Universiteit Leiden
The Netherlands

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

Novák, Z.

A PRIORI TRUTH IN THE NATURAL WORLD

A NON-REFERENTIALIST RESPONSE TO BENACERRAF'S DILEMMA

PROEFSCHRIFT

TER VERKRIJGING VAN
DE GRAAD VAN DOCTOR AAN DE UNIVERSITEIT LEIDEN,
OP GEZAG VAN RECTOR MAGNIFICUS PROF. MR. P.F. VAN DER HEIJDEN,
VOLGENS BESLUIT VAN HET COLLEGE VOOR PROMOTIES
TE VERDEDIGEN OP WOENSDAG 23 JUNI 2010
KLOKKE 13.45 UUR

door

ZSOLT NOVÁK

geboren te Boedapest in 1970

PROMOTIECOMMISSIE

promotor	Prof. dr. B.G. Sundholm
co-promotor	Dr. J.W. McAllister
overige leden	Prof. M. Balaguer, *California State University, Los Angeles*
	Prof. P. Horwich, *New York University*
	Prof. G. Martí, *Universitat de Barcelona*
	Prof. dr. R. van Woudenberg, *Vrije Universiteit Amsterdam*

To my Mum and Dad

ACKNOWLEDGEMENTS

To complete a dissertation is to complete a long journey. My journey was long, much longer than I originally expected. To reach this end, I received support from several people and institutions for which I wish to express my deep gratitude here.

First and foremost, I wish to thank my supervisors, Dr. James McAllister at the University of Leiden as well as Professor Nenad Miščević and Professor Howard Robinson at Central European University, for the ongoing academic and personal support that they gave me during my doctoral years in Leiden and Budapest. I also thank my promotor, Professor Göran Sundholm for his earlier comments and his more recent administrative support in the process of my graduation. Beyond them, I wish to acknowledge my special debt to András Simonyi, my closest colleague and friend since our first encounter with philosophy, whose material feedback and informed suggestions had a hardly expressible impact on this work. Finally, another person whose influence must be emphatically acknowledged is Professor Daniel Isaacson, my external supervisor, strong critic, and warm supporter in Oxford. The fact that I could not convince him of the correctness of my position provided a constant motivation for me to elaborate further on my case in the last couple of years.

Since the beginning of this project, I received comments on parts of this work in draft from various professors and colleagues. In particular, I wish to thank Professor Timothy Williamson and Professor John Campbell for their encouraging feedback during my research year in Oxford, Professor Paul Horwich, Professor Michael Devitt, and Professor Hartry Field for their constructive and helpful notes on an early exposition of my ideas in New York, Professor Katalin Farkas for her thoughtful comments on the first draft of this work, Professor Mark Colyvan for the illuminating conversation on Quine's philosophy of mathematics in Dubrovnik, and Professor Mark Balaguer for the intense discussion of my views last summer in Budapest. From among my colleagues and friends, I wish to

thank Bjørn Jespersen, Arianna Betti, Catarina Dutilh Novaes, Leon de Bruin, Gábor Zemplén, Tamás Demeter, Károly Varasdi, Zsófia Zvolenszky, András Szigeti and Péter Rauschenberger for the talks about this work or other issues in philosophy during the last few years. I also had the opportunity to present parts of this work at seminars and conferences in Lund, Lisbon, Leiden, Oxford, Dubrovnik and Budapest, where I received interesting comments from my audience as well.

Special thanks go to Professor Gábor Forrai, my supervisor and chief mentor during my undergraduate years, whose vivid mind and reliable personality were and will always remain a guiding example for me. I also benefited much from his comments on this work. In a similar vein, I wish to thank Professor János Laki, Professor Ferenc Huoranszki, Professor János Kelemen, Professor István Bodnár, Professor András Máté, Professor László Tengelyi, Professor Ágnes Erdélyi, Professor Tamás Miklós, and Professor Gábor Boros for the example and support they gave me during my undergraduate years in Budapest.

I could not have completed this work had I not enjoyed the stimulating and supportive work environment provided by my two host institutions, the Department of Philosophy at Central European University, and the Institute for Philosophy at the University of Leiden. I wish to thank both faculties for the kind support I received from them during my doctoral years. I am particularly grateful to Krisztina Biber and Lies Klumper, the coordinators of these units, for their unfailing patience and help. Lies' peerless hospitality in Leiden was an unceasing source of comfort for me from the very beginning of my stay. Last but not least, I wish to express here my gratitude to Yehuda Elkana, former president and rector of CEU, for his special attention and ongoing support in Budapest from our first meeting until the final days of his term.

Beyond enjoying these benefits, my project received generous material support from a number of other persons and institutions. In particular, I wish to thank the Hungarian Institute of Political History and its director, Professor György Földes, for an early grant in Budapest, the Hungarian Scholarship Board, the Collegium Hungaricum in Vienna and its former scientific director, Dr. Gábor Ujváry, and the Institute Vienna Circle and its director, Professor Friedrich Stadler, for supporting my research at IVC in 2000 and 2001, the Open Society Institute and the UK Foreign and Commonwealth Office for sponsoring my research with a Chevening Scholarship, and Wolfson College for providing me with a wonderful living and working environment in Oxford in 2002-2003, the Netherlands Organisation for International Cooperation in Higher Education for facilitating my research at the University of Leiden by offering me a Huygens Scholarship for 2004-2005, the Leiden University Fund for financing my research trip to New York in 2006, and the Open Society Foundation for awarding me a Kathleen Wilkes Scholarship that enabled my participation at the Annual Philosophy of Science Conference in Dubrovnik in 2007.

On a more personal level, first, I wish to thank my two paranimfs, Marije Martijn and Jur Koksma, for their invaluable assistance in the graduation process, the unforgettable moments they shared with me in the last couple of years, and, most importantly, the warm friendship that helped me so much in finding a second home in the Netherlands. Beyond them, I wish to thank my friends in Budapest, as well as those whom I met during the completion of this project in Vienna, Oxford, Leiden and Amsterdam, for their unfailing love and joyful company, without which I could hardly have kept myself at my computer for such a long time.

My cousin, Zoltán Novák, provided indispensable help to me in designing and editing this volume, for which I am very grateful too.

Finally, my greatest debt is owed to my family, especially to my parents whose unconditional love and support I am afraid I shall never be able to duly return, and to whom, with full awareness of this debt, I lovingly dedicate this work.

