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

Citation

Novák, Z. (2010, June 23). *A Priori truth in the natural world : a non-referentialist response to Benacerraf's dilemma*. Retrieved from https://hdl.handle.net/1887/15729

Benacerraf's Dilemma and the Explanatory Challenge
to the Standard Referentialist Construal of Truth

Introduction

It is an entrenched and plausible view in philosophy that we can gain knowledge of objective truths by evidence other than sense experience.[1] The clearest candidates of this type of knowledge are our claims about abstract (i.e. non-spatiotemporal) domains, as in pure logic and mathematics. Beyond these paradigm instances, there are some other, more contestable examples as well, including our widely shared normative claims or value judgements in ethics, aesthetics and epistemology, and the descriptive statements of metaphysics.

Once we believe in the possibility of *a priori* knowledge acquisition, it becomes natural to ask ourselves: what is happening here, *how* do we learn what is objectively and necessarily the case without relying on the deliverances of sense perception? Clearly, any response to this question will draw heavily on what can be thought of the nature of those conditions whose obtaining or absence is supposed to determine the truth value of the relevant claims. A proper explanation of *a priori* knowledge requires an appropriate conception of the meaning of *a priori* beliefs and the nature of *a priori* truths.

In philosophy of mathematics, the mutual dependence of theories of meaning and truth, on the one hand, and theories of

---

[1] Substantial parts of this chapter will appear in Novák and Simonyi (2010b). The final characterisation of the doctrinal map to be presented in section 2 owes much to the discussions I had with my coeditor while preparing the introduction for the previous volume.

knowledge acquisition, on the other, has long been an established part of common sense. This is in great part due to two brilliant articles by Paul Benacerraf, published in 1965 and 1973, which have influenced virtually every writer on the subject since.

The first of these became the groundbreaking work of mathematical structuralism. According to a structuralist, the subject matter of mathematical theories is not a single domain of abstract individuals that, beyond having certain relations to each others, also possess some further properties, which distinguish the system in which they feature from other isomorphic ones. Rather, it is either all systems of individuals exemplifying an abstract structure or the structure itself whose elements are merely positions in the structure lacking any further individuating property, so that questions about what mathematical objects really are cannot be answered beyond what the theory says about the defining relations of these objects to each other.[2]

Benacerraf's second paper explicates a dilemma which can be seen as fuelling much of the debates and inventions in early-twentieth-century works on the foundations of mathematics. The dilemma is the following. If we maintain that the truth value of our mathematical beliefs is determined by the obtaining or absence of those abstract and non-epistemic conditions that these beliefs purport to be about (i.e. whether certain mathematical objects possess certain mathematical properties), then we find ourselves unable to understand how we can, by means of natural cognitive mechanisms, discover whether or not these conditions obtain. On the other hand, if we suppose that knowledge requires appropriate causal contact between knowing minds and the obtaining truth conditions of true beliefs, then we seem to be forced to conclude that the truth conditions of our established mathematical theories cannot be construed along the standard referentialist lines. Summing up, in philosophy of mathematics, our standard referentialist conception of truth seems to be incompatible with our standard causal theory of knowledge.[3]

---

[2] Benacerraf (1965).
[3] Benacerraf (1973).

The significance of Benacerraf's observations is not confined to the philosophy of mathematics. Similar questions can be raised in the philosophy of any discourse in which we are supposed to acquire knowledge of causally inert subject matters. Logic, for instance, is mostly supposed to concern with inferential relations among propositions. Is there anything more to being a proposition than having certain inferential relations to other propositions? Are propositions, together with their inferential relations, real entities existing in an abstract (platonic) realm, or are they merely projected by human minds? Can we maintain that the truth value of our logical beliefs is determined by the obtaining or absence of those mind-independent conditions that these beliefs purport to be about? If we maintain this referentialist construal of logical truth, can we properly explain how our reasoning capacities could inform us about the obtaining or absence of such causally inert truth conditions? Isn't it the case that any causal account of how we actually *discover* objective logical truths undermines the adequacy of the standard referentialist construal of the truth conditions of logical beliefs?

In this chapter, I shall advance what I take to be the most influential explanatory challenge to the standard referentialist construal of truth in the semantics of our paradigm *a priori* discourses (i.e. in pure logic and mathematics). In section 1, I shall reconstruct Benacerraf's epistemological challenge to the standard platonist construal of mathematical truth in two forms: first, in the original, and then, in a slightly modified form that relies on a weaker epistemological premise than the original one. Having advanced these reconstructions, I shall generalise the case, and turn it into an argument against any construal that is realist and referentialist about truth in the semantics of discourses in which we are supposed to acquire knowledge or reliable beliefs about causally inert domains.

In section 2, I shall spell out the most important semantical and epistemological assumptions underlying Benacerraf's original dilemma, and then review the available theoretical positions that a referentialist may take in response to his challenge to standard referentialism about truth. Some of these options will be rejected

at the outset by reference to some basic observations about our cognitive practice in pure logic and mathematics. Two semantical and two epistemological responses, however, will be regarded as *prima facie* promising ways of defending the standard view against its non-referentialist alternative. This result will set the stage for the following three chapters of this work, in which I shall show that none of these referentialist construals satisfies the entire set of adequacy conditions put forward in chapter 2. In view of the inadequacy of the available referentialist responses to Benacerraf's dilemma, I shall conclude that the suitable answer to his puzzle is the adoption of a non-standard account of the nature of truth, which allows for a non-referentialist construal of this semantic property within discourses about causally inert domains. In chapter 7, I shall elaborate such a non-referentialist construal, and demonstrate its adequacy against the background of the explanatory requirements presented in chapter 2.

