something that is not manifest in public linguistic practice can hardly be acquired by the observation of this practice alone. So, unless one believes that our knowledge of meaning is grounded in something more than the observation of linguistic practice, one can conclude that the correctness of premise (9*) entails the correctness of premise (7*) as well.

Well, I think that the intuitive problem with Dummett's anti-realist semantics is that it conflicts not merely with our actual linguistic practice, but also with our knowledge of our own actual understanding. The fact, for instance, that we do not licence the identification of truth with provability or supportability by evidence is not merely a fact about our current linguistic practice, but also something rooted in a reflectively observable feature of our own understanding. In principle, of course, an anti-realist may concede that there might be realist aspects of our private, idiosyncratic understanding, and then insist that these aspects are not communicable, and hence cannot be constitutive of the public meaning of our representations. But reflection also informs us that it is because of these realist features of our own understanding that we overtly refuse the revisions suggested by the anti-realist. In other terms, we seem to have good reasons to suppose that the earlier highlighted aspects of our actual linguistic practice are, indeed, the manifestations of a realist understanding.

Moreover, if we reflect upon our own understanding, we may also observe that our ideas of the truth conditions of effectively not decidable sentences are far more articulated than what is implied by a statement of their mere recognition transcendence. It is not only that we do not identify the truth conditions of the claim 'Caesar sneezed 15 times on his 19th birthday' with any presently observable conditions, whose obtaining or absence we can establish by means of our actual methods and epistemic capacities, but also we can distinguish between the former conditions and the truth conditions of the claim 'Napoleon sneezed 15 times on his 19th birthday'. This may also strengthen the realist's conviction that Dummett's argumentation must be in some way mistaken. Of course, a proper realist response to this argumentation must include an account of how, indeed, we can manifest our ideas of recognition transcendent truth conditions by means of the public use of our linguistic representations.

One commonly recognised feature of our understanding, which is nevertheless strikingly neglected by Dummett while he considers the ways in which we can manifest our semantical knowledge in linguistic practice, is the compositionality of this understanding. Presumably, the reason for which Dummett neglects this characteristic in the current context is that its existence seems to imply nothing about the relation of truth conditions to our actual epistemic capacities. The truth conditions of truth-apt representations can be determined by the correct declarative use conditions of their components and the way they are combined in a realist as well as in an anti-realist semantical framework. Nevertheless, if compositionality holds, then a realist may argue that beyond the two methods specified by Dummett there is also a further way in which we can demonstrate our knowledge of truth conditions: namely, by manifesting our knowledge of the correct declarative use conditions of the semantically basic constituents of our truth-apt representations, on the one hand, and our knowledge of those functions that determine the correct declarative use conditions of our complex representations in the light of those of their constituents, on the other.[24]

Of course, drawing attention to this third way of displaying knowledge of truth conditions, in itself, does not provide us with the required explanation of how we can manifest knowledge of recognition transcendent truth conditions. A brief reflection upon the semantical structure of our effectively undecidable sentences, however, may illuminate the explanatory benefit of the previous observation. Consider, for instance, the sentence 'Caesar sneezed 15 times on his 19th birthday'. While on a realist interpretation the truth conditions of this sentence are certainly recognition transcendent in character, so that our idea of them cannot be displayed by the declarative use of the sentence, our knowledge of the semantic content of some constituent expressions, such as 'sneeze' or 'birthday', is easily demonstrable by applying them systematically in various other, effectively

[24] In his review of the various realist attempts at answering Dummett's two arguments against semantic realism, Hale regards this line of reasoning as a potentially fruitful way of responding to the acquisition argument. Hale (1997), 279.

decidable, sentences. Supposing that our idea of the truth conditions of a sentence is determined by our ideas of the (constant) correct declarative use conditions of the sentence's constituents and the way we compose these constituents into a meaningful whole, this demonstration will also amount to a demonstration of some part of our idea of the truth conditions of any sentence composed of these constituents. If this is so, however, then the only thing that a realist needs to show is that her ideas of the recognition transcendent truth conditions of our effectively undecidable sentences, such as 'Caesar sneezed 15 times on his 19th birthday', have no parts that cannot be manifested (and acquired) in this way (i.e. via the application of the relevant sub-sentential constituents in a systematic way in various other, effectively decidable, sentences).

Dummett might think that his opponents will never succeed in showing this, because ideas of recognition transcendent conditions can in principle never be composed of ideas of recognisable conditions. In the case of our previous example, for instance, he may wonder how we could ever manifest our knowledge of the correct declarative use conditions of the name 'Caesar' or the inflection '-d' by applying these constituents in other, effectively decidable, sentences, if the conditions in question are construed along realist lines. Similar questions can be raised regarding the key expressions of other problematic discourses as well. One may wonder, for instance, how we could ever manifest our knowledge of the correct declarative use conditions of expressions such as 'a prime number', 'an experience of a flower', or 'an electron' by applying these constituents in effectively decidable sentences, if the conditions in question are understood in a realist way. Apparently, Dummett may have nothing to object to the idea that we can manifest our knowledge of truth conditions, due to the compositional character of our understanding, by manifesting our knowledge of the correct declarative use conditions of sub-sentential expressions. What he must deny, in defence of the key premises of his arguments against semantic realism, is that the latter conditions could be composed into larger complexes whose obtaining might not be recognised by means of our actual methods and epistemic capacities. In the case of the name 'Caesar', for instance, this means the denial of the realist claim that the correct declarative use conditions of this expression, of which our knowledge can be made manifest by using this term in decidable declarative sentences, can be combined with those of 'sneeze', 'birthday' etc. to generate recognition transcendent truth conditions for the claim 'Caesar sneezed 15 times on his 19th birthday'. Notice, however, that this denial may seem utterly justified if we accept Dummett's core idea that any semantical knowledge that we can communicate by the disciplined declarative use of a piece of representation must be knowledge of conditions that can recognisably obtain at the time of application.

Despite the initial plausibility of this reasoning, Dummett's negative verdict on the realist claim that our knowledge of recognition transcendent truth conditions can be manifested by displaying our understanding of sub-sentential expressions is far from being justified. The fact that the correct declarative use conditions of our semantically basic (i.e. implicitly introduced) representations cannot be fully recognition transcendent, because our ideas of these conditions can be effectively communicated only if we have the opportunity to apply the relevant symbols exactly when these conditions recognisably obtain, by no means implies that the combination of these conditions cannot result in larger complexes whose obtaining can no more be recognised by our actually available methods and epistemic capacities.[25]

[25] Despite its partly epistemological phrasing, the same insight seems to underlie Russell's famous distinction between knowledge by acquaintance and knowledge by description: "The distinction between *acquaintance* and *knowledge about* is the distinction between the things we have presentations of, and the things we only reach by means of denoting phrases. […] In perception we have acquaintance with objects of perception, and in thought we have acquaintance with objects of a more abstract logical character; but we do not necessarily have acquaintance with the objects denoted by phrases composed of words with whose meanings we are acquainted. To take a very important instance: there seems no reason to believe that we are ever acquainted with other people's minds, seeing that these are not directly perceived; hence what we know about them is obtained through denoting.

Consider, for instance, the correct declarative use conditions of our terms about mental states. On a realist construal, these conditions can be specified by reference to those mental characteristics that the terms in question purport to be about. On Dummett's anti-realist construal, in contrast, they must be understood in terms of observable patterns of behaviour. Dummett's reason for this construal is that, apparently, the only conditions whose obtaining can recognisably coincide with the disciplined application of these terms in effectively decidable declarative sentences about human minds are behavioural in character. On a brief reflection upon the case, however, we may realise that Dummett's judgement at this point is premature.

First, there are effectively decidable declarative sentences involving terms about mental states whose disciplined application actually excludes the anti-realist interpretation of these terms in the light of what we know of the semantic content of the other constituents of these sentences. The disciplined use of the expression 'is happy' in effectively decidable declarative sentences, for instance, could be thought to coincide systematically with the obtaining of some recognisable behavioural conditions only if the content of some other terms, such as 'pretends', were different from that which we attribute to them on the basis of their observable declarative application in various other effectively decidable sentences. On the standard interpretation of this term, the disciplined application of the expression 'is happy' in sentences such as 'John merely pretends that he is happy' cannot correlate with the obtaining of those behavioural conditions that we would normally regard as a sufficient observable ground for accepting the sentence 'John is happy', because on the standard interpretation the truth of the former sort of sentences assumes that the behavioural conditions in question obtain, while the sentence 'John is happy' cannot be declaratively applied.

All thinking has to start from acquaintance; but it succeeds in thinking *about* many things with which we have no acquaintance". Russell (1905), 479-480.

Second, there are other, non-behavioural, conditions whose obtaining recognisably and typically coincides with the disciplined use of our terms about mental states in various effectively decidable sentences. These are introspectively or reflectively observable internal states in humans (and maybe in other animals), whose obtaining is (also recognisably, at least in our own case) among the causal antecedents of the individuals' overt behaviour. Of course, many times when our mental terms are declaratively applied we cannot recognise the obtaining of those internal conditions. This, however, does not endanger the successful manifestation of our ideas of them. If they obtain in other people as well, and their obtaining correlates with the obtaining of the same type of behavioural conditions in the case of most human beings, then our knowledge of these conditions can be communicated by the disciplined declarative application of the relevant mental terms, even if the conditions in question obtain within private realms. With some idealisation, the instruction might take the following form.

First, we could turn the attention of our audience to situations in which the intended mental conditions obtain by applying the relevant mental terms in effectively decidable sentences whose correct declarative applicability indeed typically coincides with the obtaining of some publicly recognisable behavioural conditions (e.g. 'John is happy'). Focusing on the recognisable common aspects of these situations, our audience can find in principle two sorts of conditions that might be taken as the intended correct declarative use conditions of our mental terms: Dummett's behavioural conditions, whose obtaining in all these situations is recognisable for everyone, and the intended mental conditions, whose obtaining is recognisable for everyone only in those of these situations in which the mental terms under scrutiny are applied in sentences about the interpreter herself. Having reached this stage, we can eliminate Dummett's candidates by using the same mental terms in sentences whose correct application undermines the anti-realist hypothesis (e.g. 'John merely pretends that he is happy'). In view of this new evidence from our linguistic practice, our audience will ideally understand that our ideas of the correct declarative use conditions of mental terms are to be construed along realist lines.[26]

What the previous method clearly demonstrates is that, contrary to what Dummett presumably supposes, ideas of recognition transcendent truth conditions can be composed of ideas of recognisable correct declarative use conditions. Of course, a proper defence of realism along these lines in the semantics of other problematic discourses, such as those about past events, about the external world, about unobservable entities, about values and normative properties, or about causally inert abstract entities, would require a parallel account of how our knowledge of the respective realistically construed conditions can be manifested in linguistic practice by the disciplined application of the key terms of these discourses in effectively decidable sentences about the designated domains. The crucial tasks to be accomplished, in each case, will be fundamentally the same. On the one hand, we must somehow communicate, by the disciplined use of some effectively decidable sentences, that the correct declarative use conditions of the problematic terms are not anti-realist in character. In the case of our terms about mental states, this task could be accomplished by composing decidable sentences about mental states, which included expressions such as 'pretends'. On the other hand, we must somehow show, again, relying strictly on the disciplined use of effectively decidable sentences, which other conditions we regard as the declarative

use conditions of those terms. In our previous example, this could be done by the declarative use of sentences about mental states exactly when some behavioural conditions obtained, because the obtaining of these conditions, as a matter of fact, typically coincided with the obtaining of some introspectively or reflectively recognisable conditions in the characterised subject's mind.

While the first task can be productively accomplished by composing and endorsing effectively decidable sentences whose truth presupposes, as in the case of 'John merely pretends that he is happy', the contrast between the actually intended and the anti-realistically construed truth conditions of the relevant problematic representations, the achievement of the second aim (i.e. the positive presentation of the actually intended conditions) may require the exploitation of further concept-characterising resources from the instructor. This is mainly because, contrary to the mental case, the disciplined application of our terms about past events, causally effective unobservables or causally inert abstract entities in effectively decidable sentences does not coincide in a directly recognisable manner with the obtaining of any of these intended realist conditions. Accordingly, we cannot hope to manifest our positive ideas of these conditions by appealing to our audience's recognitional abilities at the moment of our instructive utterances, as we could previously, in the case of our terms about mental states.

One alternative resource that I think a realist may rely on in these further occasions is our ability to introduce ideas of not directly observable contents by applying content-subtracting operations on ideas of actually recognisable conditions. In particular, we can characterise our ideas of past, unobservable or abstract conditions by recalling ideas of actually recognisable aspects of the world and then subtracting from them the elements of presentness, observability by actual epistemic

[26] Notice that if the intended mental conditions did not obtain in other people as well, or their obtaining did not correlate with the obtaining of the same behavioural conditions as in our own case, then at the end of the first stage of our instruction our audience would ideally not hesitate, but assume that the correct declarative use conditions of our mental terms are among those behavioural ones that Dummett's anti-realists talked about. In the second stage of the process, however, this hypothesis would be undermined, so at the end our audience would presumably be left with no idea about the semantic content of our mental expressions. The fact that we can understand each other while using this mental vocabulary is, accordingly, supporting evidence for the common belief that other people also possess the intended internal mental states.

capacities, or spatiotemporality, respectively.[27] A sufficient characterisation of these realist notions may, of course, require some further, highly delicate maneuverings with our ideas of actually recognisable conditions, but the conclusion to be drawn is the same: contrary to what Dummett might suppose, the composition of our ideas of actually recognisable aspects of the world may provide us with ideas of conditions whose obtaining or absence is no longer directly recognisable to us. If this is so, however, then the key premises of Dummett's acquisition and manifestation arguments prove to be false: our knowledge of truth conditions whose obtaining or absence cannot be recognised by means of our actual methods and epistemic capacities can nevertheless be communicated by the disciplined use of sub-sentential expressions in effectively decidable declarative sentences.