Budapest, May 2010

To suppose that philosophy divides into separate compartments labeled "philosophy of mind", "philosophy of language", "epistemology", "value theory", and "metaphysics", is a sure way to lose all sense of how the problems are connected, and that means to lose all understanding of the sources of our puzzlement.

Hilary Putnam
(*The Threefold Cord*, p. 69)

TABLE OF CONTENTS

INTRODUCTION

The difference between *a priori* and empirical truth, knowledge, justification and evidence is one of the central contrasts encountered in modern philosophy. Still, if someone asks what philosophers commonly mean by the apriority of a piece of truth or knowledge, an honest answer can hardly include more than the classic negative characterisation of this property: a certain truth is meant to be *a priori* in so far as one can acquire *a priori* knowledge of it, and a piece of knowledge is *a priori* if it can be justified without reliance on the deliverances of experience. If the interrogator wants to learn something positive about the nature and extension of the alleged non-experiential form of justification, then she must soon realise that her interlocutors have only vague and divergent views about this subject. Even worse, they have no clear and commonly held conception of what counts as a piece of experience, and they have no consensus either about the sense in which (or the extent to which) the evaluation of an *a priori* belief must be independent of what we can learn from experience.

What is relatively clear and commonly accepted is that the paradigms of *a priori* truth, knowledge, justification and evidence occur in pure logic and mathematics. Even if one denies the existence of *a priori* truth and knowledge, what one usually wants to emphasise and argue for is that our knowledge of logical and mathematical truths is also based on experience. In view of this implicit agreement, one may wonder why it is so hard to develop a positive account of the nature and extension of *a priori* truth and knowledge. Couldn't we simply examine the cognitive mechanisms underlying logical and mathematical belief formation and decide which features of these mechanisms we regard as the defining characteristics of apriority? The decision in question would determine also which other truths can be known without reliance on experience, so we could establish a relatively sharp line between experiential and non-experiential justifications (or pieces of evidence) as well.

Simple as it may sound, the proposal just mentioned proves quite hard to realise. This is mainly because philosophers have no clear notion of what knowledge acquisition in pure logic and mathematics consists in, and thus what exactly it is that could be systematically examined at the beginning of a search for an appropriate positive account of *a priori* truth and knowledge. In contrast to the case of empirical belief formation, whose types (such as perception, recollection, and introspection) are commonly supposed to occur in the empirically observable spatiotemporal world, it is rather obscure which facts we actually interact with, and by means of which cognitive mechanisms, when we develop our beliefs in pure logic and mathematics.

In philosophy of mathematics, the problem has been nicely articulated by Paul Benacerraf in a paper on mathematical truth. In Benacerraf's reconstruction, what is puzzling about mathematical knowledge is that the standard view of what one knows in pure mathematics seems to be incompatible with the received (scientific) conception of the nature of human knowledge acquisition. According to the latter conception, human knowledge is a natural phenomenon, which requires causal interaction between the knowing mind and the obtaining truth conditions of the known propositions. The standard (referentialist) view of these truth conditions, however, is that they are always identical with those states of affairs that the relevant beliefs are about. The truth conditions of the belief *that zebras are herbivores*, for instance, are supposed to be the conditions that zebras are herbivores. If these conditions obtain, then the belief is supposed to be true, if not, then false. In the case of pure mathematics, this means that the relevant conditions are abstract mathematical conditions, whose obtaining or absence has no causal impact on anything else in the world. If mathematical knowledge amounted, indeed, to knowledge of the obtaining of such conditions, and knowledge required causal contact, then mathematical knowledge would be impossible. But this sounds absurd. An account that entails that we cannot know that two plus three equals five can hardly be regarded as adequate. Consequently, at least one of the two fundamental assumptions

must be wrong. Either knowledge does not require causal interaction between minds and obtaining truth conditions, or the truth conditions of mathematical beliefs are not identical with what these beliefs purport to be about (and thus what we know in pure mathematics is not the obtaining of the intended abstract states of affairs).

Note that the same problem arises in the case of any knowledge candidate that is supposed to be about real entities that have no causal impact upon our actual cognitive capacities. Logical beliefs, for instance, are often supposed to be about abstract and mind-independent propositions and their inferential relations, conditions that obtain in a causally inert domain too. Ethical and other normative properties are also sometimes construed as causally inert real entities. Finally, beliefs about what must or may be the case are sometimes understood as beliefs about conditions obtaining in various equally real but non-actual worlds that are also causally separated from the one in which we develop these beliefs.

Faced with the difficulty, philosophers have tried to save their uniform referentialist conception of truth in a number of different ways. Deflationists about truth have argued that a proper theory of truth is metaphysically neutral, and has no substantive role in the systematic explanation of cognitive and other truth-related phenomena. Less esoterically, their view is that, for instance, by maintaining that the truth conditions of the belief *that zebras are herbivores* are the conditions that zebras are herbivores we do not say anything about what we think we actually discover (by real cognitive mechanisms) when we learn that zebras are herbivores. Anti-realists about truth hold that the truth conditions specified in terms of the relevant intended referents are in some way epistemic, so their obtaining or absence need not causally affect the mind for knowledge to be realised. Finally, platonists about logical and mathematical truth either assume that we have special cognitive capacities to acquire knowledge of the obtaining of causally inert (extra-mental) truth conditions, or simply deny that knowledge requires interaction between obtaining truth conditions and the knowing mind.

Notably, none of these reactions helped the community develop a clear understanding of what knowledge acquisition in pure logic and mathematics consists in. To say that mathematical states of affairs are deflated or in some sense mental, or that their obtaining can be grasped by the exercise of some specific capacity, such as intuition, or that mathematical knowledge does not require any contact between obtaining truth conditions and knowing minds does not clarify what philosophers (or scientists) could examine in order to develop a positive account of the nature and extension of *a priori* truth and knowledge. Moreover, the proposals just summarised give rise to a number of other explanatory puzzles that are at least as disturbing as the one they were supposed to resolve.

But how about the alternative route? Why couldn't philosophers simply abandon the standard referentialist construal of the truth conditions of these problematic beliefs? What is so implausible in the assumption that the truth conditions of correct logical and mathematical beliefs obtain in the actual spatiotemporal world, rather than in the abstract domains that these beliefs purport to be about, and we can acquire knowledge of these truths by means of some natural cognitive mechanisms that are (in principle) no less observable than those underlying the existing types of empirical belief formation?