1. Benacerraf's Epistemological Challenge to Platonism about Mathematical Truth and Its Updated Generalisation

In his paper on mathematical truth, Paul Benacerraf argued that an adequate theory of the subject should meet at least two important requirements. First, it should be in conformity with our general conception of truth (in Benacerraf's terms, our "theory of truth theories"), which ensures that the suggested truth conditions of mathematical claims are, indeed, conditions for their *truth*, rather than merely conditions for their formal derivability or theoremhood. Since the only available topic-neutral theory of truth seemed to be Tarski's semantic account, which defines truth for the sentences of a given language recursively in terms of reference or satisfaction, Benacerraf's first constraint eventually requires that our theory of mathematical

truth be a specification of Tarski's general referentialist conception of truth.[4]

Second, a proper account must be compatible also with a reasonable epistemology, which explains *how* we could acquire the body of mathematical knowledge that we apparently have. Since, on Benacerraf's view, acquiring knowledge of some truths presupposes the (metaphysically thick) existence of a certain relation between the available evidence for and the obtaining truth conditions of the beliefs under consideration, what his epistemological constraint amounts to is that a proper theory of mathematical truth must specify the truth conditions of our mathematical beliefs such that they can stand in a suitable relation with our actual evidence for these beliefs. Benacerraf's formulations also reveal that, for him, the obtaining relation in question must be a causal one, so what he means by a "reasonable epistemology" is eventually a causal theory of knowledge.[5]

In view of these adequacy conditions, Benacerraf's central claim in the paper is that none of the available accounts of mathematical truth can be regarded as fully adequate, since virtually all of them satisfy one or another of these requirements at the expense of the other.[6] The standard platonist construal of mathematical truth, for instance, is mainly motivated by the intention to meet the first, semantical expectation. The truth conditions that it attributes to mathematical beliefs, however, seem to be beyond the reach of human cognitive capacities. In

---

[4] Benacerraf (1973), 666.
[5] Benacerraf (1973), 667, 671-673.
[6] Benacerraf (1973), 661. Notice that Benacerraf's argumentation is similar to the one that I pursue in this work. His objection to the theories of mathematical truth advanced so far is that they are equally unable to satisfy all those explanatory requirements that an adequate account of this subject is supposed to satisfy. His semantical and epistemological *desiderata* eventually coincide with the first (fit with a general theory of truth) and the fourth (knowledge) from among the *explananda* presented in chapter 2, while his platonist reading of the standard referentialist account in philosophy of mathematics suggests that he observes the theoretical constraint resulting from the acknowledgement of the second *explanandum* (objectivity of truth) as well.

contrast, the alternative non-standard accounts are mostly inspired by the intention to meet the second, epistemological requirement. In order to achieve this aim, however, these conceptions abandon the referentialist presuppositions of Tarski's topic-neutral theory of truth.

One may notice that Benacerraf's *desiderata* are in fact substantive specifications of the general requirements that a proper theory of mathematical truth must fit with a general theory of truth, on the one hand, and support a reasonable account of mathematical knowledge acquisition, on the other. In the case of the semantical *desideratum*, for instance, one may accept the general requirement that a proper understanding of mathematical truth must show a minimal homogeneity with our general conception of truth, without supposing that the general notion in question is what Tarski's referentialist account specifies. Benacerraf himself is fully aware of this possibility, as it appears from the following formulation:

> If, on the other hand, mathematese is not to be analysed along referential lines, then we are clearly in need not only of an account of truth (i.e. a semantics) for this new kind of language, but also for a new *theory of truth theories* that relates truth for referential (quantificational) languages to truth for these new (newly analysed) languages.[7]

Still, he neglects the alluded non-referentialist strategy in philosophy of mathematics, because no one seemed to take it seriously before:

> However, I do not give this alternative serious consideration in this paper because I don't think that anyone has ever actually chosen it. For to choose it is explicitly to consider *and reject* the "standard"

interpretation of mathematical language, despite its superficial and initial plausibility, and then to provide an alternative semantics as a substitute.[8]

As regards the epistemological constraint, one may, again, accept that the possibility of mathematical knowledge presupposes that our actual evidence for our mathematical beliefs stands in a suitable relation with the obtaining truth conditions of these beliefs, without supposing that the substantive relation in question is causal in character.[9] What I wish to emphasise at this point is that the solution that I shall propose to the dilemma will follow the former, neglected route. In particular, I will argue that there is an alternative realist semantic account of truth, which specifies the truth conditions of truth-apt representations in terms of the *correct declarative use conditions*, rather than in terms of the intended referents, of these symbols.

Despite the contestability of the above specific assumptions, today it is widely acknowledged that Benacerraf's criticism of the available theories of mathematical truth had a great impact on the subsequent development of philosophy of mathematics. His considerations, for instance, inspired the formulation of the so-called *epistemological argument* against the standard platonist construal of mathematical truth.[10] Since the argument will play a crucial role in my case against the standard referentialist construal of truth, in the following few paragraphs, I shall reconstruct it in three different forms: first, in its original form; second, in a slightly modified form that relies on a weaker epistemological premise than the first; and third, in its most generic form that seems to provide us with good reason for rejecting any construal that is realist and referentialist about truth in the semantics of

---

[7] Benacerraf (1973), 669. (Unless indicated otherwise, italics are kept as they appear in the original.)

[8] Benacerraf (1973), 669.

[9] A more radical reaction can even reject the general assumption, and maintain that mathematical knowledge does not require any contact between the mind and the obtaining truth conditions of true beliefs. I will discuss these reactions to Benacerraf's specific requirements a bit later in greater detail.