Summing up, in this section I examined two influential arguments, propounded by Michael Dummett, to the effect that our actual understanding of truth cannot be recognition transcendent in character. First, I provided a brief reconstruction of the arguments, and argued that the most plausible way in which Dummett's opponents might answer these challenges is to query the adequacy of their key premises, and insist that we can acquire and manifest in use knowledge of truth conditions whose obtaining or absence cannot be recognised by means of our actual methods and epistemic capacities. Before providing an outline of why I think a realist may reasonably reject Dummett's key premises, I made two important observations from the perspective of our broader dialectical interest. First, I argued that Dummett's notion of realism does not coincide with the concept that I introduced under the same name earlier in this work. The most important difference between Dummett's construal and mine is that Dummett's anti-realists need not deny the objectivity of truth (i.e. the conceptual possibility of mistaken views about truth). Consequently, the explanatory considerations that I advanced in chapter 4 in support of a general realist construal of truth do not disqualify Dummett's semantic anti-realism from among the viable theories of meaning and truth. Second, I also observed that in the semantics of discourses about the specified recognition transcendent domains the conception that Dummett's anti-realists advocate about truth is clearly non-referentialist in character, in so far as it holds that the truth conditions of the sentences within these discourses are not to be construed in terms of the relevant (recognition transcendent) subject matters. In view of this non-referentialist commitment, I argued that Dummett's anti-realism is not a suitable doctrine for a referentialist to adopt in response to the original or modified and generalised form of Benacerraf's challenge in the semantics of discourses in which we are supposed to acquire knowledge or reliable beliefs about causally inert domains. Having clarified the relation of the two construals, I finally examined Dummett's core motivation for adopting the crucial premises of his arguments, and argued that it is based on a limited view of our capacity to introduce new ideas of truth conditions. Once we realise that, relying on the compositionality of understanding, we can manifest our knowledge of truth conditions by displaying our knowledge of the correct declarative use conditions of sub-sentential expressions, and also that by composing ideas of recognisable conditions we can generate ideas of no longer directly recognisable conditions, we will have no more reason to query the communicability of our realist (or in my terminology: our referentialist) ideas of truth and correct declarative use. In harmony with this result, as I already indicated in chapter 3 and will emphasise in chapter 6 too, my problem with a platonist construal of paradigm *a priori* truths is not that ideas of platonic truth conditions cannot be acquired from, and made manifest in, our publicly observable linguistic practice, but instead that the obtaining (or absence) of such conditions can have, by definition, no impact upon our cognitive processes, so it cannot appear in an explanation of the possibility of *a priori* knowledge either. With these conclusions in mind, I shall turn now to the work of

[27] In chapter 7, I shall provide a more detailed account of how we can generate ideas of abstract (i.e. non-spatiotemporal) objects and properties from earlier acquired notions of observable features of our direct natural environment.

another major opponent of traditional semantic realism, and examine Hilary Putnam's "internal realist" arguments against "metaphysical realism" and the standard realist (viz. correspondence) theory of truth.

2. Putnam: Arguments against Metaphysical Realism

Another highly influential case against the standard realist notion of truth has been put forward by Hilary Putnam in his writings since the mid-70s.[28] Putnam baptises the philosophical position he wishes to attack the perspective of "metaphysical realism" (or "external realism"), a doctrine according to which:

> …the world consists of some fixed totality of mind-independent objects. There is exactly one true and complete description of 'the way the world is'. Truth involves some sort of correspondence relation between words or thought-signs and external things and sets of things.[29]

Instead of this deeply entrenched realist doctrine, Putnam suggests us to adopt an alternative philosophical perspective, which he coins "internal realism". According to this view:

> …*what objects does the world consist of?* is a question that it only makes sense to ask *within* a theory or description […] there is more than one 'true' theory or description of the world. 'Truth […] is some sort of (idealized) rational acceptability – some sort of ideal coherence of our beliefs with each other and with our experiences *as those experiences are themselves represented in our belief system* – and not correspondence with mind-independent or discourse-independent 'states of affairs'.[30]

[28] Classical expositions of the argument can be found in Putnam (1977) and Putnam (1981). Since the latter work provides a more complete formulation of the case, in this section I shall mainly rely on this. For succinct critical discussions of Putnam's argumentation see Lewis (1984), and Hale and Wright (1997b). For a recent defence of internal realism see Forrai (2001).

[29] Putnam (1981), 49.

Putnam's main charge against his opponents is that their view presupposes that there are determinate referential relations between our words and the mind-independent entities of the world, otherwise it would be senseless to maintain that truth is correspondence between endorsed representations and obtaining aspects of the external world, but it does not support any explanation of how these relations could be fixed after all. On his view, the main advantage of his internalist perspective is that it does not require from its proponents a similar account, since it takes reference as a trivial relation between the elements of our conceptual scheme (or their linguistic expressions), on the one hand, and the corresponding objects carved out from reality by this conceptual scheme, on the other. As Putnam formulates it:

> In an internalist view also, signs do not intrinsically correspond to objects, independently of how those signs are employed and by whom. But a sign that is actually employed in a particular way by a particular community of users can correspond to particular objects *within the conceptual scheme of those users*. 'Objects' do not exist independently of conceptual schemes. *We* cut up the world into objects when we introduce one or another scheme of description. Since the objects *and* the signs are alike *internal* to the scheme of description, it is possible to say what matches what. Indeed, it is trivial to say what any word refers to *within* the language the word belongs to, by using the word itself. What does 'rabbit' refer to? Why, to rabbits, of course! What

[30] Putnam (1981), 49-50.

> does ‘extraterrestrial’ refer to? To extraterrestrials (if there are any).[31]

As Bob Hale and Crispin Wright’s critical review nicely illuminates, Putnam’s dialectic can be reconstructed as a conjunction of three sub-arguments. The first is usually called the “permutation argument”, and it purports to show that, in a metaphysical realist perspective, reference cannot be determined by fixing the truth conditions (i.e. the truth values in every possible world) of every syntactically correct sentences in view of all ideally available observational data and theoretical constraints upon this evaluation.[32] The second could be called the “ain’t in the head argument”, and it is designed to show that in a metaphysical realist perspective, reference cannot be determined by our narrow intentional states or anything else obtaining in our head.[33] Finally, the third is sometimes denoted as the “just more theory argument”, and it is meant to show that a metaphysical realist cannot explain how reference is determined by an appeal to a causal (or any other natural) relation between our words (or mental symbols) and the mind-independent world either.[34] Putting together these three sub-arguments, Putnam arrives at the conclusion that a metaphysical realist cannot provide a suitable account of what determines the alleged referential relations between our representations and the objects and properties of the external world.

Before examining the three arguments, it may be worth clarifying how Putnam’s construal of the division between metaphysical realism and internal realism relates to the distinction I introduced earlier between realism and anti-realism about truth. On a first impression, it appears that Putnam’s contrast corresponds more to my distinction than Dummett’s construal did before. First of all, Putnam’s internalist reduces truth to a *prima facie* epistemic property, that of (idealised) rational acceptability, which qualifies her position as anti-realist in the earlier specified sense of the term. Second, she does not query the referentialist idea that the truth conditions of our declarative sentences are to be specified in terms of the relevant intended subject matters. Accordingly, Putnam’s case against metaphysical realism appears as a genuine challenge to realism (rather than, as Dummett’s arguments, merely to referentialism) about truth.

By the epistemisation of the intended subject matter of our beliefs, internal realism seems to provide a suitable referentialist answer to Benacerraf’s (updated and generalised) epistemological challenge in the semantics of our paradigm *a priori* discourses. If the subject matter of our claims about abstract domains is not thought to be platonic (i.e. mind-independent) in character, then a proper explanation of knowledge (or reliable belief formation) within these discourses does not require the existence of an information-conveying mechanism between our minds and a platonic realm.

On a closer look, however, we may also find some discrepancies between the construals just compared. First, contrary to what I supposed of an advocate of realism about truth, Putnam’s opponent is meant to understand truth always in terms of reference. As we have seen, on his view, the existence of a determinate referential relation between our representations, on the one hand, and some elements of the external world, on the other, is a precondition for the existence of truth in any correspondentialist sense of the term. In contrast, on my

[31] Putnam (1981), 52.

[32] The so-called model-theoretic argument advanced in Putnam (1977) is meant to establish the same point. As Hale and Wright rightly observe, however, the model-theoretic resources exploited in this paper are neither necessary nor sufficient for developing the case into a conclusive form. Hale and Wright (1997b), 428-429. Lewis also notes that the “real model theory adds only a couple of footnotes that are not really crucial to the argument”. Lewis (1984), 68. The core idea supported by the argument has previously been defended in Quine (1960b) and Quine (1975) as well. As I shall show in due course, Quine’s famous reinterpretations are less artful, but also less challenging than the ones obtained by Putnam’s permutations.

[33] Putnam’s famous Twin Earth and Brains-in-a-Vat thought experiments are meant to illustrate this second part of his reasoning. The Twin Earth argument first appears in Putnam (1975b), while the core idea of the Brains-in-a-Vat hypothesis can be traced back to Harman (1973).

[34] Putnam (1977), 18, Putnam (1981), 45-46, Putnam (1989).

construal, the precondition is merely the existence of a substantive semantic relation between the above *relata*, which need not involve our conscious referential intentions as a constitutive element. On this view, a condition may have the required semantic relation to a symbol even if while using the latter we never consciously think of the former. It is also important to see, however, that the recognition of this difference does not neutralise Putnam's attack *vis-à-vis* the realist position advocated in this work. Clearly, if Putnam is right in maintaining that his opponent cannot explain how the alleged referential relations could be determined by any means including our conscious referential intentions, then a realist who has to explain the emergence of fixed semantic relations that need not even be referential in character will hardly be able to deliver the required account either. Her dialectical position may be even weaker than that of Putnam's opponent: in those cases in which the truth conditions of our beliefs are supposed to be non-referential in character, her explanation cannot draw on the circumstance that the beliefs in question are (thinly) about the obtaining of some states of affairs.

Beyond observing the previous discrepancy, one may add that the compared construals diverge in a more substantive manner as well, which might even query the significance of Putnam's argumentation to the debate over the correctness of realism about truth as it has been understood in this work. In particular, one may argue that Putnam's idea of internal realism does not render this philosophical perspective an anti-realist position in the sense of the term specified earlier, and the real target of Putnam's dialectic is not the realist construal of truth in general, but merely the traditional correspondentialist version of this doctrine. The proposal seems to be supported by Putnam's terminology as well. Internal realism, one could remark, must be nothing but a specific form of realism. Of course, the crucial motive behind the adoption of this interpretation is not terminological. The suggestion is rather based on some of those formulations that Putnam applies while characterising the perspectives of metaphysical and internal realism. The most

important of them has already been quoted at the beginning of this section. Speaking about the metaphysical commitments of his favoured internalist perspective, Putnam makes the following claims:

> 'Objects' do not exist independently of our conceptual schemes. *We* cut up the world into objects when we introduce one or another scheme of description.[35]

What the proponents of the previous interpretation may find worth noticing in this formulation is that it does not embrace the radical anti-realist idea that the world owes its existence entirely to the human mind. Putnam's conception seems to be rather that the world is somehow there even independently of our thoughts. What he claims is merely that we have good reasons to believe that it is not cut into objects (and properties) before we introduce one or another of our schemes of description.[36] Now if this is true, then, similarly to Dummett's anti-realist, an internal realist may accept the realist notion that the truth conditions of our beliefs, once they are identified by our conceptual schemes, obtain (if they do) independently of what anyone ever believes about this particular circumstance. According to this perspective, our epistemic involvement is confined to the constitution of the truth conditions of our beliefs, so it does not concern the actual obtaining or absence of those conditions in the world thus conceptualised. Clearly, if this interpretation of Putnam's

[35] Putnam (1981), 52.

[36] The clearest example of Putnam's hesitation in relation to the idea of a mind-independent world appears in a paragraph where he discusses Kant's notion of the noumenal world: "Today the notion of a noumenal world is perceived to be an unnecessary metaphysical element in Kant's thought. (But perhaps Kant is right: perhaps we can't help thinking that there is *somehow* a mind-independent 'ground' for our experience even if attempts to talk about it lead at once to nonsense.)" Putnam (1981), 61-62. Later Putnam explicitly endorses the realist bit (Putnam (1988), 114, Putnam (1992), 58, Putnam (1999), 6, 18, fn. 7 on 178), and regrets "having spoken of 'mind dependence' in connection with these issues" in Putnam (1981) (Putnam (1999), fn. 8 on 178).

formulations were correct, then his case against metaphysical realism would not qualify as a challenge to realism about truth in general.