According to Benacerraf, the main reason for which philosophers had better insist on the standard (broadly Tarskian) referentialist understanding of truth is that this theory is the only available, articulate and *prima facie* plausible, general conception of the subject. In absence of such a construal, one can hardly motivate the assumption that certain conditions are the *truth* conditions (as opposed to, say, the rational acceptability conditions) of some beliefs.

Beyond this conceptual point, there are other explanatory challenges as well that an advocate of the non-referentialist option has to face. One may wonder, for instance, what makes it the case that the truth value of logical and mathematical beliefs seems necessary in character if the truth conditions of these beliefs are supposed to obtain contingently in the spatiotemporal

world. Further, one may ask how people could acquire *a priori* knowledge of any fact in the natural world. Finally, one may wonder how the obtaining of some presumably finite natural conditions could guarantee the truth of a theory about an infinite and abstract domain.

In view of these and other explanatory difficulties, today most philosophers believe that the proper response to Benacerraf's dilemma must preserve the standard referentialist construal of truth. My primary purpose in this work is to show that this belief is mistaken.

In my view, the appropriate reaction to Benacerraf's challenge is neither the deflation or epistemisation of truth, nor the postulation of exotic cognitive capacities, nor the denial of the received contact theory of knowledge. Instead, the proper response to the puzzle is to recognise that the truth conditions of some beliefs are not identical with those conditions that these beliefs purport to be about. Beyond explaining why I think that the conceivable referentialist strategies to escape the dilemma are equally mistaken, in this work I shall also put forward a particular non-referentialist construal of logical and mathematical truth that arguably satisfies *all* major explanatory requirements one can reasonably set for such an account. The construal will provide a relatively clear notion of what logical and mathematical knowledge acquisition consists in, and it will inspire a scientifically specifiable positive view of the nature and extension of *a priori* truth, knowledge, justification and evidence as well.

The central claim of the envisaged construal is that the truth conditions (as opposed to the intended referents) of logical and mathematical beliefs are natural conditions in human heads. In particular, they are analytic relations among representations that subjects develop in their heads in the course of their cognitive engagement with their direct natural environment. In view of this assumption, I shall call the proposal a *representationist* construal of the relevant truths. The construal preserves the realist idea that the truth value of logical and mathematical beliefs is independent of anyone's actual opinion of this issue. It subscribes to the moderate empiricist claim of the analytic nature of logical and

mathematical beliefs, but it identifies the factual basis of analytic truths in the actual natural world. As an essential component, the construal assumes that a subject can detect the obtaining or absence of the relevant representational conditions in her head without actually thinking about them. This is how these conditions can serve as non-referential truth conditions for those beliefs. Finally, being natural in character, the obtaining or absence of these conditions can causally influence the subject's cognitive capacities, which means that the construal is compatible with the received "contact theory" of knowledge acquisition as well. Truth and knowledge in pure logic and mathematics emerges as a natural phenomenon, which can be subject to systematic empirical investigation.

Having defended this particular non-referentialist construal of the paradigms of *a priori* truth and knowledge, at the end of this work, I shall argue that the critical feature of representationality can be more than a contingent characteristic of *a priori* truths. In fact, there are good reasons for us to take this feature as an essential, defining trait of apriority, and thus as a necessary characteristic of *a priori* truths. According to the resulting representationist construal of apriority, a certain truth qualifies as *a priori* (i.e. *a priori* knowable) exactly when it consists in the obtaining or absence of some conditions in the realm of representations within a subject's head, while a piece of evidence, or the justification that it provides for a belief that is based on its recognition, or the knowledge achieved by this justification, qualifies as *a priori* exactly when it is generated by an (in principle) observable cognitive mechanism within a subject's head that conveys reliable information of the obtaining or absence of representational truth conditions to the subject's mind.

This work is divided into seven chapters. The first two contain preliminary material for the argumentation developed in the remaining five. The preliminary chapters are meant to clarify the major conceptual and methodological assumptions of the subsequent reasoning. The argumentative part starts with chapter 3, which presents Benacerraf's dilemma and delineates those theoretical options that one might adopt in response to the

dilemma. The remaining four chapters of the work are devoted to the evaluation of the *prima facie* viable response candidates. Chapters 4, 5 and 6 contain my arguments against the standing referentialist responses, while chapter 7 focuses on the specification and defence of the proposed non-referentialist alternative. To conclude this introduction, let me summarise what I shall do in the seven chapters in somewhat more detail.

In chapter 1, I shall put forward the central tenets of this work and elucidate seven key notions whose proper understanding may be essential for a heedful assessment of the position to be advocated here. In section 1, I shall address some issues concerning our idea of *apriority*. In section 2, I shall summarise what I maintain, and will also defend, of the notion of *truth*. In section 3, I shall clarify the sense in which I will talk about *realism* and show how this sense can be retained and communicated in the face of the most recent (quasi-realist) challenge to metaphysical thought. In section 4, I shall explain what I understand by *reference* (and referentialism about truth) in this work. In section 5, I shall specify what I will mean by *abstract* and *natural* referential domains. Finally, in section 6, I shall review what the representationist construal of *a priori* truth and knowledge to be advocated here presupposes in philosophy of mind concerning the existence, the nature, and the semantic content of *representations*.

In chapter 2, I shall turn to the most important methodological assumptions of this work. In section 1, I shall advance those general methodological principles that I think should govern theory formation about any particular segment of the world, and specify what the application of these principles amounts to in the context of the current investigation. The primary purpose of this section is to clarify why I believe that the best way to start an inquiry into the nature of *a priori* truth is to compile a relatively extended list of the most obvious and striking characteristics of truth in our paradigm *a priori* discourses, and then regard the potential to support a reasonable explanation of all these characteristics as a minimal condition of adequacy for an account of *a priori* truth in general. In section 2, I shall put

forward such a list of the major *explananda* for a construal of *a priori* truth. It will be against the background of these explanatory adequacy conditions that, in the remaining five chapters, I shall evaluate the alternative construals of *a priori* truth.