[10] As it was mentioned in chapter 2, this argument is not the only case that one may reasonably raise against a platonist construal of mathematical truth.

discourses in which we are supposed to acquire knowledge or reliable beliefs about causally inert domains.[11]

As a first approximation, Benacerraf's original argument can be reconstructed in the following way:

1. Human beings exist entirely within space-time.
2. If there exist any abstract mathematical objects and properties, then they exist outside space-time.
3. If there exist any abstract mathematical objects and properties, then human beings cannot have knowledge of them.
4. If mathematical platonism is correct, then human beings cannot have mathematical knowledge.
5. Human beings have mathematical knowledge.
6. Mathematical platonism is not correct. [12]

The first two premises of the argument are nearly uncontroversial.[13] The crucial step is obviously the third. If one granted that, then the argument would simply go through, since (4) follows from (3) and, together with the commonly accepted (5), it also entails (6).[14] The conception that underlies Benacerraf's adoption of (3) is what is usually known as the causal theory of knowledge. According to this view, a subject knows a certain fact only if she is causally related to that fact in an appropriate way. In other words, the existence of an appropriate causal link between the subject's belief state and the obtaining truth conditions of her

belief is a necessary condition of her possessing a piece of knowledge.[15]

The standard platonist response to this epistemological challenge is to reject the causal theory of knowledge and, thus, to resist (3). One received consideration that platonists tend to invoke in support of this reaction is that the indispensability of mathematics in the empirical sciences provides us with good reason for believing in the existence of mathematical knowledge, and thus in the existence of mathematical entities, even if these entities cannot stand in a causal relation with our knowing minds.[16] In response, some anti-platonists observed that Benacerraf's epistemological challenge can be reconstructed in a weaker form as well, which does not invoke the contested causal theory of knowledge.[17] In this form, the challenge relies on two other assumptions: first, that the existence of an information-conveying relation between obtaining mathematical truth conditions and human beings is a prerequisite for a suitable explanation of the reliability of mathematical beliefs; and second, that in absence of the latter explanation, there any reason for believing in the existence of mathematical knowledge tends to be

---

[11] For such reconstructions, see Field (1989), Balaguer (1998), and Hale and Wright (2002). Further reflections on Benacerraf's paper can be found in Morton and Stich (1996).

[12] The particular reconstruction adopted here is taken from Balaguer (1998), 22.

[13] Nevertheless, Gödel (1944), Gödel (1951) and Gödel (1964) as well as Brown (1991) and BonJour (1998) seem to question the adequacy of (1). I shall return to these theoretical options in the course of my discussion of the available platonist replies to Benacerraf's challenge in chapter 6.

[14] Although radical sceptics and error theorists may reject (5), their view is virtually never endorsed in present-day philosophy.

---

[15] The classical formulation of the theory can be found in Goldman (1967), and it emerged as an attempt to correct the traditional analysis of the concept of knowledge, challenged by Gettier (1963) and somewhat earlier by Russell (1912), ch. 13.

[16] The consideration is known as the Quine-Putnam indispensability argument, and it occurs, among others, in Quine (1948), Quine (1960a), Putnam (1971), and Putnam (1975d). In contrast, Field (1980) argued that mathematics is in fact dispensable in the empirical sciences, and the usefulness of mathematical theories in scientific explanations is due to their conservative character and contribution to substantial simplifications, rather than to their correspondence to the facts of a platonic realm. Beyond reasoning from indispensability, there are various other proposals for explaining the possibility of mathematical knowledge in absence of causal links between mathematical beliefs and their allegedly platonic truth conditions, which I shall discuss in chapter 6, where I develop my arguments against platonist construals of knowable truths in general.

[17] Field (1989), 25-30. For a slightly different but still similar weakening of the original argument, see Balaguer (1998), 23-24. For a recent defence of Field's version against the objections formulated by Burgess and Rosen (2005), see Liggins (2006).

undermined. In more explicit terms, the reconstruction may run as follows:

1. Human beings exist entirely within space-time.
2. If there exist any abstract mathematical objects and properties, then they exist outside space-time.
3. If there exist any abstract mathematical objects and properties, then they cannot stand in any information-conveying relation with human beings.
4. The existence of an information-conveying relation between obtaining mathematical truth conditions and human beings is a prerequisite for a suitable explanation of the reliability of mathematical beliefs.
5. If mathematical platonism is correct, then there can be no suitable explanation of the reliability of mathematical beliefs.
6. If there can be no suitable explanation of the reliability of mathematical beliefs, then that tends to undermine any reason for believing in the existence of mathematical knowledge (and thus, in a platonist semantic framework, in the existence of mathematical objects and properties).
7. If mathematical platonism is correct, then any reason for believing in the existence of mathematical knowledge tends to be undermined.
8. Human beings have reason for believing in the existence of mathematical knowledge.
9. Mathematical platonism is not correct.[18]

---

[18] Field (1989), 26. Unsurprisingly, semantical platonists have various responses to this weaker form of Benacerraf's epistemological challenge as well. In the following section, I shall say more about these responses, and in chapter 6 I shall discuss them in detail.

Finally, I wish to observe that the significance of the above arguments is not confined to the philosophy of mathematics. On a brief reflection upon the crucial premises, it may become clear that parallel objections can be made to the standard referentialist and realist construal of truth in the semantics of any discourse in which we are supposed to acquire knowledge or reliable beliefs about causally inert domains. Some philosophers, for instance, maintain that our logical, moral, aesthetic or epistemological claims are also truth-apt, and their truth value is determined by the obtaining or absence of the respective causally inert logical, moral, aesthetic or epistemological states of affairs. Clearly, in its most generic form, Benacerraf's epistemological challenge applies to the referentialist and realist construals of these logical, moral, aesthetic and epistemological truths as well. The reasoning, in this most generic form, may run as follows:

1. If there exist any causally inert objects and properties, then they cannot stand in any information-conveying relation with human beings.
2. The existence of an information-conveying relation between obtaining truth conditions of some sort of beliefs and human beings is a prerequisite for a suitable explanation of the reliability of this sort of beliefs.
3. If the truth conditions of beliefs about causally inert objects and properties are to be construed along the standard referentialist and realist lines, then there can be no suitable explanation of the reliability of these beliefs.
4. If there can be no suitable explanation of the reliability of beliefs about causally inert objects and properties, then that tends to undermine any reason for believing in the existence of knowledge of these entities (and thus, in a referentialist and realist semantic framework, in the existence of causally inert objects and properties).