There are three important notes to be made in response to the last observation. First, we must keep in mind that the conception put forward of the nature of *a priori* truth (and the nature of truth in general) in this work is not merely a realist account, but it also includes the idea that truth is a certain correspondence between what is (either in a referentialist or in a non-referentialist sense) endorsed by the declarative use of our truth-apt representations, on the one hand, and what actually obtains in the spatiotemporal world, on the other. Accordingly, Putnam's case against the metaphysical realist perspective would retain its significance to the theory of *a priori* truth advocated in this work, even if the above (realist) interpretation of his ideas proved to be correct.

Second, it may be also worth noticing that if Putnam's internal realism were understood along the previous lines, then his identification of truth with idealised rational acceptability would not amount to the epistemisation of truth and falsity. The move would rather reveal that Putnam's notion of idealised rational acceptability is to be understood in a realist way, standing for a non-epistemic property of some representations. Once the truth conditions of a truth-apt representation were identified by our classificatory scheme, the truth value of this representation would be determined by the world independently of what anyone would ever believe about this circumstance.[37]

In the semantics of our paradigm *a priori* discourses, this moderate realist construal of truth would entail an internalist form of platonism, the idea that the abstract truth conditions of the relevant *a priori* claims are cut out from reality by our prevailing conceptual schemes. Now, one may wonder whether this "internal platonist" construal of our paradigm *a priori* discourses could provide us (and the advocates of referentialism) with a suitable response to Benacerraf's original or modified and generalised epistemological challenge presented in chapter 3. Well, I believe that it cannot. But before showing exactly why, let me advance the third important note indicated above.

To put it briefly, if Putnam's idea is, indeed, that there is a mind-independent world, which has no features to be properly represented independently of the classificatory work of human minds, then he must answer at least two natural questions that may emerge concerning his view. First, one may wonder how an amorphous world could impose any constraint upon human concept formation.[38] If the world has no features (contrasts and similarities) independently of our actual classificatory judgements, then what makes it the case (in that world) that not all classificatory judgements are equally acceptable? Otherwise, if the world does not impose any constraint upon our classificatory work, then what motivates us to believe in its real existence after all? Emphasising the difference between internal realism and radical relativism, Putnam says the following about the existing constraints upon human belief formation:

[37] As Putnam formulates, "the two key ideas of the idealization theory of truth are (1) that truth is independent of justification here and now, but not independent of *all* justification. To claim a statement true is to claim it could be justified. (2) truth is expected to be stable or 'convergent'; if both a statement and its negation could be 'justified', even if conditions were as ideal as one could hope to make them, there is no sense in thinking of the statement as *having* a truth-value". Putnam (1981), 55-56. In a later text, he admits that his formulations in Putnam (1981) were slightly misleading, as they could suggest that he took the idea of idealised rational acceptability (or that of better and worse epistemic situations) more basic than the concept of truth. In fact, the suggestion he wants to make is "that truth and rational acceptability are *interdependent* notions" (i.e. "that the dependence goes both ways: whether an epistemic situation is any good or not typically depends on whether many different statements are *true*"). Putnam (1988), 115.

[38] Putnam himself does not use the term 'amorphous' in his works. What he explicitly denies is the mind-independent identity of objects and properties. Nevertheless, the target of his argumentation is a conception of the world, which supports a correspondence theory of truth. Since the existence of any shape or structure in the mind-independent world is sufficient for the adequacy of such a theory, we have reasons to suppose that according to Putnam's internalist picture the world in itself is not merely void of objects and properties (understood as classes of objects in different possible world), but also entirely shapeless or structureless (i.e. amorphous in the received sense of the term).

> Internalism does not deny that there are experiential *inputs* to knowledge; knowledge is not a story with no constraints except *internal* coherence; but it does deny that there are any inputs *which are not themselves to some extent shaped by our concepts*, […] or any inputs *which admit of only one description, independent of all conceptual choices.* Even our description of our own sensations […] is heavily affected (as are the sensations themselves, for that matter) by a host of conceptual choices.[39]

Notice, however, that these statements do not answer the question formulated above. They only specify the inputs constraining our substantive belief formation, supposing that we already possess a certain conceptual framework, but they do not tell us anything about what makes any of these frameworks (any of the "conceptual choices") more adequate than others.[40] Moreover, the suggested inputs are meant to be experiential, which means that they are conceptualised elements of our experience, rather than emanating from a world that is supposed to exist independently of human minds. So, the idea that they constrain our beliefs about the world does not tell us anything about how the world itself could influence these beliefs.[41]

The second natural question to be answered by Putnam, if the suggested realist interpretation of his position is correct, is inspired by the central realist tenet that the world could exist even if it were not actually thought of by human minds. In principle, every realist must face the question why she believes in the existence of this world and what she thinks about the nature and the emergence of its relation to human minds. Putnam's question to his opponents was more specific. He wanted to hear an account of what could fix the apparent referential relations between our concepts or words, on the one hand, and their subject matter, on the other, if the latter are construed along the metaphysical realist lines. In an internalist perspective, this specific question does not arise, since the purported referents of our representations are not meant to possess identity conditions independently of the adopted conceptual scheme. Nevertheless, if the suggested realist construal of Putnam's internalist position is correct, then the internalist idea that "we cut up the world into objects when we introduce one or another scheme of description" cannot be taken as a suitable account of reference either, unless Putnam explains how he thinks our conceptual schemes can cut up the world into objects and properties.[42]

Both questions concern the (realistically construed) internalist's idea of those relations which are supposed to hold between a (purportedly amorphous) mind-independent world, on the one hand, and human minds, on the other.[43] Clearly, these questions will equally emerge if one classifies part of this mind-

[39] Putnam (1981), 54.

[40] Note that a pragmatist answer, according to which a given scheme of description is superior to another if and only if its adoption is more fruitful than the other's in view of our prevailing cognitive purposes, will not do unless it specifies the relevant purposes as well. If the latter have anything to do with the allegedly amorphous mind-independent world, then it will be hard to see why such a world would favour one scheme to the other. If the purposes in question are, in contrast, specified in mentalist terms (e.g. in terms of predictive success), then the subsequent problem presented in the main text arises.

[41] As Fichte rightly observed, transcendental idealism can be advanced without assuming the existence of a mind-independent world. A more recent example of this idealist perspective on truth, knowledge and existence is Goodman (1954), Goodman (1978).

[42] The same point seems to be recognised by Hale and Wright too. Hale and Wright (1997b), 446.

[43] By the mid-1990s, Putnam also recognised that his internal realist metaphysics, which retained the representationalist idea that the perceptual inputs of our minds "are the outer limit of our cognitive processing" and "everything that lies beyond those inputs is connected to our mental processes only causally, not cognitively", had no proper account of the relation of our knowing minds to the mind-independent world either (Putnam (1999), 12-20). His adoption of what he calls 'natural realism', the doctrine that "successful perception is a *sensing* of aspects of the reality "out there" and not a mere affectation of a person's subjectivity by those aspects", was largely motivated by the conviction that this move can resolve the "antinomy" created by the received representationalist forms of realism. In the concluding part of this section, I shall briefly explain why I think that the adoption of this new philosophical perspective provides no adequate response to the problem of reference either.

independent world as causally inert and abstract in character, and adopts an internalist version of platonism in the semantics of our paradigm *a priori* discourses. So, an advocate of internal platonism must explain not merely how we usually discover what obtains in the alleged platonic realm, but also how the objects and properties of this realm are cut out from the mind-independent world by our prevailing conceptual schemes, and how the amorphous world-part which is supposed to be cut into platonic objects and properties constrain the development of the relevant conceptual schemes. In absence of any interaction between this platonic realm and the rest of the conceptualised world, these further *explananda* render the dialectical position of an internal platonist at least as hopeless as that of her metaphysical realist opponent.

Summing up, we can conclude that Putnam's argumentation is significant to the main concern of the current work. Clearly, it challenges the realist conception put forward earlier of the nature of *a priori* truth (and the nature of truth in general), in so far as this includes the idea that truth is a certain correspondence between what is endorsed by the declarative use of our truth-apt representations, on the one hand, and what actually obtains in the spatiotemporal world, on the other. Since Putnam's internalist construal of truth is referentialist in character, his position could, in principle, also be hoped to support a suitable referentialist answer to Benacerraf's (original or modified and generalised) epistemological challenge in the semantics of our paradigm *a priori* discourses. However, we have also acknowledged that, contrary to what is suggested by its standard classification (i.e. the idea that it includes an anti-realist conception of truth), Putnam's internalist perspective may be compatible with realism about truth, the idea that the truth conditions of our beliefs (once they are identified by our conceptual schemes) obtain (if they do) independently of what anyone ever believes about this circumstance. If the reading which supports this compatibility is correct, then Putnam's argumentation cannot be regarded as challenging more in the semantical account advocated in this work than its commitment to a correspondence theory of truth.

On such a construal, internalism would not provide an escape to the advocates of referentialism from Benacerraf's epistemological challenge in the semantics of our paradigm *a priori* discourses. The idea that abstract objects and properties are cut out from an amorphous mind-independent world by our prevailing schemes of description does not remove the conceptual obstacles from the path of a proper explanation of *a priori* knowledge, and it raises a few specific puzzles concerning the relation of human minds to the mind-independent world. In particular, an internal realist should somehow explain how an amorphous world could impose any constraint upon human concept formation, and how our prevailing conceptual schemes could cut up an independent world into objects and properties.

With these conclusions in mind, we can turn now to the three sub-arguments of Putnam's case against metaphysical realism and the correspondence theory of truth.

As we have seen, the so-called "permutation argument" was designed to show that the referential relation between our representations, on the one hand, and the allegedly mind-independent objects and properties of the world, on the other, cannot be determined by fixing the truth conditions of every syntactically correct sentence in view of all ideally available observational data and theoretical constraints upon this assignment.

How can we fix the truth conditions of our sentences? Well, Putnam's idea is that we can do this by fixing the truth value of the relevant sentences in every possible world.[44] A possible world

[44] One might think that this strategy needs no further justification, since the truth conditions of a certain sentence can be understood as the set of those possible worlds in which the sentence in question is true. In my view, this reasoning is mistaken. According to their received construal, truth conditions are entities that may or may not obtain, among others, in the actual world. In contrast, sets of possible worlds are entities that may or may not contain, among others, the actual world. The two sorts of entities cannot be identified with each other. If Putnam's strategy is correct, then it is correct because our notion of truth conditions and notion of possible worlds are related in a way which ensures a certain one-to-one correspondence between sets of possible worlds, on the one hand, and sets of

is meant to be constituted by those (ontologically basic) conditions which are supposed to obtain in that world.[45] Accordingly, each set of these conditions uniquely determines a possible world. By specifying the truth value of a certain atomic sentence *S* in every possible world, we actually determine the set of those possible worlds in which *S* is true (i.e. in which its truth conditions obtain). Let us call this set *T*. It is clear that by knowing *T* we can develop some idea of the truth conditions of *S*. We will certainly know, for instance, that the conditions in question must be among those conditions which obtain in every possible world within *T*. Otherwise *T* would contain a world in which *S* is false.[46] Let us call the largest set of these conditions *C*. Now, the crucial question is, of course, can our knowledge of *T*

those conditions whose collective obtaining can be necessary and sufficient for the truth of our representations, on the other. It is the existence of such a relation that I shall briefly demonstrate in this paragraph.

[45] For the sake of simplicity, I will grant here that we have an intuitive notion of which conditions the basic constituents of possible worlds are. In possession of our ideas of basic conditions, we can develop the required notion of every possible world. In fact, as I shall point out in the concluding part of this section, our ideas of basic (or any other) conditions (and possible worlds), properly understood, already presuppose the existence of determinate semantic relations between our representations and the represented aspects of the world, and thus cannot serve as explanatory resources in a maximally informative account of reference determination.