In chapter 3, I shall set the stage for the argumentation advanced in the last four chapters by presenting 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 and then developing an exhaustive doctrinal map of the conceivable responses to this challenge. In section 1, I shall reconstruct an updated and generalised version of Benacerraf's original dilemma about mathematical truth, which demonstrates that, unless the received contact theory of knowledge is false, a substantive realist and referentialist construal of truth in the semantics of discourses that involve knowledge or reliable belief formation about causally inert domains cannot be adequate. In section 2, I shall spell out the most important semantical and epistemological assumptions of this case, and review those theoretical positions that one might take in response to it. Some of these alternatives I shall eliminate as non-starters. The remaining options I shall divide into two groups: four of them I shall classify as *prima facie* plausible referentialist responses, and one as the non-referentialist alternative. Having developed this doctrinal map, in the final part of this chapter, I shall briefly explain my argumentative strategy in the rest of this work.

In chapter 4, I shall argue against the deflationist responses to Benacerraf's dilemma, which attempt to save the idea that truth conditions can always be specified in terms of intended subject matters by denying that our notion of truth represents a substantive property whose nature and metaphysical status can be further characterised, and thus by refraining from a realist interpretation of the truth conditions of our beliefs. My main objection to this strategy will be that an advocate of deflationism can provide no suitable explanation of the objectivity of truth, which means that her construal of *a priori* truth violates at least one major adequacy condition set for such an account in chapter 2. In section 1, I shall develop my case by examining the most

fully elaborated version of deflationism to date, Paul Horwich's minimal theory of truth. Beyond arguing for the claim that a proper explanation of objectivity requires a substantive realist construal of truth, I shall also show that Horwich's positive "use-regularity conception" of meaning together with his commitment to what is expressed by the uncontroversial instances of the equivalence schema 'it is true *that p* if and only if *p*' entails a realist use-theoretic conception of truth that preserves the idea that truth and falsity characterise their bearers independently of what anyone actually thinks or knows of this particular circumstance. A further important aspect of this conception is that it does not stipulate the referential character of truth, and thus remains compatible with the non-referentialist construal of the paradigms of *a priori* truth that I shall argue for in chapter 7. In the second section of chapter 4, I shall defend the received realist correspondence theory of truth against the so-called "slingshot arguments", which are meant to demonstrate that all true beliefs refer to the same thing, and therefore cannot be made true by the obtaining of their own separate truth conditions. As I shall show, the arguments exploit the collapse-generating aspects of our highly coarse-grained (Fregean) notion of reference, which is clearly different from that fine-grained concept of symbols-world relation which is operative in a referentialist construal of truth.

In chapter 5, I shall turn to the conceivable anti-realist responses to Benacerraf's dilemma, which might be thought to be able to save the standard referentialist construal of truth by maintaining that the truth conditions of our beliefs are in some way epistemic, so their obtaining or absence need not causally affect the mind for knowledge to be realised. My main objection to these conceptions will be the same as my objection to their deflationist alternatives: in so far as one abandons realism in the sense clarified earlier, one can provide no suitable explanation of the objectivity of truth, which means that one's construal of *a priori* truth must violate a major adequacy condition set for such an account in chapter 2. In the two extensive sections of chapter 5, I shall defend the realist conception advocated in this work against a number of influential arguments that are often regarded

as anti-realist challenges to this position. In section 1, I shall examine Michael Dummett's acquisition and manifestation arguments against the standard realist assumption that the truth conditions of our beliefs can be verification-transcendent in character. First, I shall show that the real target of Dummett's criticism is not so much the realist as the referentialist construal of truth, so his semantical programme cannot help the advocates of referentialism escape Benacerraf's dilemma. Second, I shall argue that Dummett's cases rely on a limited view of our capacity to develop and communicate new ideas of truth conditions, and therefore cannot demonstrate the inadequacy of standard referentialism in the semantics of discourses about verification-transcendent domains. In section 2, I shall turn to Hilary Putnam's internal realist argumentation against metaphysical realism and the correspondence theory of truth. First, I shall observe that Putnam's internalist perspective cannot help the referentialist escape Benacerraf's dilemma either, since it also embraces realism in the relevant sense of the term, and thus leads the referentialist to the same explanatory difficulties as its metaphysical realist counterpart. In spite of this, I shall admit that Putnam's reasoning is still significant for the concerns of the current work, since it queries the viability of the correspondence theory of truth, something that is clearly endorsed by the adoption of the realist use-theoretic construal advocated in chapter 4. In the second part of this section, therefore, I shall examine Putnam's three sub-arguments for the claim that a metaphysical realist cannot explain how representations can refer determinately to particular aspects of the world, and specify why I think that the argumentation fails to eliminate a broadly causal theory of reference determination. The main tenet that I think the results of this chapter illuminate is that the real problem with referentialism in the semantics of our paradigm *a priori* discourses is not that the advocates of this doctrine cannot explain how we can develop ideas of causally inert domains, but instead that they cannot explain how we can acquire knowledge of, or develop reliable beliefs about, such domains.

In chapter 6, I shall focus on the standing platonist responses to Benacerraf's dilemma. As mentioned before, these responses attempt to save the standard referentialist construal of truth by querying some of the epistemological premises of Benacerraf's case: they either assume that subjects can have a non-causal epistemic contact with platonic domains, or deny that knowledge requires interaction between obtaining truth conditions and the knowing mind. What I intend to show in this chapter is that these responses cannot fulfil their dialectical role either. In section 1, I shall briefly review the most important explanatory considerations in favour and against the platonist construal, and argue that in absence of a viable account of knowledge or reliable belief formation about abstract domains, a platonist may not be able to explain some other striking characteristics of our paradigm *a priori* beliefs either. In section 2, I shall examine the proposed platonist epistemologies, and expound why I think that these accounts cannot save the adequacy of standard referentialism about truth either. Putting it briefly, the non-causal contact theories will be rejected on the ground that they are *ad hoc* and uninformative, and they open the door for parallel stipulations in the case of knowledge claims about any exotic domain, while the no-contact theories will be abandoned because they provide us with no reason for taking anything that occurs in our mind as a reliable indicator of the actual obtaining or absence of the alleged platonic truth conditions of our logical and mathematical beliefs. With the elimination of these epistemological responses, I shall complete my case against the standard referentialist interpretation of the truth conditions of our paradigm *a priori* beliefs. If the conclusions of my three polemic chapters are correct, then the adequate response to Benacerraf's dilemma must be non-referentialist in character.