5. If the truth conditions of beliefs about causally inert objects and properties are to be construed along the standard referentialist and realist lines, then any reason for believing in the existence of knowledge of these entities tends to be undermined.
6. Human beings have reason for believing in the existence of knowledge of causally inert objects and properties.
7. The standard referentialist and realist construal of the truth conditions of beliefs about causally inert objects and properties is not correct.[19]

With this reconstruction, I conclude the presentation of what I take to be the most influential explanatory challenge to the realist (i.e. platonist) version of standard referentialism in the semantics of discourses about causally inert domains. In the following section, I shall spell out the most important semantical and epistemological assumptions of Benacerraf's dilemma, and then review and briefly assess those theoretical positions that a referentialist might take in response to the above challenge to the realist form of standard referentialism about truth. The doctrinal map to be developed here will also clarify the relation of these referentialist responses to the non-referentialist alternative, which I shall briefly characterise in the concluding part of this chapter.

---

[19] From the currently relevant epistemological perspective, causally inert values and normative properties occurring in the actual spatiotemporal world as well as causally effective objects and properties occurring in realistically construed but non-actual spatiotemporal worlds can be taken as causally inert entities too. Accordingly, in its most generic form, the epistemological argument provides a challenge to those referentialist and realist construals of normative and modal truths as well, which understand the truth conditions of normative and modal claims in terms of the above spatiotemporal entities. The best-known example of such a construal in the semantics of modal claims is Lewis's referentialist, still anti-platonist realism about modal truths. Lewis (1986). An early form of this "causal argument" for a naturalist theory of possibility appears in Armstrong (1989), 3-13.

2. Referentialist Responses to Benacerraf's Dilemma and the Non-Referentialist Alternative

One way to design an exhaustive doctrinal map of the conceivable responses to the presented forms of Benacerraf's dilemma is to identify those crucial presuppositions that are responsible for the observed tension between our standard semantical and epistemological conceptions in the philosophy of discourses about causally inert subject matters. Corresponding to the horns of the dilemma, I shall classify these assumptions into two major categories: semantical and epistemological assumptions.

The most fundamental semantical assumption behind Benacerraf's original case is that mathematical claims express genuine propositions that are truth-apt, some being perhaps true while others false.[20] I shall call this first tenet *cognitivism* in the semantics of mathematics and the other problematic discourses in general.

The second semantical assumption underlying Benacerraf's dilemma is that metaphysical and epistemological considerations may impose substantive constraints upon a proper theory of meaning and truth. In particular, truth and falsity are substantive properties that play an important explanatory role, among others, in our account of knowledge acquisition about various domain.[21] I shall call this second tenet *substantivism about truth* in the semantics of the relevant discourses.

Benacerraf's third semantical assumption is that truth in mathematics is a real, non-epistemic property.[22] In other terms, the truth conditions of mathematical claims obtain (or not) independently of anyone's actual knowledge of, or capacity to

---

[20] The assumption is not explicitly stated, but clearly implied in Benacerraf's paper. Benacerraf (1973), 666. Some of the following assumptions are also implicit Benacerraf's text. To compensate the lack of explicit formulations, I shall provide more than one reference whenever possible.
[21] Benacerraf (1973), 661, 662, 671.
[22] Benacerraf (1973), 664, 665, 668, 674, 675, 676.

recognise, this particular circumstance, so no epistemic fact involving the truth value of a mathematical claim is constitutive of the obtaining or absence of the claim's truth conditions. An ideal thinker can still be claimed to be able to know all mathematical truths, but the conceptual ground of this claim is not an epistemic construal of truth, but instead a realist construal of being an ideal thinker. Generalising from the mathematical case, I shall call this third tenet *realism about truth* in the semantics of discourses about causally inert subject matters.

The fourth semantic assumption, explicitly discussed in Benacerraf's paper, is that the truth conditions of mathematical claims can be specified in terms of the intended subject matters of these claims (i.e. in terms of mathematical objects possessing mathematical properties).[23] The assumption is independent of the previous two, since it does not imply anything substantive concerning the nature of the intended subject matters.[24] What it does imply is adherence to the standard referentialist construal of mathematical truth in conformity with our notion of truth in the semantics of other segments of natural language. Following Benacerraf's terminology, I shall call this tenet, generally, *referentialism about truth*, emphasising that the term 'referentialism' has no substantive metaphysical implications here (i.e. that an advocate of this tenet need not commit herself to any conception concerning the metaphysical status and nature of the relevant subject matters).[25]

The fifth semantic assumption, also explicitly touched upon by Benacerraf, is that the subject matters of mathematical expressions are the kinds of entities they are normally taken to be.[26] For instance, numbers and geometrical objects are abstract individuals that are causally inert and have no location in physical space and time. Again, this assumption is clearly independent from the earlier ones. One may maintain that mathematical claims are about abstract (i.e. non-spatiotemporal) states of affairs without subscribing to a substantive realist interpretation of mathematical objects and properties and also without adopting the referentialist idea that the truth conditions of these claims have to be understood in terms of these abstract subject matters. Again, generalising from the mathematical case, I shall call this fifth tenet *non-revisionism about subject matter* in the semantics of the relevant discourses.

On the epistemological side, Benacerraf's most fundamental assumption is that at least some mathematical beliefs qualify as knowledge.[27] I shall call this first epistemological tenet, properly generalised, *anti-scepticism* in the epistemology of the relevant discourses. Since both tenets presuppose the truth of the known propositions, this assumption implies also that the truth conditions of at least some of our mathematical beliefs actually obtain.

The second epistemological assumption behind Benacerraf's case is that the acquisition of knowledge requires an appropriate causal link between the knowing mind and the obtaining truth conditions of the known propositions.[28] I shall call this second epistemological tenet *a causal theory of knowledge acquisition*. As we have seen, the modified form of the argument does not rely on this assumption. Instead, it rests on the conviction that the obtaining of an information-conveying contact between the above *relata* is a precondition for reliable belief formation, and therewith for the legitimacy of our beliefs in the existence of knowledge. I shall call this weaker version of Benacerraf's second epistemological tenet, interchangeably, *a contact theorist account of*

---

[23] Benacerraf (1973), 665, 672, 677, 678.

[24] Benacerraf is apparently aware of this independence. Benacerraf (1973), 664.