[46] Notice that this reasoning presupposes that atomic sentences have no disjunctive truth conditions. If they had, the disjuncts would not obtain in all possible worlds in which these sentences were true. Suppose, for instance, that the sentence ‘Napoleon had a red nose’ is an atomic sentence. One may hold that the truth conditions of this sentence can be only disjunctively characterised: in the simplest case, the sentence is true either if Napoleon’s nose is light red or if Napoleon’s nose is dark red. Clearly, if this construal were correct, then neither of the specified conditions would obtain in every possible world in which the sentence ‘Napoleon had a red nose’ is true. The simplest response to this challenge is to stipulate that for each atomic condition there is an atomic sentence which represents the obtaining of this condition, so that a sentence with disjunctive truth conditions will be always logically equivalent with the disjunction of some atomic sentences. With reference to these logical relations, then, we can separate a set of atomic sentences in the strict sense of the term, whose members no longer possess disjunctive truth conditions. (Thanks to András Simonyi for reminding me of this complication.)

and *C* also help us determine which elements of *C* are the actual truth conditions of *S*? Well, it is easy to show that it can. What has to be realised is merely that *S*’s truth requires the obtaining of every condition within *C*. This already implies that the truth conditions of *S* include all elements of *C*. Suppose that there is a condition in *C*, whose obtaining is not necessary for *S*’s truth. If so, then there must be at least one possible world in which this condition does not hold, whereas *S* is true. This world must, of course, belong to *T*, since *T* contains all worlds in which *S* is true. But then the condition can still not be an element of *C*, since *C* includes only those conditions which obtain in every world within *T*. Our assumption has led to a contradiction. So, we can conclude that *S*’s truth requires the obtaining of every condition in *C*, and thus the elements of *C* can be legitimately taken as the truth conditions of *S*. Since these elements are fixed by the determination of *S*’s truth value in every possible world, we can also assume that, as Putnam suggests, the truth conditions of our sentences can be fixed by the determination of their truth value in every possible world.[47]

Why does Putnam think that fixing the truth value of our sentences in every possible world cannot determine which objects and properties the sub-sentential components of these sentences refer to? Well, his main reason is that, as he claims, the intended referents of our sub-sentential expressions can always be subjected to a permutation which keeps the truth value of every sentence composed of these expressions in every possible world

[47] Note, however, that this method of specifying truth conditions will not work in the case of those representations whose truth or falsity is necessary in character. Take, for instance, our mathematical sentences. Those which are true are true, while those which are false are false in every possible world. Applying the suggested method in this case would imply, among others, that all true mathematical sentences have the same truth conditions. The account of *a priori* truth advocated in this work provides a simple explanation of this delimitation: the truth conditions of our *a priori* claims cannot be specified by designating those possible worlds in which the claims in question are taken to be true, because these conditions are not referential in character, while the possible worlds in which these claims are either equally true or equally false are all meant to be variants of the world that these claims purport to be about.

invariant. He illustrates his point by the following example.[48] Consider the sentence ‘A cat is on a mat’. On the standard interpretation, this sentence is true in exactly those possible worlds in which there is at least one cat on at least one mat at some place and time. Moreover, the previous distribution of truth values is, among others, due to the fact that the expression ‘cat’ refers to cats, while the expression ‘mat’ refers to mats in all possible worlds where there are cats and mats. Now, consider what happens if we alter the referential relations of these two expressions at least in some possible world in the following manner. Take those possible worlds (say, the ‘*A*-worlds’) in which there is at least one cat on at least one mat and there is at least one cherry on at least one tree, and suppose that the expression ‘cat’ refers to cherries, while the expression ‘mat’ to trees in these worlds. In other possible worlds, let the expressions refer to cats and mats, respectively, as before. Does this alteration of the referential relations of these expressions influence the truth conditions of the sentence ‘A cat is on a mat’? Apparently not. For, according to the suggested non-standard interpretation, whenever the expression ‘cat’ refers to cherries while the expression ‘mat’ to trees, there will be at least one cherry on at least one tree, which ensures that the sentence ‘A cat is on a mat’ will be true in these worlds just as much as it was when the two expressions were supposed to refer to their standard referents.

Putnam shows that “a more complicated reinterpretation of this kind can be carried out for all the sentences of a whole language”, and that, consequently, “there are always infinitely many different interpretations of the predicates of a language which assign the ‘correct’ truth-values to the sentences in all possible worlds, *no matter how these ‘correct’ truth-values are singled out*”.[49]

One thing that may disturb some readers in Putnam’s non-standard interpretations is that, contrary to their standard

[48] Putnam (1981), 33-35. I will present the example in a slightly modified form, which nevertheless preserves the author’s original intention.

[49] Putnam (1981), 35.

counterpart, they do not seem to assign the same objects, as referents or extensions, to our expressions in every possible world. As Hale and Wright observe:

> The kind of reinterpretation illustrated by the cats-and-cherries example sustains continuity in truth-value only because it is required to be sensitive to *what is actually the case*: for instance, ‘a cat is on a mat’ is true, under the illustrated reinterpretation […] only because what it says is *constrained to vary* as a function of which [types of possible worlds] the actual world belongs to.[50]

[50] Hale and Wright (1997b), 435-436. According to Hale and Wright, one can always specify some permutation-based reinterpretations that preserve the truth value of our sentences in every possible world and also retain the uniformity of reference across these worlds. As far as I can see, however, the interpretations that these authors have in mind ensure merely the uniformity of the reference of our proper names. The way they construe the referents of our predicate terms entails that these referents cannot be the same in every possible world. To use a very simple example, suppose that there is a limited world in which there are three individuals (*a*, *b*, and *c*) and two properties (*E* and *F*). Our language contains three names (‘John’, ‘Mary’ and ‘Paul’) to name the individuals, and two predicates (‘is plamp’ and ‘is blamp’) to refer to the properties. Suppose that the default interpretation of the language is that ‘John’ refers to *a*, ‘Mary’ refers to *b*, ‘Paul’ refers to *c*, ‘is plamp’ refers to *E* and ‘is blamp’ refers to *F* in every possible world. Under this interpretation, the truth condition of the sentence ‘John is plamp’ is the condition that *a* is *E*, the truth condition of the sentence ‘John is blamp’ is the condition that *a* is *F*, etc. (We could specify these conditions by listing those possible worlds in which the sentences in question are true.) What Hale and Wright suggest is that we can uniformly change the referential relations of our terms without altering the truth conditions of the sentences of our language. We can design, for instance, an alternative interpretation, according to which ‘John’ refers to *b*, ‘Mary’ to *c* and ‘Paul’ to *a* in every possible world, while none of our sentences will change their truth value in any of these words, if we cleverly change the referents of our predicates ‘is plamp’ and ‘is blamp’. Of course, in order to meet the uniformity condition, the new referents of these predicate terms should also be constant in every possible world. The trick applied by Hale and Wright is the following: first, they identify the default referents of the relevant predicates, nominalistically, with their extensions (i.e. with the individuals falling under the predicates); second, they identify the old names of the individuals within these extensions; third, they identify the new referents of those names; and fourth, they stipulate that the latter individuals constitute the new extensions of the predicates.

To this, Putnam may respond that the notion of identity on which the observation is based is perspectival. Against the background of our default construal of identity, the non-standard assignments offered are non-uniform indeed, but if we replace this notion with another, in view of which cherries in *A*-worlds become identical with cats in others while cats in *A*-worlds with cherries in others, then the same non-standard interpretation will turn into one which assigns uniform referents to our expressions while our standard construal will fail to do so.[51]

If we want to grasp the real specificities of Putnam's non-standard interpretations, we must eliminate the previous perspectival elements from our formulations. So, let us recast the point in terms of the relation of truth conditions and assigned referents. Earlier we observed that the truth conditions of a sentence can be always identified with the largest collection *C* of those conditions which obtain in those possible worlds in which the sentence in question is supposed to be true (i.e. in worlds within *T*). This fact is clearly independent of how we construe the identity of those conditions in terms of which we develop our ideas of possible worlds. As we have seen, it is also one that must be acknowledged by Putnam as well. Now, let us distinguish between two sorts of interpretations. Q-interpretations are those which assign to each sub-sentential expression in every possible world a referent that is in some way a component of the conditions in *C*, while P-interpretations are those which assume that some of these expressions in some possible world refer to something that is not a component of the conditions in *C*. An important difference between the two classes is that while interpretations in the Q-class support the idea that the truth conditions of our sentences are compositionally determined by the conditions that their constituents refer to in every possible world, those in the P-class undermine the principle of compositionality in every world in which they assign to some constituent expression something that is not a component of the conditions in *C*.

Relative to our default notion of identity, the standard construal of the sentence 'A cat is on a mat', according to which 'cat' refers to cats and 'mat' refers to mats in every possible world, is clearly a Q-interpretation. The conditions which obtain in exactly those possible worlds in which the sentence 'A cat is on a mat' is true can be specified (among others) in terms of the above referents: (1) there is at least one object which is a cat; (2) there is at least one object which is a mat; and (3) at least one object which is a cat is on at least one which is a mat. We may also add that Quine's famous reinterpretations, which replaced the intended referent of the term 'rabbit' either with three-dimensional spatial cross-sections of four-dimensional space-time rabbits, or with particular exemplifications of the universal *rabbithood*, or with undetached rabbit-parts, are also examples of

Clearly, the method ensures that the truth value of our sentences will be preserved after the reinterpretation in every possible world. Nevertheless, with the extensional construal of properties, Hale and Wright destroy the uniformity of predicate reference under every interpretation. On an extensional construal, for instance, the original referent of our predicate 'is plamp' is the set {*a*} in a world in which only *a* is *E*, and the set {*a*, *b*, *c*} in a world in which *a*, *b* and *c* are equally *E*. This variation in predicate reference is what is preserved by the permutation suggested by Hale and Wright (1997b), 437. Notice that if predicates are supposed to refer to properties, rather than to individuals possessing those properties, then the referent of our predicate 'is plamp' will prove to be uniform according to its standard interpretation (viz. *E* in every possible world), but also lose its coherent interpretability after the execution of Hale and Wright's permutations.

[51] The idea shows up in Putnam's reasoning against the charge that his non-standard interpretations assign referents to our expressions on the basis of their (the objects') extrinsic, rather than intrinsic, properties. As he formulates the core point of the charge, "[i]n the actual world, every cherry is a cat*; but it would not be a cat*, even though its intrinsic properties would be exactly the same, if no cherry were on any tree". And his reply: "The upshot is that viewed from the perspective of a language which takes 'cat*', 'mat*', etc., as primitive properties, it is 'cat' and 'mat' that refer to 'extrinsic' properties […] while relative to 'normal' language, language that takes 'cat' and 'mat' to refer to cathood and mathood […] it is 'cat*' and 'mat*' that refer to 'extrinsic' properties. Better put, being 'intrinsic' and 'extrinsic' are relative to a choice of which properties one takes as *basic*; no property is intrinsic or extrinsic in itself". Putnam (1981), 37-38. Notice, however, that in other passages Putnam's formulations suggest that he would not subscribe to this interpretative manoeuvre. As I shall point out in fn. 59 below, his brain-in-a-vat thought experiment actually assumes that (in a metaphysical realist perspective) identity across possible worlds is fixed before the beginning of referent assignment.

Q-interpretations.[52] The non-standard referents that they assign to our sub-sentential expressions do not cease to be a component of the truth conditions of those sentences that are composed of these expressions.

In contrast, the permutations underlying Putnam's reinterpretations do not observe the previous constraint upon referent assignment, since in some possible world they assign a referent to some expressions, which are not a component of the truth conditions of those sentences that the expressions in question figure in. In the case of our particular example, Putnam's interpretation assumes that in *A*-worlds the term 'cat' refers to cherries, objects that, on our default notion of identity, are not a component of the truth conditions of the sentence 'A cat is on a mat' (or any other sentence including the term). The objection that in *A*-worlds the relevant truth conditions are not those specified in the previous paragraph, but instead some others including the presence of at least one cherry on at least one tree is, in view of our previously chosen notion of identity, a non-starter. If the objection were correct, then the latter conditions should obtain, as we have seen, in all possible worlds in which the sentence in question is true. But this cannot be the case. In at least some possible worlds that are not *A*-worlds the sentence 'A cat is on a mat' is true even if in that world there is no cherry on any tree whatsoever. Putting it briefly, relative to our default notion of identity, Putnam's permutations fall into the category of P-interpretations.

Of course, by replacing this default notion with some other construal of identity, Putnam could save any of his alternatives from falling into the category of P-interpretations, but the replacement would help him out only because it would also change the truth conditions of the sentences under scrutiny. In the case of his particular example, for instance, the suggested reinterpretation could qualify as uniform only if cherries in *A*-worlds were considered to be identical with cats in others. On such a construal, however, the *A*-world manifestation of those conditions which obtain in exactly those possible worlds in which the sentence 'A cat is on a mat' is true would be different from what is implied by our default interpretation. If Putnam subscribes to the applicability of this queer interpretative manoeuvre, then he can no longer consistently maintain his initial assumption that the truth conditions of our sentences can be fully determined by the specification of the truth value of these sentences in every conceivable real-world situation. Another important observation concerning this move is that no notion of identity will save for him more than one of these assignments. So, even if we acknowledge that none of his alternatives is a P-interpretation in an absolute sense, nevertheless we can maintain that, disregarding some specific cases (including Quine's non-standard construals), Putnam's permutations cannot provide us with alternative interpretations that would equally support, on a given construal of identity, the semantic principle of compositionality.

Summing up, Putnam's permutation argument demonstrates that the correct declarative use of our sentences (i.e., ideally, the specification of the truth value of our sentences in every possible world) does not fix the reference of the sub-sentential expressions of these sentences, supposing that one of the following two conditions obtains: (1) the specification of the truth values in question does not fix the truth conditions of the sentences either; (2) the referents assigned by the suggested non-standard interpretations need not be uniform, and a component of the truth conditions of the sentence they figure in, in every possible world. From a metaphysical realist perspective, the obtaining of (1) is certainly counterintuitive, in so far as identity is meant to be an objective feature of the basic constituents of the world. To be sure, a metaphysical realist need not deny that the actual referents of our expressions are partly created by our classificatory scheme. Some people may perceive and conceptually distinguish features in the world that others take to be identical. Contingent cultural, biological and psychological facts, pragmatic considerations, and even arbitrary conventions

[52] Quine (1960b).

can have an impact on what we regard as different temporal parts of the same individual or different spatiotemporal instantiations of the same universal property even in a metaphysical realist perspective. What the advocates of this perspective maintain is merely that the process of conceptualisation is essentially constrained by the mind-independent similarities and contrasts of the conceptually associated aspects of the world.[53] Due to these "an sich" characteristics, for a metaphysical realist, it is simply a mistake to consider cherries and cats in *A*-worlds to be identical with cats and cherries, respectively, in all others. Nevertheless, a metaphysical realist must also admit that her idea of scheme-independent similarities and contrasts in the world is based on explanatory considerations over and above her knowledge of the declarative applicability (i.e. the truth value) of our sentences in various possible worlds. Assuming (2) may seem a mistake for an even wider audience, but the conclusion to be drawn concerning this assumption is basically the same: even if we suppose that the referents of our sub-sentential expressions are always a component of the truth conditions of the sentences built from these components, we cannot base this assumption merely on what we know of the truth value of our sentences in various possible worlds. So, whether or not one accepts Putnam's non-standard interpretations as genuine alternatives, the permutation argument seems to achieve its dialectical aim: it reveals that if we adopt a metaphysical realist conception of various aspects of the world, then the referential relations of our representations to these aspects cannot be fully determined by fixing the truth value of our sentences in every possible world. What the argument shows is that for each consistent truth value assignment there will be more than one correspondence between our representations and the aspects of the real world. If this is so, however, then an advocate of metaphysical realism or a correspondence theory of truth cannot explain in the suggested way how we can unambiguously think of and speak about the intended aspects of the world.