In chapter 7, I shall show that a specific naturalist version of non-referentialist realism about the paradigms of *a priori* truth can satisfy all major explanatory adequacy conditions set for such a construal in chapter 2. As stated above, I shall call this version a representationist construal of the relevant truths, since it

supposes that the truth conditions of our purportedly *a priori* claims about abstract domains obtain (if they do) in the realm of representations within the subjects' heads. In section 1, I shall observe that our actual cognitive and linguistic practice manifests no analytic link between the notion of truth conditions and the notion of intended referents, which means that a non-referentialist construal of certain truths cannot be rejected on conceptual grounds. In section 2, I shall elaborate the details of the envisaged representationist construal while developing an empirically confirmable naturalist account of the emergence of the semantic content of our paradigm *a priori* beliefs, one of the eleven *explananda* set for a construal of *a priori* truth in chapter 2. In section 3, I shall confront the suggested representationist construal with the other explanatory requirements specified in chapter 2, and explain how I think the construal can satisfy those requirements, and thus qualify not merely as a suitable response to Benacerraf's dilemma, but also as a minimally adequate construal of the relevant truths. Finally, in section 4, I shall briefly return to the original broader perspective of this investigation and argue that by reference to the emphasised feature of representationality philosophers can provide a minimally adequate specification of the nature of *a priori* truth, knowledge, justification and evidence in general as well. At the end of the chapter, I shall put forward the proposed definitions, and highlight what the resulting notion of apriority retains from its past epistemological connotations.

Having finished the argumentation, in the conclusion, I shall briefly review the most important assertions of this work, and elucidate what the collected findings may teach us about some neighbouring issues in the current literature and about the role of empirical inquiry in the development of concepts with major philosophical significance.

CHAPTER 1

Conceptual Preliminaries: Elucidation of Key Concepts

Introduction

In this work, I shall argue for a naturalistic construal of *a priori* truth and knowledge acquisition. The core tenet that I shall defend in the following chapters is that there is a clear semantical contrast between those claims that we typically distinguish as *a priori* and empirical, which can give rise to a substantive account of the traditional epistemological distinction between *a priori* and empirical knowledge, justification and evidence. The way I shall defend this claim is direct and simple: I shall articulate what exactly this semantical contrast consists in by providing a substantive characterisation of the nature of *a priori* truth.

Although many philosophers are sceptical about the existence of *a priori* knowledge and justification, and there is also much debate over the proper construal of apriority, virtually no one refuses the application of the term for expressing substantive positions in epistemology. This indicates that most philosophers have some, more or less explicit, idea of what apriority consists in. My primary aim in this work is to elaborate an account that preserves the categorical distinction between *a priori* and empirical beliefs, largely observes the received application of the contrasted terms, and, together with our best empirical theories of the world and human cognition, explains the most important characteristics of our purportedly *a priori* claims.

The construal that I shall propose is based on the theoretical hypothesis that the clearest examples of *a priori* claims or beliefs, our claims or beliefs in pure logic and mathematics, are attitudes to truth-apt representations of abstract domains, whose truth

value is nevertheless determined by the obtaining (or absence) of some natural conditions in human heads, in particular, some relations among our representations, which we develop in the course of our cognitive engagement with our direct natural environment. In what follows, I shall call this proposal *a representationist construal* of the relevant *a priori* truths. Although I shall acknowledge the constitutive role of some epistemic facts or events in the emergence and existence of these truths, nevertheless I shall maintain that our actual knowledge of the obtaining of the suggested conditions is definitely not among these facts. Accordingly, my proposal will be, in a sense specified below, realist about these truths.

An important semantical implication of the above hypothesis, which constitutes the second major tenet of this work, is that truth within our paradigm *a priori* discourses (that is, again, in pure logic and mathematics) cannot be referential in character. In other terms, in contrast to the case of our empirical beliefs about the obtaining of some causally effective spatiotemporal conditions, the truth conditions of our logical and mathematical beliefs, as well as those of our presumably justified beliefs about causally inert entities in general, cannot be adequately specified in terms of the relevant intended subject matters.[1] In what follows, I shall call this semantical tenet *a non-referentialist construal* of the relevant *a priori* truths.

After presenting my arguments for the above hypotheses about the cited paradigms of *a priori* truth, in the last section of chapter 7, I shall suggest that being representational might be more than a contingent characteristic of some *a priori* truths. In fact, by reference to this property we can provide a useful and adequate real definition of *a priori* truth in general. According to this construal, a certain truth is *a priori* if and only if the conditions whose obtaining it actually consists in are representational (i.e. obtaining in the natural realm of

[1] In the semantics of pure logic, this tenet applies if it is supposed that logical beliefs have platonic subject matters (e.g. Fregean thoughts or propositions with inferential relations among each other) or have no subject matter at all.

representations within human heads). In harmony with this definition and the existing conceptual link between our notion of *a priori* truth, on the one hand, and our notions of *a priori* knowledge, justification and evidence, on the other, we can stipulate also that a piece of evidence, or the justification that it provides for a true belief that is based on its recognition, or the piece of knowledge that we acquire by this justification, is *a priori* if and only if it is generated by an appropriate causal mechanism within human heads that conveys information of the obtaining or absence of truth conditions within a subject's system of representation to her knowing mind.[2]

Before starting my reasoning, however, in this chapter, I wish to say a few words about the intended meaning of the central terms of the previous proposals. In section 1, I shall start with some remarks on the notion of *apriority*. The main question that I shall address here is whether our understanding of this notion in terms of independence of experience indeed amounts to a purely epistemological distinction, or rather it relies on further contrasts that are best classified as ontological or semantical in character. Another question that I shall raise in this section is whether there is a reasonable way to start an inquiry into the nature of *a priori* truth, knowledge, justification and evidence without having a sharp and commonly shared notion of exactly which truths, beliefs, justifications or epistemic grounds are *a priori* in character.