[25] Putting stress upon the conceptual independence of this tenet from the former two may be significant in the light of two relatively entrenched terminological conventions in present-day philosophy: first, the characterisation of construals explaining meaning and truth without any reference to intended subject matters as anti-realist accounts in semantics; and second, the characterisation of reference as a substantive relation between representations and represented entities, a notion clearly distinguishable from the deflated concept figuring in the label suggested here for the standard (broadly Tarskian) conception of truth.

[26] Benacerraf (1973), 673, 675.

[27] Benacerraf (1973), 673.

[28] Benacerraf (1973), 671, 672.

reliable belief formation, or *a contact theorist account of knowledge acquisition*.

Adopting the five semantical assumptions, we must conclude that the truth conditions of our claims about causally inert subject matters obtain (or not) without exerting any influence upon other existents, including our knowing minds. Adopting the two epistemological assumptions, on the other hand, we must conclude that at least in some cases there is an information-conveying mechanism between the obtaining truth conditions of our beliefs about causally inert entities and our actual evidence in support of these beliefs. The two conclusions clearly contradict each other: while the semantical assumptions suggest that there can be no contact between the truth conditions of beliefs about causally inert subject matters and the human minds, the epistemological assumptions imply that at least in some cases this contact obtains.

The conceivable responses to Benacerraf's original or modified and generalised dilemma can be classified also into two major categories: those that reject some of the semantical assumptions specified above, and those that abandon some of the epistemological assumptions.

Among the semantical responses, the most radical is the rejection of cognitivism concerning the problematic types of claims. If a claim is not an endorsement of a genuine proposition, and therefore it cannot be true or false, then it cannot qualify as a piece of genuine knowledge either. Of course, the systematic nature of our linguistic practice may still call for a proper explanation, but this account need not involve reference to the obtaining of any truth conditions. The best example of this *non-cognitivist* treatment of an otherwise problematic discourse is Hare's prescriptivism in metaethics, but the same strategy has been traditionally attributed to metaethical emotivists, such as Ayer and Stevenson, and more recently to metaethical expressivists, like Blackburn and Gibbard.[29] In philosophy of

mathematics, Hilbert's instrumentalist account of what he called ideal (infinite) mathematics is sometimes regarded as an instance of non-cognitivism in the sense specified above.[30] On the current understanding, the crucial tenet of non-cognitivism is that the linguistic practice under scrutiny does not serve the expression of genuine beliefs, because it is not regulated by the detection of the obtaining or absence of some semantically significant conditions that could be regarded as conditions of truth.

A less radical semantical response to Benacerraf's problem is to deny the correctness of the second, substantivist assumption, and adopt a *deflationist* position in the semantics of the relevant discourses. Deflationists maintain that a proper theory of truth and reference is orthogonal to both our conceptions of the metaphysical status and nature of truth- or declarative application conditions and our theories of how we acquire knowledge of the obtaining or absence of these conditions. In other terms, semantics is autonomous *vis-à-vis* metaphysics and epistemology. The main reason for this is that, on a deflationist understanding, truth is not a substantive property, so there is nothing to say about its nature and its relation to our epistemic capacities. Our notion of truth is fully characterised by the instances of Tarski's Disquotation Schema or its counterpart for propositions as primary truthbearers. A deflationist may still wonder how we can acquire knowledge of causally inert subject matters, and maybe even admit that, indeed, there is something theoretically puzzling in this phenomenon. Nevertheless, contrary to Benacerraf's claim, she can maintain that no response to this challenge can undermine the adequacy of referentialism about truth, since playing a substantive explanatory role in theories of knowledge is not a prerequisite for a condition to become constitutive of the truth conditions of a truth-apt representation. Classical versions of deflationism include Ramsey's redundancy theory, Strawson's performative theory, and Quine's disquotational theory, while the most influential recent forms of deflationism are Grover, Camp

---

[29] Hare (1952), Ayer (1946), Stevenson (1944), Blackburn (1993), and Gibbard (1990).

[30] Hilbert (1925).

and Belnap's prosentential and Horwich's minimal theories of truth.[31] Beyond these clearly anti-substantivist examples, as we have seen in chapter 1, deflationist conclusions can be derived from Blackburn's (purportedly substantive anti-realist) "quasi-realist" programme in semantics as well: if all distinctive claims of a realist can be endorsed, on some re-interpretation, by an anti-realist as well, then it may seem quite natural to question the intelligibility of the very contrast between realism and anti-realism, and opt for a deflationist theory of truth.[32]

The third available semantical reaction to Benacerraf's dilemma is to accept the cognitivist and substantivist assumptions, but deny the adequacy of realism, and adopt an *anti-realist* position about truth in the semantics of discourses about causally inert subject matters. Anti-realists about truth maintain that truth is a substantive epistemic property. In other terms, they hold that the truth conditions of a certain class of claims do not obtain independently of our capacities for recognising these truths (i.e. that some epistemic facts concerning the truth values of these claims are constitutive of the obtaining or absence of those truth conditions). Anti-realism in semantics and metaphysics has always found its basic motivation in epistemological considerations. No wonder that the doctrine may appear as a solution to Benacerraf's dilemma as well. If the truth value of our claims about causally inert subject matters is construed in epistemic terms, then the explanation of knowledge acquisition need not invoke an information-conveying link between the knowing mind and something whose existence is fully external to it. Anti-realist replies may differ in their stance to Benacerraf's fourth and fifth semantical assumptions (i.e. whether they maintain or reject referentialism about truth and non-revisionism about subject matter in the semantics of the relevant discourses). It may be worth noting, however, that some influential doctrines from among those which are often classified

as anti-realists about truth are arguably realist in the currently adopted sense of the term. Putnam's internal realist epistemisation of truth, for instance, is sometimes presented as a representative of (a referentialist and non-revisionist form of) anti-realism concerning this entity. Putnam, however, has never claimed that epistemic states are constitutive of the obtaining or absence of referential truth conditions.[33] Dummett's verificationist theory of truth cannot be regarded as realist in the current sense either, since the conditions that he takes to be the truth conditions of our beliefs are supposed to obtain also independently of anyone's actual knowledge of, or capacity to recognise, this particular circumstance.[34] Three further examples, whose anti-realist status is contestable, are Gibbard's projectivist semantics in metaethics, Blackburn's quasi-realist construal of our claims about moral and modal states of affairs, and Peacocke's conceptualism in the semantics of *a priori* discourses.[35] More plausible examples of anti-realism about truth include the construals of subjective idealists, Carnap's conventionalism about *a priori* (analytic) truth, and maybe Brouwer's intuitionist theory in philosophy of mathematics.[36]