Another option to explain how referential relations between representations and the represented aspects of a purportedly mind-independent world are fixed is to hold that what determines these relations is to be found in our head. In particular, it may be thought that a mental or linguistic symbol is standing for some aspects in the world, because of the way we actually think about, and intend to refer to, these external entities.[54] Putnam's second argument, which I called the "ain't in the head" argument, purports to show that this type of account of reference determination also fails to be adequate: arguably, our thoughts, beliefs, referential intentions and other intentional states cannot fully determine the referential relations of our representations to the intended aspects of a mind-independent world either.

Putnam advances various examples that are meant to illustrate his general point. First, he observes that ordinary indexical words, such as 'I', 'this', 'here' and 'now', are trivial counterexamples to the assumption that our mental states can fix the reference of our mental and linguistic representations:

[53] What the metaphysical realist is supposed to deny is not that there are alternative correct characterisations of the world, but merely that there is more than one true *and complete* description of the way the world is. Putnam does not seem to appreciate enough the difference between the two statements. Putnam (1981), 49. The association of metaphysical realism with the idea that "there is one definite totality of objects that can be classified and one definite totality of all properties" occurs in Putnam's more recent writings as well. Putnam (1999), 7. Notice, however, that the latter claim is more specific and indeed less plausible than the other core tenet typically associated with metaphysical realism, according to which some identities and contrasts in the world are independent of our prevailing classificatory scheme, and thus there is an absolutely good sense in which our true representations can be said to correspond to something in the mind-independent world. In this section, I shall suppose that Putnam's argumentation attacks this minimal tenet of metaphysical realism. See also fn. 82 below.

[54] The most typical argument for the claim that narrow mental states determine reference runs from the premises that (1) knowing the sense of a certain expression is being in a narrow mental state and that (2) sense determines reference.

> I may be in the same mental state as Henry when I think 'I am late to work' […] and yet the token of the word 'I' that occurs in my thought refers to me and the token of the word 'I' that occurs in Henry's thought refers to *Henry*.[55]

Putnam's claim is not necessarily that his and Henry's *global* mental states are the same when they entertain the thought 'I am late to work' applying the indexical element to themselves, although he does believe that global mental states, narrowly understood, cannot determine reference either.[56] His point here is rather that shared ideas of the referential power of indexicals, in themselves, cannot fully determine the referent of these pieces of representation. The case, however, is far from being a trivial falsifier of the assumption that our intentional states can determine the referential relations between our representations and the intended aspects of a mind-independent world. In particular, mostly, it is far from obvious that our referential intentions accompanying the correct meaningful application of an indexical concept or term in various contexts entirely coincide. Clearly, they must have some common characteristics, with reference to which they can be regarded as falling into the same kind. For instance, we may observe that all subjects who apply the indexical term 'I' in standard sentential contexts and situations intend to refer to themselves. This fact, however, does not entail that there are no differences between these intentions which could in principle explain the difference between the actual referents of these terms in the relevant contexts. For instance, it seems rather plausible to maintain that Putnam's and Henry's referential intentions while entertaining the indexical thought mentioned in the previous example are, at least in normal circumstances, never entirely the same: while Putnam's intention is to refer to Putnam (rather than to Henry), Henry's is to refer to Henry (rather than to Putnam). As we shall see, referring to the case of indexicals is rather unfortunate also because Putnam's reasons for maintaining that nothing in our head can determine the referential relations of our representations to the represented aspects of the world have nothing to do with the specific semantic features of indexical concepts and expressions.

We can find out more about what Putnam might have in mind by considering his second illustrative case, commonly known as the Twin Earth thought experiment:

> Twin Earth is very much like Earth […] Suppose […] that there are English speakers on Twin Earth (by a kind of miraculous accident they just evolved resembling us and speaking a language which is, apart from a difference I am about to mention, identical with English as it was a couple of hundred years ago). I will assume these people do not yet have a knowledge of Daltonian or post-Daltonian chemistry. So, in particular, they don't have available such notions as 'H_2O'. Suppose, now, that the rivers and lakes on Twin Earth are filled with a liquid that superficially resembles water, but which is *not* H_2O. Then the word 'water' as used on Twin Earth refers *not* to water but to this other liquid (say, XYZ). Yet there is no relevant difference in the mental state of Twin Earth speakers and speakers on Earth (in, say, 1750) which could account for this difference in reference. The reference is different because the *stuff* is different. The mental state by itself, in isolation from the whole situation, does not fix the reference.[57]

This time, the crucial point emphasised by Putnam is that the referent or extension of some representations is not fully determined by what we actually know or stipulate of the intended entities at a certain time. The concepts and terms whose

[55] Putnam (1981), 22.

[56] Putnam (1981), 22., fn. 1.

[57] Putnam (1981), 19 and 22-23.

referential properties the example explores are our representations of natural kinds. What this second example reveals is that Putnam's negative tenet is intimately related to his observation that in a metaphysical realist perspective the semantic relations under scrutiny often transcend our actual conceptions of the intended subject matters. Notice, however, that in the case of our natural kind expressions this transcendence is permitted by our referential intentions themselves. While speaking about water, Twin Earth speakers as well as speakers on Earth before 1750 intend to refer to something whose nature is not yet discovered by contemporary natural sciences. The fact that the above speakers refer to different stuffs in their natural environment is partly due to their referential intentions. The existence of such intentions may, of course, undermine the radical claim that our beliefs and referential intentions alone determine the reference of any actual piece of representation. Putnam's negative tenet, however, assumes much more than this. It claims that (if metaphysical realism is true, then) our beliefs and referential intentions *cannot* determine the reference of *any* kind of representation purportedly referring to an aspect of the external world. The above illustration fails to support the modal aspect as well as the universal scope of this claim. A metaphysical realist could accept everything that is stated in the thought experiment and still believe that with the growth of knowledge we can sharpen our referential intentions such that there remains no room for our environment to influence the extension of our terms, and also that many of our actual representations are already used with such intensions.

Putnam's third illustrative case in support of his claim that reference cannot be fully determined by anything in the head purports to show that the above gap between our actual intentional states, on the one hand, and the alleged referential relations between our representations and the represented aspects of a mind-independent world, on the other, is present necessarily and not only in the case of our natural kind expressions, but generally in the case of any representation that is supposed to stand for something in a "ready-made" world. The example in

question is commonly known as the brain-in-a-vat thought experiment:

> [I]magine that a human being […] has been subjected to an operation by an evil scientist. The person's brain […] has been removed from the body and placed in a vat of nutrients which keeps the brain alive. The nerve endings have been connected to a super-scientific computer which causes the person whose brain it is to have the illusion that everything is perfectly normal. There seem to be people, objects, the sky, etc; but really all the person […] is experiencing is the result of electronic impulses travelling from the computer to the nerve endings. The computer is so clever that if the person tries to raise his hand, the feedback from the computer will cause him to 'see' and 'feel' the hand being raised. Moreover […] the evil scientist can cause the victim to 'experience' (or hallucinate) any situation or environment the evil scientist wishes. […] It can even seem to the victim that he is sitting and reading these very words about the amusing but quite absurd supposition that there is an evil scientist who removes people's brains from their bodies and places them in a vat of nutrients which keep the brains alive.[58]

The first thing that Putnam observes concerning the above scenario is that its obtaining is fully compatible with the laws of nature, as far as we know them, in the actual world. The second is that despite the "physical possibility" of the described situation, a metaphysical realist could never consistently state that she is a brain in a vat. This is because her perspective actually implies that she is either not a brain in a vat or she is, but then her thought that she is is not about the intended scenario. Rather it is about those aspects of the external world that stand in the same kind of

[58] Putnam (1981), 5-6.

relation with her ideas in her actual world as our concepts and words are supposed to stand with their referents in a "normal" physical environment.[59] The fact that the narrow mental states of the envatted brain are, *ex hypothesi*, qualitatively indistinguishable from those of another in a normal human head does not guarantee that the corresponding ideas of these brains refer to the same objects and properties in the external world. As Putnam rightly observes, the intrinsic qualitative features of our representations do not guarantee the representational aspects of these entities. If they did, then a perfect copy of a caricature of Winston Churchill accidentally created by an ant crawling on a patch of sand would amount to a genuine representation of Churchill independently of anyone's taking it to be so.[60] Clearly, in a metaphysical realist perspective, the referent of a certain piece of representation is singled out by a substantive relation between some aspects of the world and the latter, rather than by some intrinsic property of the latter. What Putnam's thought experiment shows us is that if we adopt his opponent's perspective, and suppose that our concepts and words refer to some aspects of a ready-made world in virtue of a specific (causal) relation between the former and the latter, then the very same sceptical considerations that can be used to challenge the metaphysical realist's belief in the possibility of knowledge undermine the assumption that our narrow intentional states can fix the reference of our representations as well. In so far as the relevant external correlates of our (fixed) narrow mental (or neural) states can so radically vary as in Putnam's thought experiment, a metaphysical realist would hardly be correct in assuming that any of these narrow states can ever determine what our representations refer to in a mind-independent world.

[59] Elaborating on the consequences of his opponent's perspective, Putnam sometimes argues that due to the presented semantical problems, in a ready-made world "it is not possible after all that we are Brains in a Vat". Putnam (1981), 51. As far as I can see, the conclusion is a *non-sequitur* and not even needed for Putnam's case against metaphysical realism. Another point to be noted here, already touched upon in fn. 51 above, is that Putnam's thought experiment presupposes that the referents of the corresponding concepts of a normal and an envatted brain are objectively different. Trees and birds in normal possible worlds cannot be identified with particular programmes or electronic signals in those inhabited with envatted brains. This may suggest that Putnam would not opt for the indicated queer reinterpretation of the notion of identity in order to save the principle of compositionality in his permutation argument either. Due to this resistance, he cannot help but accept that the non-standard interpretations he has in mind fail to observe the principle that the referents of the basic constituents of a truth-apt representation must be the components of the truth conditions of this representation.

[60] Putnam (1981), 1-2 and 12-13.

In view of the previous illustration, Putnam's argument against the idea that reference can be determined by the way we actually think about, or intend to refer to, the entities of a mind-independent world can be reconstructed in the following way:

1. According to metaphysical realism, reference is a substantive relation between representations and (potentially obtaining) represented aspects of the world.
2. Intentional states narrowly understood are identified with reference to their intrinsic properties, rather than to their substantive relations to (potentially obtaining) aspects of the world.[61]
3. Consequently, reference construed along the metaphysical realist line cannot be fixed by intentional states narrowly understood.

The argument seems trivially correct. It relies on the truism that the intrinsic properties of an entity, in themselves, cannot account for its relations to others. Clearly, this truism applies to everything, including every representation, whether indexical, standing for a natural kind or any other feature of the world. If a

[61] Intentional states can be identified by reference to their external semantic content, their substantive relation to some aspects of the world as well. As Putnam rightly observes, however, explaining reference in terms of such "impure" mental states would be circular, since being in such states presupposes reference as an integral component. Putnam (1981), 41-43.

metaphysical realist wants to explain what fixes the referential relation of our representations to the (potentionally obtaining) represented aspects of the world, then she must come up with a story which invokes both *relata* and some facts that establish the relevant relation between them.

As we have noted, Putnam's brain-in-a-vat thought experiment draws heavily on the idea that the required facts must involve a suitable sort of causal link between our narrow intentional states, on the one hand, and the represented aspects of the world, on the other. The reason for which we are supposed to accept that brains in a vat cannot think of their own predicament is that the latter circumstance is not among the distinctive causal antecedents of their (narrow) thought that they are brains in a vat.[62] Accordingly, the most natural response from the advocates of metaphysical realism to Putnam's arguments that neither our knowledge of the truth value of our sentences in every possible world, nor our narrow intentional states can determine the reference of our concepts or thoughts and their linguistic expressions is that a proper account of this semantic relation must invoke a suitable causal link between the former representations and the entities that they stand for in a mind-independent world.