In section 2, I shall turn to the second key notion of this work: the intended concept of *truth*. Issues about truth take centre stage in present-day analytic philosophy. What are the entities that this concept primarily and derivatively applies to? Does it represent a substantive property of these entities, or it is merely a useful logical device? If it represents a property, then can we say something about the nature and metaphysical status of that

[2] An important consequence of this construal of apriority is that claims *about* the obtaining or absence of these truth conditions in the actual world will also qualify as *a priori*, even if *their* truth conditions remain to be understood in referential terms.

property, or it is a primitive feature that cannot be further characterised? If there is something more to say about it, then what exactly should we say? Are there epistemic properties that are constitutive of the property of truth? Is anyone's knowledge of some truths constitutive of the existence of those truths? If truth is not an epistemic property, then what can be said about the nature of those conditions whose obtaining or absence determines the truth value of the relevant truthbearers? Can that obtaining or absence be inaccessible to the human mind, and if so, then in what sense of inaccessibility? Can the conditions in question obtain necessarily? Can they obtain outside the spatiotemporal world? Can they obtain in a world that is not actual? And how does truth relate to the semantic content of its bearer and the bearer's more basic representational constituents? Finally and, from the perspective of this work, most importantly, how does truth relate to the existence of the intended referent or subject matter of its bearer and the bearer's constituents? In the following chapters, I shall attempt to answer each of these questions in more or less detail. In this chapter, I shall merely present a brief sketch of these answers, which may provide the reader with a preliminary grasp of what I shall argue for somewhat scattered in the remaining part of this work.

In section 3, I shall focus on the third key notion of this work: the adopted concept of *realism*. First, I shall offer an explicit characterisation of what I mean by realism, a specification that will preserve the concept's primarily metaphysical character. On the other hand, I shall recall two reasons for which the resulting metaphysical doctrine concerning a certain object, property, condition or domain may become a semantically significant conception as well. Afterwards, I shall contrast the specified notion of reality with the idea of objectivity, and emphasise that in this work objectivity will be invoked as a major *explanandum*, while reality as an indispensable *explanans*. Having clarified this, I shall briefly examine the quasi-realist observation that a thoughtful anti-realist about a certain domain can to a large extent mimic the linguistic practice of her realist opponent while specifying her view of the semantic features of her

representations of this domain. In contrast to those who believe that the correctness of this observation undermines the communicability or intelligibility of the metaphysical doctrines denoted by the terms 'realism' and 'anti-realism', I shall argue that not only can we understand the alternative senses in which realists and quasi-realists can apply the same words while explicating the content of their own semantical commitments, but there are even ways in which we can communicate these senses, and thus express our metaphysical views in terms of a public vocabulary in a sufficiently transparent manner. Finally, I shall distinguish two senses in which one can adopt realism concerning a given discourse (and thus become, for instance, an 'ethical realist' or a 'mathematical realist'), and observe that adopting realism in the one sense is a logically independent commitment from adopting realism in the other.

In section 4, I shall clarify the sense in which I intend to talk about *reference* and *referentialism about truth* in this work. First, I shall endorse what virtually everyone assumes about reference: the idea that it is an asymmetric "word-world" relation between meaningful symbols, on the one hand, and what the users of these symbols intend to think of or speak about while applying these representations in particular contexts in the actual world, on the other. Second, I shall recall a number of alternative construals of this relation, with different implications concerning the exact circle of referring entities, and emphasise that the notion adopted in this work is a relatively inclusive and deflated one, which acknowledges the referential power of definite descriptions, adjectives, adverbs and other meaningful expressions as well, and presupposes neither the reality nor the actual existence of these symbols' intended subject matters or referents. Third, I shall highlight the essential role of our referential intentions in the determination of the nature and identity of actual referents, and explain why I shall largely neglect in this work those conceptions that attempt to save the received referentialist construal of truth in our paradigm *a priori* discourses by adopting a revisionist view about the nature of the relevant intended subject matters. Finally, I shall observe that the suggested deflated notion of reference

stands for a very fine-grained asymmetric "word-world" relation, which must not be conflated with its relatively coarse-grained Fregean counterpart. With these observations in mind, I shall conclude that my primary opponents in this work will be those philosophers who believe that an adequate theory of truth can specify the truth conditions of every truth-apt mental or linguistic symbol in terms of those conditions or entities that the constituents of this symbol refer to in the specified deflated, essentially intentional and fine-grained sense of the term.

In the short characterisation I gave of the construal of *a priori* truth, knowledge, justification and evidence that I should defend in this work, there were three further terms whose meaning may call for preliminary clarification. In section 5, I shall focus on two of them, and specify the sense in which I intend to speak about *natural* conditions, objects, properties and domains, on the one hand, and *abstract* conditions, objects, properties and domains, on the other.

Finally, in section 6, I shall turn to the last crucial constituent of that characterisation, and elucidate the sense in which I shall speak about *representations* in the following chapters. The primary aim of this section is to provide a clear picture of what the suggested representationist construal of *a priori* truth and knowledge presupposes concerning the existence, the nature, and the semantic content of representations. In particular, there will be six major assumptions emphasised and discussed in this part. First, I shall grant that the construal presupposes the existence of both mental and physical representations, vindicate the first half of this commitment against the eliminativist challenges advanced in the philosophy of mind, and finally note that the core message defended in this work could be formulated in a way that is compatible with an eliminativist stance towards mental representations as well. Second, I shall notice that the construal also assumes the compositionality of semantic content, which is often believed to run counter to a connectionist picture of the computational architecture of the mind. In contrast to the latter conviction, I shall adopt the opposite view and maintain that compositional semantic contents can emerge in a connectionist

computational system as well. Third, I shall register that the construal takes the (truth-related part of the) semantic content of our mental and physical symbols to be explicable in terms of the correct declarative use conditions of these representations. Nonetheless, I shall hasten to add that this (broadly) use-theoretic conception of semantic content is meant to be the denial of neither a truth conditional account of meaning nor a realist construal of truth and correct use in general. In fact, as I shall emphasise, the construal assumes both that a truth-apt representation's correct declarative use conditions (in a non-embedded state) are identical with the truth conditions of this representation, and that these conditions are to be understood along the realist lines. Fourth, I shall point out (here again) that the construal abandons the standard referentialist assumption that the declarative use conditions of our mental and physical symbols can be understood in each representational context in terms of those objects and properties that these symbols purport to be about. Fifth, I shall remark that the construal also assumes the existence of narrow semantic contents, and then explain how these contents can be distinguished from each other and the syntactic objects that they are associated with. Finally, I shall observe that the construal distinguishes between the possession and the actual entertainment of representations, and assumes that the truth conditions of *a priori* beliefs obtain (if they do) within the domain of representations possessed in the thinkers' heads. As I shall argue, this assumption leaves room for a realist construal of *a priori* truth. On the other hand, it leaves open whether the domain of representations invoked in this construal can be rightly classified as mental in character.