The fourth semantical response to Benacerraf's dilemma is to reject his referentialist assumption, and adopt *non-referentialism about truth* in the semantics of discourses about causally inert subject matters. Non-referentialists maintain that the truth conditions of a certain class of claims cannot be specified in terms of the intended subject matter of the constituents of these claims. For instance, on a non-referentialist construal, the truth conditions of mathematical claims are not mathematical states of affairs, whichever way these would be further understood.

---

[31] Ramsey (1927), Strawson (1950), Quine (1970), Grover, Camp and Belnap (1975), and Horwich (1998b).

[32] Blackburn (1993).

[33] What the internal realist Putnam argues for is that the identity conditions of the intended states of affairs that can be regarded as the referential truth conditions of our beliefs are created by the classificatory work of mind. Putnam (1981).

[34] What Dummett's anti-realist assumes is that the truth conditions of our beliefs are always verifiable (i.e. that we have an effective, though fallible, method to determine whether or not they actually obtain). Dummett (1991).

[35] Gibbard (1990), Blackburn (1993), Peacocke (2005).

[36] Kant (1781/1787), Carnap (1934), Brouwer (1949).

Rather, they are conditions that may or may not obtain in a non-mathematical realm. Non-referentialism does not imply anything about the metaphysical status and nature of the relevant subject matters. Nevertheless, it makes realism about truth compatible with anti-realism, fictionalism, eliminativism or quietism about subject matters. Of course, as Benacerraf rightly observed, an advocate of this position must explain what makes her preferred non-referential truth conditions qualify as conditions of *truth*. Once she can deliver this explanation, she can construe the relevant conditions, without reducing the corresponding subject matters, either in anti-realist or in epistemologically unproblematic realist terms. Examples of this non-referentialist strategy may include Dummett's verificationist construal of truth in discourses about epistemically inaccessible domains, Blackburn's and Gibbard's projectivist theory in metaethics, and Putnam's and Hellman's modal structuralism in philosophy of mathematics.[37] If conative attitudes, possession conditions of concepts and analytic links within our personal system of representation are construed realistically, then a number of influential accounts that have been developed as an alternative to the epistemologically problematic realist and referentialist construals of truth can be classified as realist in the semantics of discourses about causally inert entities.

The fifth semantic strategy that may be adopted in reply to Benacerraf's dilemma is to reject the fifth semantical assumption specified above, and embrace a *revisionist* construal of the subject matter of the problematic discourses under consideration. In the case of mathematics, for instance, this would amount to the view that mathematical claims are not about abstract states of affairs whose constituents are causally inert and have no spatiotemporal location, but instead they are either about some aspects of the natural world, or about some concepts in an active intellect, or about some other entities that can influence the human mind. Alternatively, a revisionist can take mathematical claims to be about *in re* or *ante rem* structures, rather than about a single system of individuals whose members, beyond having certain relations to each others, also possess some intrinsic properties that distinguish the system they constitute from isomorphic systems of other individuals. In other terms, she can take these claims to be either about all systems of individuals exemplifying a certain structure or about the structure itself that can be exemplified by those systems. Revisionism in itself does not imply anything about the metaphysical status of the relevant subject matters. For instance, a structuralist interpretation of mathematics is compatible with a deflationist, an anti-realist and a realist construal of mathematical referents as well. Nonetheless, a major motive behind a revisionist construal of the subject matter of mathematics and other discourses about *prima facie* causally inert subject matters is that this construal allows for the wedding of a substantive realist and referentialist understanding of truth with a causal contact theory of knowledge acquisition in the philosophy of the relevant discourses. Theories falling into this class include Mill's and Kitcher's referentialist naturalism and various forms of structuralism in philosophy of mathematics.[38]

In case one does not want to follow any of the five semantical strategies characterised so far, one may try to answer Benacerraf's original or modified and generalised dilemma by querying at least of the epistemological assumptions of the case. The most radical epistemological strategy is to deny the possibility of knowledge about causally inert entities. If our

---

[37] Dummett (1991), Blackburn (1993), Gibbard (1990), Putnam (1967), Hellman (1989). Benacerraf himself is rather sceptical about the prospects of this strategy, but he considers Putnam's modal structuralism also a possible attempt in this direction. Benacerraf (1973), 669. Shapiro (2000) also emphasises the logical independence of what he calls "realism in ontology" and "realism in truth-value", and he takes Chihara (1990) and Hellman (1989) as representatives of the strategy of adopting realism in truth-value without realism in ontology. Shapiro (2000), 32-33. Note, however, that Shapiro's notion of realism in truth-value is not synonymous with the notion of realism about truth as it is understood in this work. The contrast is spelled out in fn. 7 in chapter 2.

[38] Mill (1843), Kitcher (1984). Influential structuralist accounts include Benacerraf (1965), Resnik (1997), Shapiro (1997). For a recent defence of structuralism, see Isaacson (forthcoming).

received theories of such entities do not qualify as knowledge, then the adoption of Benacerraf's five semantic assumptions concerning the claims and implications of these theories remains compatible with the standard causal theory of knowledge. A limited form of scepticism may result from an error-theorist view of our received beliefs concerning causally inert entities. Such views have been defended by Mackie in metaethics and Field in philosophy of mathematics.[39] An error-theorist argues that our received conceptions of a certain domain are equally false, and thus cannot qualify as knowledge, because the world does not contain those individuals and properties whose existence is required for their truth. Note, however, that an error theorist need not assume that the existence of the relevant entities is a precondition of any truth about the corresponding domains. In absence of this assumption, she may maintain, for instance, that negative existential beliefs about causally inert entities are still true, and as such potentially qualifying as knowledge. A limited scepticism like this cannot resolve Benacerraf's dilemma. In order to save the compatibility of the standard realist and referentialist semantics with the standard causal theory of knowledge, one must deny the existence of any type of knowledge of the relevant causally inert domains.