Putnam's third argument, sometimes denoted as the "just more theory" argument, purports to show that a metaphysical realist cannot explain how reference is determined by appealing to a causal (or any other natural) relation between our words (or mental symbols) and the mind-independent world either.[63] The argument can be best reconstructed from the following lines:

> Suppose there is a possible naturalistic or physicalistic *definition* of reference, as Field contends. Suppose
>
> *x refers to y* if and only if *x bears* R *to y*
>
> is true, where *R* is a relation definable in natural science vocabulary without using any semantical notions (i.e. without using 'refers' or any other words which would make the definition immediately circular). If (1) is true and empirically verifiable, then (1) is a sentence which is itself true even on the theory that reference is fixed as far as (and *only* as far as) is determined by operational *plus* theoretical constraints. […] If reference is only determined by operational and theoretical constraints, however, then the reference of '*x* bears *R* to *y*' is *itself* indeterminate, and so knowing that (1) is true will not help. Each admissible model of our object language will correspond to a model of our meta-language in which (1) holds; the interpretation of '*x* bears *R* to *y*' will fix the interpretation of '*x* refers to *y*'. But this will only be a relation *in each admissible model*; it will not serve to cut down the number of admissible models at all.[64]

As we can see, Putnam's strategy is to apply the result of his permutation argument to the suggested naturalistic account of reference itself, and conclude that the account cannot unambiguously tell us which naturalistic relation reference consists in, since it is just more theory, whose referential power is

[62] In some places, Putnam argues that due to the absence of the required link between the words of a brain in a vat, on the one hand, and the intended referents of these words in normal circumstances, on the other, "Brain-in-a-Vat Worlders cannot refer to anything external at all" (Putnam (1981), 10, 13). In other paragraphs, he concedes that due to the obtaining of the "close causal connection" between the use of those words, on the one hand, and the electronic impulses of the computer causing the envatted subjects' experience, or the programme responsible for these impulses, on the other, brains in a vat can still refer to external things, even if not to those that we are supposed to refer to in our normal physical environment (Putnam (1981), 14-15.). The common element of these reasonings is that the relation which determines what our representations refer to is a causal link between our narrow representational states, on the one hand, and the represented aspects of the world, on the other.

[63] The naturalistic account that Putnam addresses here has been put forward, among others, by Field (1972), Evans (1973) and Devitt (1981).

[64] Putnam (1981), 45-46.

limited the same way as that of our other (non-semantical) theories of the world.

The immediate response a metaphysical realist may give to this argument is that it conflates the original question of what fixes reference with another of whether we can unambiguously specify those factors or mechanisms which contribute to reference determination. The charge is, however, clearly illegitimate. Putnam is, apparently, fully aware of the conceptual difference between the above questions.[65] His point is not that the two questions are the same, but rather that they cannot be answered independently of each other.

A better response to the argument is to ask why a metaphysical realist could not assume, while giving her account of reference determination, that she can determinately refer to the intended aspects of a mind-independent world.[66] As Hale and Wright aptly formulated it:

> The metaphysical realist […] takes up the challenge to say what constitutes determinate relations of reference, only to find that no sooner has he opened his mouth than Putnam gags him with the complaint that he has no right to assume any of his words to be determinate in reference. The resulting situation is therefore really no different from that generated by the boring and jejune variety of meaning-scepticism which challenges an opponent to explain how meaningful discourse is possible, but won't countenance attempted answers because to presume them meaningful is to beg the question against it. Obviously the metaphysical realist has to be presumed capable of contentful – so, determinately referential – speech if he is to respond to Putnam's challenge, or indeed to any challenge at all. The onus legitimately placed upon him is not to *demonstrate that* determinate reference is possible, but to provide a constitutive account which *explains how* determinate reference works. Accordingly, he is perfectly within his rights to assume, at least pro tem, a meta-language in which a determinate account of the putative mechanics can in principle be given.[67]

Prima facie the objection seems fair. There is no point to ask for an account of a certain phenomenon (in this case: determinate reference), if we *ab ovo* reject the possibility of any determinate thought of the realm including the *explanandum*.

Notice, however, that Putnam's dialectic is a bit more deliberate than that which Hale and Wright attribute to him. Most importantly, his claim that his opponent cannot determinately think of what she intends to invoke in her account of reference is not a groundless premise in his argument. Rather, it is based on what he thinks this opponent can legitimately hold of the relation of the intended *explanans* (viz. the causal relations between representational states and intended referents) to our actual representation of it. His reasoning can be reconstructed in the following way:

1. The reference of our notion of causal relations cannot be determined more tightly than that of our concepts of other aspects of the mind-independent world.
2. Our concepts of aspects of the mind-independent world are developed together with our empirical theories of that world, and any external factor that contributes to the determination of the reference of these concepts must exert its influence through those operational and theoretical constraints that we observe in the course of this empirical theory formation.

[65] Putnam (1981), 46.

[66] The point was stressed by Lewis (1984).

[67] Hale and Wright (1997b), 441.

3. The operational and theoretical constraints that we observe in the course of empirical theory formation are not sufficient for determining the reference of our concepts of aspects of the mind-independent world (cf. the permutation argument).
4. So, the reference of our concepts of aspects of the mind-independent world cannot be fully determined after all.
5. So, the reference of our notion of causal relations cannot be fully determined either.

Beyond the lines already quoted to introduce his argument, there are various other passages in Putnam's work which support the above reconstruction. I shall recall here only two of them:

> …let us consider the view that (1) [the sentence '*x refers to y* if and only if *x bears R to y*' – Zs. N.], understood as Field wants us to understand it (as describing the determinate, unique relation between words and their referents), is true. If (1) is true, so understood, what *makes* it true? Given that there are many 'correspondences' between words and things, even many that satisfy our constraints, what *singles out* one particular correspondence *R*? Not the empirical correctness of (1); for that is a matter of our operational and theoretical constraints. Not, as we have seen, our intentions (rather *R* enters into determining what our intentions signify). It seems as if the fact that *R is* reference must be a *metaphysically unexplainable* fact, a kind of primitive, surd, metaphysical truth.[68]

And somewhat later:

[68] Putnam (1981), 46.

> To me, believing that some correspondence intrinsically just *is* reference (not as a result of our operational and theoretical constraints, or our intentions, but as an *ultimate* metaphysical fact) amounts to a magical theory of reference.[69]

Now, if we look at the above reconstruction, we can easily realise that premise (2) is an explicit denial of what the advocate of a causal/naturalistic account of reference suggests in response to Putnam's question about reference determination. The dialectically problematic element in Putnam's reasoning is, accordingly, not so much that he queries his opponent's capacity to determinately think of what she intends to invoke in her account of reference, but instead that he does not believe that there could be anything else to be invoked in such an account beyond those operational and theoretical constraints that we observe in the course of empirical theory and concept formation. Once one adopts premise (2), there is no longer reason to examine any further account of reference determination, and Putnam's just more theory argument also survives. On the other hand, once one rejects premise (2), there remains no conclusive reason to query the naturalist's capacity to determinately refer, so Putnam's just more theory argument collapses as well.

The main question, accordingly, to be answered before assessing Putnam's third argument is whether there are good reasons for us to adopt premise (2). From the passages just quoted it seems that Putnam's primary problem with a factor whose influence on reference determination would transcend the effects of those referred to in premise (2) is that the contribution of such a factor would be a "surd" metaphysical fact, something which, *per definitionem*, would not affect our minds through the epistemically internal constraints mentioned in premise (2). What Putnam seems to insist on here is the commonly accepted methodological principle that an explanation must not invoke any

[69] Putnam (1981), 47.

condition whose obtaining cannot be detected, at least ideally, by human minds. Clearly, the relations that Putnam's naturalist opponents invoke in their account of reference determination, as external factors in general, are "surd metaphysical" or "magical" elements *from the first-personal, transcendental perspective of a conscious mind*, in so far as their obtaining and effect in the case of the subject's own mental representations do not impose any extra constraint upon her theory and concept formation. It is exactly this circumstance which enables the sceptic to advance her standard challenges to the subject's ordinary knowledge claims of the external world, and it is this magical aspect of the suggested naturalist account which seems to motivate Putnam's verdict that it does not qualify as a suitable response to his antinomy of reference in a metaphysical realist perspective.

Notice, however, that the opponent has a relatively plausible rebuttal. Namely, she can remind us that the suggested external conditions, whose obtaining and contribution to the determination of the reference of our own representations cannot impose any extra constraint upon our own theory and concept formation, are clearly detectable for us from a third-personal, empirical perspective, when we study the referential relations of other subjects' representations to the aspects of an apparently mind-independent world. It is this third-personal, empirical content of our notion of causal relations (and other elements of the external world) which enables the naturalist to reject Putnam's verdict about the surd metaphysical or magical character of her causal account of reference.[70]

Of course, the plausibility of this rebuttal depends heavily on whether we accept the naturalist assumption that the relation of our own mind to the world can be reasonably characterised by the same (operationally and theoretically constrained) vocabulary and account as the relation of other minds to their mind-independent environment. By adopting this assumption, Putnam's naturalist opponent may preserve the right to reject premise (2), and therewith Putnam's just more theory argument as well. The only thing that she must show, in order to convince us that her account cannot be dismissed merely by reference to the result of Putnam's permutation argument, is that the empirical content of our notion of causation is more specific than that of our notion of correspondence.

In view of these results, we can sum up now Putnam's argumentation against metaphysical realism and the correspondence theory of truth. His main charge against his opponents, as we saw, is that they assume that there are determinate referential relations between our representations and various aspects of a mind-independent world, but they cannot provide an explanation of how these relations could be fixed among these *relata*. His ain't in the head argument is meant to pin down that a proper account of reference cannot rely merely on what obtains in our heads. Rather, it must invoke both of the above *relata* and some facts that are responsible for the emergence of the relevant semantic relations between them. On Putnam's view, however, the only external facts that a metaphysical realist can legitimately and plausibly invoke in her account of reference determination are those correlations that obtain between the operationally and theoretically adequate use of our representations, on the one hand, and the obtaining of various aspects of the mind-independent world, on the other. His permutation argument is finally designed to show that if a certain linguistic or cognitive practice has a coherent interpretation, then it has many other coherent interpretations as well. In other terms, even if we observe all operational and theoretical constraints in the course of the application of our truth-apt representations, this practice will still not fully determine which aspects of the world the basic constituents of these representations refer to. The conclusion of this line of thought is, indeed, that a metaphysical realist cannot account for the phenomenon of determinate reference.

We saw that the first part of this reasoning, which purports to establish that reference cannot be fixed by facts within our

[70] As Putnam's brain-in-a-vat thought experiment nicely illuminates, however, the empirical contents in question cannot exclude the conceivability of the traditional sceptical scenarios.

heads alone, is obviously sound. The correctness of this claim, however, does not entail that our narrow intentional states do not contribute to the determination of the referential content of our representations at all. The condition, for instance, that our understanding is compositional in character may clearly be taken to obtain in our heads, and, as we have seen, its presence may place essential constraints upon the admissible interpretations of our mental and physical representations. In particular, the referents of some component representations cannot fail to be a component of the referent of others that are composed of the former components. If the truth conditions of a certain truth-apt representation are construed in referentialist terms, then the previous principle implies that the referents of the relevant component representations cannot fail to be a component of the truth conditions of those truth-apt representations that are composed of the former components.[71]

The ways in which the applications of our mental or physical symbols are related to each other as well as to the occurrences of some elements in our own subjective experience or imagination are also (partly) determined by facts within our heads, and the resulting relations clearly contribute to the determination of these symbols' referential links to various aspects of the external world. The referents of our analytically related concepts or expressions, for instance, cannot be independent of each other, while the occurrence of the referent of a representation with empirical content cannot fail to be reliably indicated by the presence of the designated experiential features when we are supposed to perceive this referent.[72] That the referents of our empirical

representations are (mostly) to be found in the external world, rather than in our own subjective experience or narrowly understood conscious mind seems also to be guaranteed (at least partly) by facts within our heads. Further, as we have seen, if we suppose with a metaphysical realist that sameness, difference, or degrees of similarity are not merely the products of our classificatory work, but in many cases objective features of entities in a mind-independent world, then the fact that our representations are meant to stand for the same sorts of things (or, in the case of indexicals, for the values of the same reference functions) in every possible world and in the actual world in every particular context of application also reduces the number of their admissible interpretations.[73] Finally, in possession of a basic set of symbols, whose referential relations to certain aspects of the external world have been successfully established before, we can introduce some new concepts and terms which may stand for actually non-existing objects and properties as well. The referents of these analytically introduced representations are also (partly) determined by facts within our heads.[74]

The second step of Putnam's line of thought, the assumption that the only external facts a metaphysical realist can legitimately and plausibly invoke in her account of reference determination are those correspondences that obtain between the operationally and theoretically adequate use of our representations, on the one hand, and the obtaining of various aspects of the mind-independent world, on the other, is more problematic than the first. As we saw, the plausibility of this

[71] As we noted, if compositionality holds, then the referents of our concept of cat in Putnam's *A*-words cannot be cherries, unless cherries in these worlds can legitimately be identified with cats in others.

[72] If we reject the idea that we can directly grasp the occurrence (or constant existence) of some external entities by our mind, as Putnam certainly did in his internal realist period, then we can extend the scope of our previous formulation from empirical representations to all representations of the external world. Putnam's anti-Fregean, representationalist conception of acquaintance with external objects is manifest in Putnam (1981), 27. Later he famously gave up this view (as well as his entire internal realist perspective), and adopted a "natural realist" account of perception and the relation of the mind to the world. Putnam (1999).

[73] Again, if sameness is an objective feature of entities in a mind-independent world, then our referential intentions guarantee that our concept of cat refers to the same sort of things (presumably to cats) in Putnam's *A*-worlds as in other possible worlds.

[74] As I shall show in chapter 7, our ideas of abstract (i.e. non-spatiotemporal) objects and properties arguably acquire their referential content in this indirect manner, due to some facts or events (among others) within our heads.

restriction hinges largely upon whether or not we accept the naturalist assumption that the referential relation of our representations (or narrow intentional states) to the intended aspects of the world can be characterised by one and the same unified vocabulary and account, independently of whether the former *relata* occur in our own mind or in the mind of some others. If Putnam's naturalist opponent can tell us why she thinks we should adopt the above principle, and then specify what observable characteristics a relation between the two types of entities must display in order to qualify as a causal relation, as opposed to mere correspondence, then she can neutralise this step of the argumentation, and guarantee the immunity of her account from the negative consequences of the third stage of Putnam's reasoning.