1. Apriority

The concept of apriority has had a long life in the history of philosophy. In the 17th century, the distinction between *a priori* and *a posteriori* knowledge became a central contrast within epistemology. Rationalist thinkers such as Descartes, Spinoza and

Leibniz argued that real knowledge must be based on reason rather than on experience. Their empiricist opponents heavily disputed this claim. According to them, if there are some truths that can be known independently of experience, they can at most concern the relations of our ideas, rather than the facts of a mind-independent world.

Present-day discussions of apriority are fundamentally shaped by Kant's famous exposition of the *a priori* and *a posteriori* distinction in his introduction to the *Critique of Pure Reason*. Following his rationalist and empiricist ancestors, Kant construed apriority in epistemological terms. According to his definition, a judgement is *a priori* if and only if its justification does not require reliance on sensory experience.[3] The most important novelty of his exposition was the specification of the relation of this contrast to two other (non-epistemological) distinctions, which can (at least derivatively) also characterise human judgements. The first of these is the metaphysical contrast between necessary and contingent truths. Although, according to Kant, our concept of necessity is not analytically related to our concept of a priority, the two concepts are nevertheless extensionally equivalent. *A priori* judgements are necessarily true or necessarily false, while *a posteriori* judgements express contingent truths or falsities. Consequently, necessity can be regarded as a sign of apriority.[4] The second distinction examined by Kant is the semantical contrast between analytic and synthetic judgements. In his construal, analytic judgements are those whose predicate concept is wholly contained within their subject concept, while synthetic judgements are those whose predicate concept amplifies the subject concept. According to Kant, the latter distinction is not coextensive with the one between *a priori* and *a posteriori* claims. While the possibility of *a posteriori* analytic judgements is excluded by the fact that the detection of containment relations among our concepts never relies on sensory experience, some synthetic

[3] Kant (1781/1787), 136.

[4] As is known, beyond necessity, Kant took strict universality as a sign of apriority too. Kant (1781/1787), 137-138.

judgements are clearly necessary, and therefore *a priori* in character. Kant's famous programme in his first *Critique* was to show how these *a priori* synthetic judgements are possible at all.[5]

The programme set the stage for subsequent discussions of *a priori*, and it keeps shaping our ideas of the subject even today.[6] One major issue in contemporary literature is the proper delineation of the concept of apriority. The dominant view today is that apriority is an epistemological concept which is meant to stand for a property primarily attributed to types of evidence or ways of justification. Derivatively, the property in question can be attributed to pieces of knowledge, beliefs, propositions, judgements, sentences, utterances and truths due to the relevant epistemological features of these entities.[7] Sometimes arguments and concepts are also said to be *a priori*. Arguments are regarded as *a priori* if and only if all of their premises are *a priori*, while an *a priori* concept is meant to be one that can be acquired independently of experience. It is commonly understood that an epistemological notion cannot be adequately analysed in purely non-epistemic terms. Most construals of apriority, therefore, include an epistemic concept in the proposed *analysans*. Of course, this policy will not guarantee that the notion analysed becomes clearly disentangled from all non-epistemological connotations. Moderate empiricist construals of *a priori* knowledge, for instance, developed in the era of post-Fregean analytic philosophy usually assumed that the features of analyticity and necessity were essentially linked with the epistemological feature of apriority.[8] Despite the subsequent efforts to keep the epistemological distinction from other (semantical, metaphysical or ontological) distinctions apart, the

[5] Kant (1781/1787), 141-143, 146-148.

[6] Recent works on *a priori* include Hanson and Hunter (1992), BonJour (1998), Bealer (1999), Boghossian and Peacocke (2000), Casullo (2003) and Horowitz (2006).

[7] Some philosophers do not share this view. For instance, as Casullo points out, Quinton (1972) construes apriority as a concept to be analysed in non-epistemic terms. Casullo (2003), 12-13.

[8] Ayer (1946), 16-18.

inclusion of non-epistemic notions in the analysis of the concept of apriority is still prevalent in present-day epistemology.[9]

In this work, I shall adopt the original Kantian (and currently dominant) idea that the distinction between *a priori* and *a posteriori* beliefs concerns the way in which judgements, beliefs, utterances, premises and conclusions, are justified, or in which concepts are acquired in the course of a subject's cognitive development. In particular, I will suppose that a concept or the truth value of a certain truthbearer is knowable *a priori* if and only if it can be, respectively, acquired or justified independently of experience, and that this equivalence holds because the concept of apriority is to be analysed in terms of the notion of *independence of experience*.[10] My primary aim in this section is to examine whether this *prima facie* epistemological construal of apriority is indeed independent of notions that are usually classified as ontological or semantical in character.

In order to see how, following this construal, one could still consistently hold that the difference between those beliefs that are supposed to be *a priori* and those that are not might be partly ontological or semantical, rather than purely epistemological,

[9] For instance, BonJour argues that "a proposition is justified *a priori* when and only when the believer is able, either directly or via some series of individually evident steps, to intuitively 'see' or apprehend that its truth is an invariant feature of all possible worlds, that there is no possible world in which it is false". BonJour (1985), 192. Further analyses of the notion of apriority that include a non-epistemic concept in the *analysans* can be found in Chisholm (1989), 28, and Plantinga (1993), 105. For a critical discussion of these proposals, see Casullo (2003). Suggested examples of *a posteriori* necessary truths include the propositions *that water is H_2O* and *that Hesperus is identical with Phosphorus* while a proposed case of *a priori* contingent truths is the proposition *that the standard meter bar in Paris is one meter long*. Kripke (1972/1980), Putnam (1975c), and Kitcher (1980).

[10] The acquisition of a certain concept or the justification of a certain belief, in turn, is taken to be independent of experience if and only if the evidence on which the development of the concept or the establishment of the belief relies is not experiential in character. Beyond this negative characterisation, purely or partly epistemological analyses of apriority often include some positive claims about how *a priori* propositions *are* knowable or justifiable. According to BonJour, for instance, apriority means *being based merely on the deliverances of reason or rational insight*. BonJour (1998), 11.

consider what fixes the extension of the intended concept of being independent of experience. Supposing that the semantic content of a complex piece of representation is determined by that of its components, we can clarify what independence of experience means by elucidating the content of the notions of independence and experience.