The second epistemological strategy that can be adopted in response to Benacerraf's case is to insist on the adequacy of the five semantic and the first epistemological assumptions, and query the idea that the acquisition of knowledge, or the reliability of belief formation, requires an appropriate causal link between knowing minds and obtaining truth conditions. Instead of admitting the general adequacy of this causal account, the proponents of this position may argue that in the case of discourses about causally inert subject matters the contact between minds and obtaining truth conditions is not causal in character. I shall call this alternative *a non-causal contact theory of knowledge acquisition or reliable belief formation*. The classic example of

this strategy is Gödel's quasi-perceptivist account of mathematical knowledge, while more recent instances of this category include BonJour's account of rational insight and James Brown's (Gödelian) view of mathematical knowledge and the nature of thought experiments.[40]

The third epistemological option one can choose in response to the dilemma is to deny the adequacy of the second epistemological tenet even in its weaker form, and subscribe to *a no-contact theory of knowledge acquisition or reliable belief formation*. Advocates of this position maintain that, although a proper explanation of knowledge acquisition may require an account of how the epistemic grounds of a knowing mind for adopting a certain class of true beliefs can reliably indicate the obtaining of the truth conditions of these representations, nevertheless, at least in the case of our discourses about causally inert subject matters, this account need not invoke the existence of any contact between these grounds and those conditions. Examples of this category include Wright's and Hale's neo-Fregean abstractionist, Balaguer's full-blooded platonist, Katz's and Lewis's necessity-based, and Shapiro's and Resnik's structuralist strategy to account for the possibility of mathematical knowledge.[41] In case one takes his holistic view of science together with his conception of ontological commitment seriously, Quine's empiricist epistemology also qualifies as a no-contact theory of mathematical knowledge.[42]

From among the five semantic and three epistemological responses reviewed in the previous paragraphs, seven seem to provide sufficient theoretical resources for the advocates of standard referentialism to maintain their view in the face of the epistemological challenge presented in section 1. The four semantic responses that they can give are that our epistemologically problematic beliefs about causally inert domains

---

[39] Mackie (1977), Field (1980).

[40] Gödel (1944), BonJour (1998), Brown (1991).
[41] Wright (1983), Hale (1987), Balaguer (1998), Katz (1981), Lewis (1986), Shapiro (1997), Resnik (1997).
[42] Quine (1948), Quine (1951).

are not truth-apt at all, or that they are true or false only in a deflated sense, or that they are made true or false by the epistemically construed (i.e. merely projected, non-platonic) obtaining or absence of their referential truth conditions, or that they are made true or false by the real obtaining or absence of their referential truth conditions under a revisionist, non-platonistic interpretation. Either way, truth can be kept to be understood in terms of intended referents. The three epistemological responses that may be endorsed by an advocate of standard referentialism are that our beliefs about causally inert domains are not reliable or not qualifying as knowledge, or that they are reliable or known in virtue of a non-causal information-conveying contact between our minds and the relevant platonic truth conditions, or that they are reliable or known in virtue of something other than a contact between our minds and the relevant platonic truth conditions. Either way, again, the standard referentialist construal of truth appears to be preserved without leaving important *explananda* unexplained.

In the case of our paradigm *a priori* discourses, at least two of these referentialist responses, viz. non-cognitivism and scepticism, seem clearly inadequate. A theory of truth whose viability presupposes that our accepted claims in pure logic and mathematics are not truth-apt, or cannot be regarded as expressions of genuine knowledge or reliable beliefs is clearly incompatible with our cognitive and linguistic practice and our fundamental belief in the existence of truth and knowledge in pure logic and mathematics.

Revisionist responses to Benacerraf's dilemma can be also disqualified on relatively simple considerations. The general problem with these reactions, as it was mentioned in chapter 1, is that they must either deny that our referential intentions accompanying the meaning-conferring and meaning-communicating applications of our mental and physical symbols play an essential role in the determination of the nature and identity of what these symbols actually refer to or simply run counter an existing and dominant cognitive and linguistic practice in pure logic and mathematics. Beyond this general objection, I

may add that the most popular forms of revisionism in philosophy of mathematics (i.e. those *ante rem* and *in re* structuralist construals that imply that mathematics is about abstract entities) have no better reply to Benacerraf's original or modified challenge than the traditional account that maintains a non-revisionist construal of the subject matter of mathematics. Accordingly, the arguments that I shall advance in the following three chapters against the alternative non-revisionist forms of standard referentialism equally apply to these "moderate" revisionist strategies as well. As to the "radical" versions of revisionism in the semantics of these paradigm *a priori* discourses, which suppose that the subject matter of pure logic and mathematics is not even abstract in character, one may further object that such theories face serious troubles while trying to account for the necessity of logical and mathematical truths and the infinity of logical and mathematical domains. Due to these considerations, in the remaining part of this work, I shall largely disregard these moderately and radically revisionist forms of standard referentialism as well.[43]

The remaining four referentialist responses to Benacerraf's original or modified and generalised dilemma are *prima facie* more promising, so they will receive longer discussions in this work. These constitute the content of the following three chapters. The primary purpose of these chapters is, nevertheless, to show that none of these replies can save the adequacy of a referentialist construal of the paradigms of *a priori* truth in the light of the adequacy conditions set for such an account in chapter 2. In chapters 4 and 5, I shall argue against the adequacy of the remaining two semantical responses, and the resulting deflationist and anti-realist forms of standard referentialism, respectively. The main reason for which I believe we had better reject these theories is that they cannot explain the objectivity of truth and falsity (i.e. the fact that the truth value of our beliefs is independent of what anyone ever believes about them). What a

---

[43] Nevertheless, in chapter 6, I shall briefly examine the viability of a structuralist response to Benacerraf's epistemological challenge in philosophy of mathematics.

successful account of this *explanandum* seems to require is that truth be a real property, which characterises its bearers in virtue of the obtaining of some semantically relevant conditions (i.e. the bearers' truth conditions) in the actual world. This, however, is something that neither deflationists nor anti-realists should be willing to endorse: the former because they refuse any substantive claim about the nature of truth, while the latter because they believe in the adequacy of an epistemic construal of this semantic property.