The idea that the relation of our own mind to the external world can be characterised by the same account as the relation of other minds to their environment seems to be based on empirical evidence and some fallible explanatory considerations, which fit into a larger whole supporting our beliefs in both the existence of a mind-independent world with its *an sich* properties and the existence of other minds with their private, qualitative features.

First, we observe a number of strong correlations between the occurrences of various reoccurring features within our own conscious mind (such as our feeling of a certain sort of pain, or our possession of a certain type of visual experience), on the one hand, and the occurrences of some reoccurring features within a specific segment of this internal realm (namely that which we conventionally describe as our experience of our own body), on the other. This observation provides us with a suitable epistemic ground for assuming that there is an intimate relation between our own mental life and our body as it appears within our experience.

Second, we observe a significant degree of similarity between our own body and that of some others as they appear in our experience. This observation (together with the previous one) provides us with an epistemic ground for assuming that our own mental life is not the only one intimately associated with a body

that appears in our experience, and it gives rise to the pivotal naturalist conviction that our own consciousness has the same sort of relation to the empirical world as those others which are supposed to be associated with other bodies within this world. Due to these considerations, Putnam's naturalist opponent may argue that her account of reference is not surd metaphysical or magical in character, because the obtaining and effect of the relations she invokes in this account can impose a distinctive constraint upon our theory and concept formation from a third-personal, empirical perspective.[75]

Now, how about this distinctive constraint? Is there a way to distinguish empirically a causal relation between two types of entities from a mere correspondence between the occurrences of these *relata*? Well, one important element that we may intuitively expect from a correspondence for qualifying as a constitutive part of a causal relation between two sorts of entities is the existence of some further correspondences that obtain between the occurrence of the former entities, on the one hand, and the occurrence of some other densely ordered spatiotemporal characteristics, which can be regarded as intermediate links in the causal chain connecting the cause with the effect, on the other.[76] Furthermore, it can be also supposed that the spatial order of the occurrence of these intermediate elements has a certain relation to the temporal order of them: as a first approximation, for any pairs of elements in the alleged causal chain, one's occurrence is

[75] Of course, no empirical evidence can guarantee that our beliefs in the existence of a mind-independent world with its *an sich* properties, and the existence of other minds with their private, qualitative features are actually true. Solipsism and idealism are consistent metaphysical conceptions, whose correctness cannot be excluded on empirical considerations. The fallibility of an explanatory hypothesis, however, does not mean that our epistemic grounds for its adoption are illegitimate. For a short review of the naturalistic methodology underlying this claim, see the first section of chapter 2.

[76] The exact content of the notion of densely ordered is to be specified in view of our prevailing concepts of space and time. Intuitively, what the suggested constraint purports to guarantee is that there are no gaps among the elements of a causal chain. Notice that the resulting construal of causality is incompatible with the idea of distant causation.

spatially closer/farther than the other's to the occurrence of the cause if and only if one's occurrence is temporarily closer/farther than the other's to the occurrence of the cause.[77] Clearly, these constraints are fully compatible with an empiricist (Humean) perspective on causality, and they seem to be sufficient to distinguish the intended natural link from simple correspondences between the relevant terminal *relata*.

If this elucidation of our empirical notion of causality is correct, then Putnam's just more theory argument cannot be sound: the fact that the truth values of our truth-apt representations in every possible world cannot uniquely determine the referential relation of our atomic representations to some aspects of the mind-independent world will not imply that the existing causal links between these entities cannot single out the relevant semantic relations either. This is because Putnam's idea that any correspondence between the operationally and theoretically correct use of our representations and the obtaining of some aspects of the mind-independent world can be taken as a constitutive part of a causal link between these elements is false. Our notion of cat, for instance, cannot be claimed to stand in a causal relation with cherries in *A*-worlds, because the relation of cherries to other minds in such worlds does not meet the specified (observable) criteria of causal relations.

Summing up, since Putnam's naturalist opponent has good reasons to suppose that the account of reference determination she suggested has a specific empirical content, which is clearly absent in the case of the correspondence-based, model-theoretic explanation addressed by the permutation argument, she can argue that the crucial premise underlying Putnam's just more theory argument (viz. premise (2) in the previous reconstruction) is obviously false: the operational and theoretical constraints that we observe in the course of empirical theory formation are not the only constraints that may contribute to the determination of the reference of our semantically basic representations. External factors can contribute as well. Although this contribution is magical from a first-personal perspective, we have good reasons to believe in their existence in view of our third-personal empirical theories of the relation of human minds to their natural environment.[78]

Turning to the third part of Putnam's argumentation, we saw that his permutation argument can demonstrate that the reference of our semantically atomic symbols cannot be determined merely by the specification of the truth value of our truth-apt representations in every possible world. On the other

[77] It is easy to see that the notion of spatial distance in the above formulation cannot mean the length of the shortest line connecting the location of the relevant element of the causal chain with that of the terminal cause. The notion must be rather understood as the length of the route between the element in question and the terminal cause *through the spatial coordinates of the earlier elements of the chain*.

[78] In so far as our empirical theories allow some variation concerning the way a certain body of human experience may be produced in the actual world, as we supposed while discussing Putnam's brain-in-a-vat thought experiment, the naturalist proposal entails a certain epistemological limitation upon our knowledge of the actual referents of our representations. In particular, the proposal implies that we have no ideas of these referents beyond the fallible view that they are the actual common external causes of those classes of experiential features whose occurrence is taken to be a reliable indicator of the correct declarative applicability of the relevant representations. Note, however, that this limitation does not imply that the reference of our representations is indeterminate. Whatever the causal antecedent of a certain class of experiential features may be, the referent of the representation associated with this class may still be fully determinate. Further, the fact that we cannot fully exclude that we are actually brains in a vat does not mean that we have reasons to suppose that we are. Some advocates of the causal account argued that we could have some experience which would support that assumption. Devitt (1984), 63. But there seem to be at least two reasons for querying the correctness of this position. First, what we have granted by admitting the conceivability of Putnam's scenario is that any experience that is causally brought about in our mind by the obtaining of some external conditions in the world in normal circumstances can be caused by a computer as well in the suggested sceptical situation. Accordingly, no experience can be consistently taken as a sign that the causal antecedents of our experience have changed, and we have become a brain in a vat, rather than a normal person living in a normal environment. Second, as Putnam rightly observes, on the causal account, if we were a brain in a vat, we could not think that we are in the sense we do when we are not. What this means is that no one can consistently think, on any ground, that she is actually a brain in a vat in the sense she means that under normal circumstances.

hand, we observed also that Putnam's non-standard referent assignments typically presuppose that one of the following two conditions obtains: (1) the specification of the truth values in question does not fix the truth conditions of our thoughts or sentences either; (2) the referents assigned by the suggested non-standard interpretations to our concepts or words need not be uniform and a component of the truth conditions of the sentence composed of these representations, in every possible world. The obtaining of the second condition, however, would violate the commonly accepted principle of the compositionality of reference, while the obtaining of the first would undermine both the metaphysical realist view that some identities and degrees of similarity are real properties (i.e. not merely projected by our mind), as well as Putnam's own assumption stated explicitly at the beginning of his argument. What these observations suggest is that a metaphysical realist can refute Putnam's argumentation also by first clarifying why we should doubt the obtaining of both (1) and (2), and then arguing that in absence of these conditions, the specification of the truth value of our truth-apt representations in every possible world fully determines the referential relation of our atomic symbols to the distinguished aspects of the mind-independent world.[79]

So, what can be said in support of the claim that neither (1) nor (2) obtains in the actual world? Well, our primary ground for believing in the compositionality of reference seems to be our reflective evidence of our own understanding. Apparently, it is an observable fact about our own understanding that we intend to use our concept of cat in a way which implies that its referent can never fail to be a component of the truth conditions of our synthetic judgements involving this particular representation.[80] Since this observation concerns a feature of our own narrow intentional states, we may suppose that the fact observed obtains within our own head, and illustrates the earlier indicated point that the correctness of Putnam's ain't in the head argument does not exclude that some facts within our head may effectively contribute to the determination of the reference of our atomic representations.

The metaphysical realist belief that identity and similarity are also intrinsic features of (the aspects of) a mind-independent world, rather than exclusively the product of the classificatory work of our minds, seems to be based on a wider evidential ground. First, we found explanatory considerations to the effect that Putnam is presumably wrong when he denies that "there are any inputs *which are not themselves to some extent shaped by our concepts*".[81] As we noted, one important problem with the idea that the features of our experience, or the representable aspects of the world, owe their identity and similarity relations to our classificatory work is that this tenet does not leave room for an explanation of misclassification (as opposed to mere misjudgement of what obtains in the world). Apparently, the world or our experience without preconceptual characteristics cannot impose any constraints upon our concept formation. The acceptance of this point seems to provide us with sufficient ground to adopt a correspondence theory of truth.

But why does a metaphysical realist, presumably together with her internalist opponent, believe in the existence of a mind-independent world (i.e. anything beyond the actual content of a subject's experience or conscious mind)? Well, we saw that a

[79] As we noted, Q-interpretations do not require the obtaining of either of these conditions. If Quine were right in assuming that his non-standard referent assignments do not affect the truth value of any truth-apt representation in any context of application in any possible world, then a moderate form of the permutation argument could still be advanced to challenge the idea that reference can be fixed by the specification of these truth values.

[80] In chapter 7, I shall argue that the truth conditions of our analytic claims about the spatiotemporal world are not referential in character, so what the same compositionality principle says about the referent of our concept of cat in analytic propositional contexts is merely that it is meant to be constitutive of the referent, rather than the truth conditions, of the embedding representation. A parallel principle applies to the declarative use conditions of our representations, whether or not the conditions in question are referential in character.

[81] Putnam (1981), 54.

realist has empirical grounds to believe that some parts of the spatiotemporal world (viz. human bodies) are intimately associated with what we call (narrowly understood) states of human minds. Our experience of the relation of these specific parts to the rest of that world suggests also that what obtains in the latter segment has no constitutive reliance on what obtains in the former. In other terms, our experience provides sufficient grounds for us to suppose that the representable aspects of the external world exist independently of the activities of human minds.[82]

In view of these results, we can conclude that although Putnam's permutation argument successfully demonstrates that the specification of the truth value of our truth-apt representations in every possible world in itself does not fully determine the reference of our semantically basic representations, nevertheless a metaphysical realist may insist that the same model-theoretic method works perfectly well in the actual world, in which reference is apparently compositional in character, and the identity or similarity of particular aspects of the world seems to be a real property, not merely the result of the classificatory

[82] As it was noted earlier, the metaphysical realist claim is not that the identification of the semantic content (the declarative use conditions) of our representations does not depend on the classificatory work of human minds at all. In fact, there are many ways in which we can conceptualise the content of our experience and therewith, indirectly, the various aspects of the external world. Some people distinguish between aspects which others classify as the same. This much is fully compatible with the deliverances of our experience as well. What the opponents of Putnam's internalist perspective maintain is that there are particular features that no one can correctly classify into different kinds, because they are objectively of the same kind, and if two similar particulars are classified into the same kind, then some further particulars cannot thereafter be classified into a different kind, because they are more similar (in the relevant respect) to each of the previous two than those to each other. The fact that we do not recognise every difference in the world and in our experience, and that we can cut these realms into individuals and properties in various different ways, does not imply that all identity or similarity in the world is imposed by the classificatory work of human minds. What Putnam argues for is clearly this second claim. The acceptance of the first is compatible with metaphysical realism and a correspondence theory of truth.

work of human minds. If so, then it is not merely that the constraints upon reference assignment imposed by the existing causal relations between our narrow intentional states and various aspects of the mind-independent world may exceed those which are imposed by the specification of the truth value of our truth-apt representations in every possible world, but also that in the actual world the model-theoretic constraints are sufficiently strong for determining the proper interpretation of our semantically basic representations.

One may think that this result undermines the significance of the suggested causal account of reference determination. If the reference of our basic representations can be determined in the actual world by the specification of the truth conditions of our declarative thoughts or sentences, then there seems to be no longer reason for invoking facts about causal relations between the world and our narrowly understood mental states in our account of reference determination. Before concluding this section, let me briefly explain why I think that this assumption is inappropriate.

Putting it briefly, the fact that the specification of the truth conditions of our truth-apt representations actually determines the referential relation of our basic representations to various aspects of the world, in itself, does not tell us how the relevant semantic relations between these *relata* emerge in the world at all. What it ensures is merely that the semantic relations determining the former semantic correlates (viz. the truth conditions) also determine the latter (viz. the referents). To say that the referential relations of our concepts or words are determined by the specification of the truth conditions of our truth-apt representations composed of these atomic constituents is to explain the emergence of a limited number of fine-grained semantic links between our representations and the world by invoking the existence of a virtually unlimited number of coarse-grained semantic relations between these *relata*. Clearly, if we do not understand how, by using concepts or words in a systematic way, we can think of or speak about some specific aspects of the world, then we will hardly understand how, by assigning truth values to declarative thoughts or sentences in a systematic way, we can endorse or deny the obtaining of some specific conditions in the world. Apparently, a satisfactory account of reference determination must go hand in hand with a satisfactory explanation of how our truth-apt representations acquire their relation to those aspects of the world whose obtaining or absence is meant to determine the truth value of these representations.[83] What this means is that Putnam's permutation argument is directed against an ill-chosen opponent. Whether or not the argument is accepted, a metaphysical realist will still owe us an account of how our (atomic and complex) representations acquire their semantic relations to the intended specific aspects of the mind-independent world.[84]

[83] One may object that the specification of the truth value of our truth-apt representations in every possible world need not invoke the existence of any relation between these representations and the mind-independent world, since the truth conditions associated in this manner with our truth-apt mental and physical symbols can be construed as "pure" mental objects (i.e. objects of narrow intentional states). Notice, however, that if the objection were correct, then the referential relations purportedly determined by the specification of these truth conditions could not connect our concepts or words with aspects of the external world either. This is exactly what Putnam's ain't in the head argument successfully pinned down before. But even if we set this consequence aside, the assumption that we can possess ideas of truth conditions without thereby being related to various aspects of the external world is something that no semantic realist would ever seriously embrace. Finally, we may note that, contrary to what Putnam believes since his "natural realist" turn, the task of explaining how we can conceive the obtaining of various conditions in the world cannot be simply eliminated by denying the existence of an interface between human minds and the external world. The task remains until we maintain that there are meaningful symbols which refer to some other independently obtaining conditions in the world, whether or not this "aboutness" is mediated by an interface and some external (non-cognitive) causal relations between the latter and the intended referents of the symbols. For a detailed characterisation of Putnam's more recent, natural realist perspective see Putnam (1999).

[84] The very same insight reveals why Davidson's influential semantical programme, the adoption of an inverted Tarskian approach and using truth to define meaning for natural languages, cannot be taken as a sufficiently informative account of meaning determination. Davidson (1984). Notice that the result of our discussion of Putnam's permutation argument to a certain extent justifies Davidson's programme. The requirement of finite axiomatisation guarantees that the theory envisaged will observe the compositionality of meaning, while the adoption of Tarski's adequacy condition ensures that the theory will specify the truth conditions of every truth-apt representations. If we add that Davidson's idea of truth and meaning is, in our fine-grained, non-Fregean sense, referentialist in character (i.e. it specifies the truth conditions and meaning of our representations in terms of the relevant intended subject matters), it becomes clear that his core tenet actually coincides with our conclusion: a theory which satisfies the suggested two conditions specifies the meaning/reference of our subsentential expressions as well. What Davidson's programme cannot suitably answer is, again, what determines the meaning and truth conditions of our actual representations (i.e. how do our mental and physical symbols acquire their semantic relations to the relevant aspects of the world). It is our interest in the latter question, which maintains the significance of a carefully formulated causal theory of reference.

By adopting the suggested causal account, a metaphysical realist may be able to explain the emergence of referential relations between our earliest empirical concepts or terms, on the one hand, and some aspects of the external, mind-independent world, on the other. By invoking the aforementioned facts within our head, she can then develop this account into a more sophisticated theory of how we acquire the capacity to think of or speak about entities that were never causally related to our conscious minds. From her metaphysical perspective she can also explain how the mind-independent world can impose substantial constraints upon the classificatory work of human minds. Finally, it must be also noted that the account does not assume that the declarative use conditions of our mental and physical symbols (and thus the truth conditions of our declarative thoughts and sentences) are to be specified in terms of the intended subject matter of these representations. In fact, as an explanation of how our basic representations acquire their referential links to various aspects of the mind-independent world, it does not imply anything about the relation of these intended conditions to the declarative use conditions of the relevant symbols. What this final observation teaches us is that the suggested causal account is not merely a viable conception of how our mental and physical symbols can acquire their referential links to various aspects of the world, but it is also compatible with the core tenet of this work, namely that the truth conditions of our paradigm *a priori* claims are non-referential in character.[85]

In view of these results, we can conclude that Putnam's argumentation fails to demonstrate that the adoption of metaphysical realism and a correspondence theory of truth undermines the possibility of a proper explanation of how our concepts and words can determinately refer to various aspects of the world.

Summing up, in this section I examined Hilary Putnam's argumentation against metaphysical realism and the standard realist (correspondence) theory of truth. First, I provided a brief reconstruction of the three major components of the case, respectively denoted as the permutation, the ain't in the head, and the just more theory arguments. Second, I examined the significance of Putnam's argumentation from the perspective of the semantical programme advocated in this work. My conclusion in this part was that Putnam's reasoning is definitely significant to the evaluation of our realist semantical programme, in so far as it queries the viability of any substantive correspondence theory of truth. On the other hand, internal realism does not seem to support a suitable referentialist response to the original or modified and generalised form of Benacerraf's challenge either, since it apparently embraces realism, in the sense specified in chapter 1, concerning causally inert subject matters just as much as concerning the intended referents of other truth-apt representations, and thus runs, in a referentialist theoretical framework, into the same explanatory difficulties as its traditional (metaphysical realist) counterpart. With these conclusions in mind, I turned then to the detailed discussion of the three arguments. First, I showed that the permutation argument correctly demonstrates that in absence of further constraints the specification of the truth value of our truth-apt representations in every possible world does not fully determine the referential relation of our atomic representations to the aspects of a mind-independent world. On the other hand, I also showed that the same model-theoretic method is sufficiently conducive if (as Putnam supposes) the specification of the above truth values identifies the truth conditions of our truth-apt representations, and the reference of our mental and physical symbols is compositional in character. Second, I examined Putnam's ain't in the head argument. I argued that, despite the misleading illustrative examples advanced, the argument successfully reveals that our narrow intentional states, in themselves, cannot fix the

[85] A brief sketch of the causal account could run as follows: by directing our attention to various features of our own experience, which constitutes a narrow intentional state, we can associate some of these feature tokens with a mental symbol (and thus with each other). Initially, our classificatory work is guided by the simplest qualitative similarities and identities among the occurring aspects of our experience, while later we discover more complex similarities as well, enabling us to keep our developing conceptual apparatus sufficiently limited in number. The associated aspects of our experience, as all other features in the world, occupy a specific place in the causal order of the universe. In other terms, they are also elements of various densely ordered spatiotemporal series of (*an sich*) characteristics reoccurring in the actual world. The more experiential feature tokens are associated (under normal circumstances) with a given mental symbol, the more specific the intended external semantic content (i.e. the referent) of that symbol becomes, because the fewer external features remain in the world that actually appear among the causal antecedents of all associated experiential features. Once the invoked causal relations establish the most basic referential links between our earliest representations and their intended external referents, we can introduce some further symbols, whose semantic relation to their intended referents no longer requires the obtaining of a causal link, ever in the actual world, between these *relata*. By composing symbols with established external referents, for instance, we can develop ideas of uninstantiated universals that could be exemplified in the spatiotemporal world. By extracting the feature of spatiotemporal locality of the established referential content of representations of the spatiotemporal world, we can develop concepts of abstract objects and properties. By stipulating the unreality of a certain composite subject matter, we can develop ideas of fictive entities. Since the account, in itself, does not imply anything about the emergence of determinate semantic relations between our mental and physical symbols and their respective declarative use conditions, the elaboration of a proper theory of truth (or correct applicability) for a certain discourse must involve either the clause that the latter relations are identical with those obtaining between the very same symbols and their intended referents, or the specification of a further account which explains the emergence of this second type of relation between our representations and some aspects of the real world. In chapter 7, I shall provide an outline of such an account in relation to our paradigm *a priori* beliefs about abstract states of affairs as well as our analytic claims about the spatiotemporal world.

reference of our representations either. The only remark that I added to this conclusion is that it is fully compatible with the weaker claim that some facts within our heads may substantially contribute to the determination of what our mental and physical symbols actually refer to. Finally, I turned to Putnam's just more theory argument, and showed that it is based on the problematic premise that any external factor that contributes to the determination of the reference of our symbols about the aspects of a mind-independent world must exert its influence through those operational and theoretical constraints that we ideally observe in the course of empirical theory formation. I argued that if the premise is true, then the argument successfully demonstrates that our alleged ability to determinately refer to particular aspects of a mind-independent world cannot be explained by invoking the obtaining causal relations between our narrowly understood mental states and the relevant aspects of the external world either. In the concluding part of this section, then, I specified why I think that Putnam's argumentation fails to demonstrate the incapacity of a metaphysical realist (and advocate of a correspondence theory of truth) to account for the emergence of determinate referential relations between our mental and physical symbols, on the one hand, and some aspects of a mind-independent world, on the other. First, I explained why I think we can reasonably believe that the specification of the truth values of our declarative thoughts or sentences identifies the truth conditions of these representations, and that the reference of our symbols is compositional in character. Thus, I provided reasons to believe that the model-theoretic method of reference determination challenged by Putnam's permutation argument is sufficiently conducive in the actual world. On the other hand, I also observed that the model-theoretic response to Putnam's original explanatory question is deeply inadequate, because it does not tell us anything about how our capacity to conceive potentially obtaining real conditions (and thus, alternative possible worlds) can emerge in the first place. What the model-theoretic response illuminates is merely that the emergence of a virtually unlimited number of coarse-grained semantic relations between our representations and the intended aspects of the world goes hand in hand with the emergence of a limited number of fine-grained semantic links between these *relata*. With this conclusion in mind, I finally specified why I think we should query the correctness (of the crucial premise) of Putnam's just more theory argument, and thus provided reasons to believe that the causal account challenged by this argument may suitably explain how we can determinately think of and speak about various aspects of a mind-independent world. In view of these results, I concluded that Putnam's argumentation failed to demonstrate the inadequacy of metaphysical realism and the standard realist (correspondence) theory of truth.

Summary

In this chapter, I examined two major groups of arguments that are usually taken as the most influential anti-realist challenges to realism about truth.

In section 1, I addressed Michael Dummett's acquisition and manifestation arguments to the effect that, contrary to what realists suggest, our understanding, and thus the truth conditions of our beliefs, cannot be verification-transcendent in character. First, I provided a brief reconstruction of the arguments. Second, I argued that despite the chosen terminology the real target of Dummett's criticism in the problematic discourses is not the realist, but instead the referentialist construal of truth. Therefore, I concluded that Dummett's semantical programme cannot help the advocates of referentialism escape the original or modified and generalised form of Benacerraf's challenge in the semantics of discourses about causally inert domains. Finally, I examined the two arguments, and showed that they rely on a limited view of our capacity to introduce new ideas of truth conditions. The crucial point that seems to escape Dummett's attention is that by composing ideas of declarative use conditions whose obtaining or absence we can recognise by means of our actual methods and epistemic capacities we can develop ideas of actually no longer recognisable truth conditions as well. In view of this result, I concluded that Dummett's arguments do not succeed in demonstrating the inadequacy of standard referentialism in the semantics of discourses about verification-transcendent domains. The real problem with referentialism in the semantics of our paradigm *a priori* discourses is, in line with this conclusion, not that a referentialist cannot explain how we could develop ideas of causally inert objects and properties, but instead that she cannot explain how we could acquire knowledge or reliable beliefs about the existence of such entities.

In section 2, I turned to Hilary Putnam's internal realist argumentation against metaphysical realism and the correspondence theory of truth. First, again, I provided a brief reconstruction of the three separable components of the case. Second, I observed that although Putnam's reasoning is definitely significant to the evaluation of the realist semantical programme advocated in this work, in so far as it queries the viability of any substantive correspondence theory of truth, nevertheless it does not seem to support a suitable referentialist response to the original or modified and generalised form of Benacerraf's challenge either, since it apparently embraces realism, in the sense specified in chapter 1, concerning causally inert subject matters, and thus runs, in a referentialist theoretical framework, into the same explanatory difficulties as its traditional (metaphysical realist) counterpart. Finally, I examined Putnam's three sub-arguments, and explained why I think that their conjunction fails to demonstrate that a metaphysical realist cannot explain how we can determinately refer to various aspects of a mind-independent world. My primary objection to Putnam's reasoning was that his just more theory argument is based on the problematic assumption that any external factor that contributes to the determination of what our mental and physical symbols refer to must exert its influence through those operational and theoretical constraints that we ideally observe in the course of empirical theory formation. After specifying why Putnam's naturalist opponents can reasonably reject this assumption, I argued that a sufficiently moderate form of the causal account can explain how we can determinately think of and speak about various aspects of a mind-independent world, and thus Putnam's case against metaphysical realism and the correspondence theory of truth does not prove to be sound either. What the suggested naturalistic account of our referential capacities shows us is, again, that the real problem with referentialism in the semantics of our paradigm *a priori* discourses is not that a referentialist cannot explain how we could determinately refer to causally inert objects and properties, but instead that she cannot explain how we could acquire knowledge or reliable beliefs about the existence of such entities.

In this and the previous chapters, I showed that the two semantical responses that at the end of chapter 3 appeared to be *prima facie* available for the advocates of standard referentialism to escape Benacerraf's original or modified and generalised dilemma in the philosophy of discourses about causally inert domains are equally inadequate in the light of those adequacy conditions that were put forward in the second half of chapter 2. The inadequacy of these deflationist and anti-realist responses lies in the fact that they are both incapable of explaining the objectivity of truth. In the following chapter, I shall turn to those referentialist theories that accept all the five semantical assumptions of Benacerraf's dilemma (i.e. endorse platonism about the truth conditions of our paradigm *a priori* discourses), and attempt to answer Benacerraf's challenge by querying one of the epistemological assumptions of his case.