The two epistemological responses to Benacerraf's dilemma cannot be rejected on the previous grounds, since the referentialist position that they attempt to save is the traditional platonist one, which provides a realist construal of the paradigms of *a priori* truth. Nevertheless, in chapter 6, I shall query the adequacy of these platonist epistemological reactions as well. My main objection to them is, briefly, that the accounts they provide of the acquisition of knowledge or the reliability of beliefs within discourses about causally inert domains are either *ad hoc* and exotic, undermining all constructive methods for examining the nature and shortcomings of the relevant forms of belief formation, or insufficient, leaving us without any positive epistemic ground for supposing that the conditions whose obtaining they stipulate to be necessary and sufficient for the truth of the relevant claims indeed obtain in the intended platonic realms. If my arguments against the advanced non-causal contact theories and no-contact theories are correct, then the epistemological challenge to this realist form of referentialism in the semantics of our discourses about causally inert domains cannot be suitably answered at all.

With the fall of these referentialist strategies, the only response that is left to Benacerraf's original or modified and generalised dilemma is the one that we were advised by him to ignore because of its superficial and initial implausibility: namely, the rejection of the standard referentialist construal of truth in the semantics of discourses about causally inert domains, and the elaboration of an alternative construal as a substitute.[44] One major advantage of this response is that it can avoid the shortcomings of its referentialist alternatives. First, beyond observing our commitment to the existence of truth and knowledge in pure logic and mathematics, a non-referentialist construal of truth in the semantics of discourses about causally inert domains enables us to endorse a substantive naturalistic account of this property without subscribing to a revisionist construal of the relevant subject matters. Second, in so far as the naturalistic construal in question embraces realism about the subject, it can also explain the objectivity of logical and mathematical truth. Third, since it locates the obtaining truth conditions of our logical and mathematical beliefs in the causally efficient spatiotemporal world, it fits well with a causal explanation of logical and mathematical knowledge, or of the reliability of logical and mathematical beliefs.

Supposing that the adoption of such a naturalistic and non-referentialist construal is, indeed, the correct response to Benacerraf's dilemma, one may still wonder *how* a theory of this kind can meet the remaining adequacy conditions enumerated in chapter 2. This will be shown in the last chapter of this work. Concluding this chapter, I shall merely anticipate a response to Benacerraf's principal objection to the non-referentialist strategy, which I presented before. As we have seen, Benacerraf's main problem with a non-referentialist construal of mathematical truth was that it did not fit with our received (broadly Tarskian) general conception of truth, and therefore could not be legitimately regarded as a theory of mathematical *truth*. A non-referentialist is, of course, permitted to respond by querying the adequacy of this general account, but in case she does so, she must explicate an alternative conception, which renders her construal a theory of (a specific kind of) *truth*. Now, the general conception that I think an advocate of non-referentialism can propose instead of the standard referentialist construal is, briefly, an "inflated" realist

---

[44] Benacerraf (1973), 669.

account, according to which truth is *the property of possessing declarative use conditions that actually obtain in the world, in so far as the bearer of this property is a sufficiently complex (i.e. truth-apt) mental or physical representation.*[45] Adopting this general concept, a non-referentialist may argue that if the construal she proposes is a construal of the correct declarative use conditions of our paradigm *a priori* beliefs, then it can be legitimately regarded as a construal of the paradigms of *a priori truth*, which thereby satisfies the first adequacy condition on our list in chapter 2.

Summary

In this chapter, I presented what I take to be the most influential explanatory challenge to the standard referentialist construal of truth in the semantics of our paradigm *a priori* discourses (i.e. pure logic and mathematics).

In section 1, I reconstructed Paul Benacerraf's epistemological challenge to platonism about mathematical truth and its updated generalisation. In its most generic form, the argument was shown to query the adequacy of a substantive realist *and* referentialist construal of truth in the semantics of discourses in which we can acquire knowledge or reliable beliefs about causally inert domains.

In section 2, I set out the most important semantical and epistemological assumptions underlying Benacerraf's dilemma in its original or its slightly modified form, and reviewed those theoretical positions that seem to be available in response to this challenge to platonism in the semantics of discourses about causally inert domains. Seven of the resulting eight positions seemed to provide sufficient theoretical resources for a referentialist to maintain her view in the face of the epistemological challenge presented in section 1. Three of these responses, however, turned out to be inadequate on relatively simple considerations: non-cognitivism was disqualified by our fundamental belief in the existence of logical and mathematical truths; scepticism by that in the existence of logical and mathematical knowledge; and finally, revisionism by our intentionalist notion of subject matter. Accordingly, the advocates of standard referentialism seemed to be left with four possible responses to Benacerraf's dilemma.

In the final part of this chapter, I provided a brief summary of the dialectical purpose of the remaining four chapters of this work. Chapters 4, 5 and 6 were said to discuss the four referentialist responses specified before, and show that neither of them can save the adequacy of standard referentialism as a universal conception of truth. The conclusion we were said to be left with after these chapters was that the proper response to Benacerraf's dilemma must be the rejection of the standard referentialist construal of truth in the semantics of those discourses in which we are supposed to acquire knowledge or reliable beliefs about causally inert domains, as in the case of pure logic and mathematics. Of course, if the conclusion is correct, then at least one non-referentialist construal of these purportedly *a priori* truths must provide an acceptable account of all major *explananda* set for such a construal in chapter 2. To demonstrate that this condition, in fact, obtains was said to be the burden of the last chapter of this work.

---

[45] In chapter 4, I shall develop this general conception of truth in detail.