As regards the first of these concepts, there are at least two important senses in which *a priori* justifications, as opposed to *a priori* concept acquisitions, are not meant to be independent of experience. First, a judgement cannot be justified, unless its constituents have been understood by the epistemic agent. As Kant already observed, however, *a priori* judgements can involve empirical concepts, which could not be acquired and understood independently of experience. If this is true, then apriority, in so far as it is applied to justifications, cannot presuppose our understanding of the relevant truthbearers independently of experience.[11] Second, there are philosophers who maintain that *a priori* beliefs can be undermined by experience. If this is right, then apriority cannot consist in immunity to experiential disconfirmation either. In order to ensure the greatest possible consensus over the intended notion of independence in the analysis of *a priori* justification, the best way to proceed is to confine its significance to the sources of positive justification. According to this construal, the justification of a certain belief is independent of experience if and only if it can be confirmed without reliance on experience.

Received interpretations of experience differ from each other in how inclusive they are concerning the notion's extension. According to the narrowest construal, the concept merely denotes perceptual experience of the external world. The

[11] Apriority, applied to concepts, rather than justifications, consists in independence of experience in this stricter sense. Propositions, knowledge claims, beliefs, judgements, sentences, utterances and truths can be *a priori*, derivatively, in two different senses: they can be *a priori* in a looser sense, due to their justifiability independently of experience, and in a stricter sense, due to their understandability independently of experience. Kant calls this stricter sort of apriority *pure* apriority. Kant (1781/1787), 137.

intermediate understanding includes perceptual experience of the subject's own bodily states as well. Finally, on the broadest interpretation, the notion applies to any conscious state or event including, beyond the deliverances of the external and internal senses, any object of reflexive attention.[12] Which, from among these alternative senses, is the one that might be constitutive of the intended notion of *a priori* knowledge and justification? Although the paradigm cases of empirical claims are taken to rely on experience in the narrowest sense of the term, the extension of *a posteriori* knowledge was never confined to justified true beliefs about the external world. Following Kant's classification, judgements relying on introspective evidence (i.e. the deliverances of our inner senses, have also been regarded as empirical).[13]

One may wonder, however, what motivates this terminological convention. In other words, why do philosophers not tend to define apriority in terms of (independence of) experience merely of the external world? The most plausible answer to this question is arguably that the resulting notion of apriority, which would apply to introspective evidence and justification as well, would serve the theoretical purposes of epistemology less than a notion that renders introspection an *a posteriori* source of evidence. Such theoretical purposes can be to have a distinction that coincides with the one between infallible and fallible beliefs, or between epistemologically more and epistemologically less fundamental classes of beliefs, or maybe between more and less reliable types of belief-forming mechanisms.

Beyond these purely epistemological considerations, one may have ontological or semantical motivations as well for refusing the above inclusive construal of apriority. First, one may observe that introspection resembles perception of the external world more than the epistemic faculty used in pure logic and mathematics, in so far as its deliverances support beliefs about contingent and natural, rather than about necessary and abstract,

[12] Boghossian and Peacocke (2000), 2.
[13] Kant (1781/1787), 155-171.

states of affairs.[14] Second, one may find introspection more similar to perception than to logical and mathematical intuition also because the latter faculty, in contrast to the former two, apparently transcends the confines of the natural world.[15]

Now, clearly, this second sort of motivation is heavily loaded with substantive theories of the truth conditions of the contrasted empirical and *a priori* beliefs. A naturalistic construal of *a priori* truth, for instance, would certainly undermine the above difference between perception of the external world and introspection, on the one hand, and logical and mathematical intuition, on the other. The conclusion that I shall draw from these observations is that the adopted negative construal of apriority in itself does not guarantee the purely epistemic character of the resulting concept. The crucial question to be answered remains whether or not our notion of experience can be construed in a purely epistemic way. In other terms, what we should clarify is whether there is a purely epistemic contrast between what we commonly regard as (a type of) experience and what we usually classify as (a sort of) *a priori* evidence. If the answer to this question is negative, then our concept of apriority will hardly qualify as purely epistemic either.

There are similar considerations to the effect that the intended notion of experience appearing in the analysis of *a priori* justification cannot be the broadest one either. Defining apriority in terms of being justified independently of any conscious state or event would render the paradigm cases of *a priori* justification, such as justifications of our mathematical and logical claims,

[14] In section 5 below, I shall specify the sense in which I am to apply the terms 'natural' and 'abstract' throughout this work.
[15] According to BonJour, "the relevant notion of experience should be understood to include any sort of process that is perceptual in the broad sense of (a) being a causally conditioned response to particular, contingent features of the world and (b) yielding doxastic states that have as their content putative information concerning such particular, contingent features of the actual world as contrasted with other possible worlds. […] And thus not only sense experience, but also introspection, memory, kinaesthesia, and clairvoyance or telepathy (should these exist) would count as varieties of experience". BonJour (1998), 8.

trivially empirical. The question, again, is why we should resist the radical empiricist proposal that even these paradigm cases of *a priori* justification are, in fact, empirical. One sort of motive may be, again, intrinsically epistemological: we may hold that in some genuinely epistemological respect the justification of, say, our mathematical or logical beliefs is different from the justification of our perceptual and introspective beliefs. Other motives may be ontological or semantical: we may observe that, due to some facts concerning the relevant truth conditions, the nature of the former class of justification is categorically different from that of the latter. Be that as it may, as before, we may conclude that in absence of a purely epistemic difference between what we commonly regard as (some type of) experience and what we usually classify as (a sort of) *a priori* evidence our concept of apriority can hardly be regarded as purely epistemic in character.[16]

The interesting debate between radical empiricists, who deny the existence of *a priori* knowledge, and apriorists, who maintain that some beliefs can be justified independently of experience, is, clearly, not terminological. Most apriorists would probably concede that there might be some similarities between what *they* call *a priori* and empirical justifications. One of these features might be that both kinds involve conscious events within the subject's mind. No apriorist would ever want to claim that this feature could not be taken as constitutive of aposteriority. On such construal, they would probably admit that all knowledge is empirical. Still, they would insist that if we compared our beliefs about various domains, then we could find a theoretically important difference between those that they want to call *a priori*, on the one hand, and those that they propose to regard as empirical, on the other. Many of them might even insist that the difference in question is purely epistemological in character.

[16] In so far as we maintain that the distinction between *a priori* and *a posteriori* beliefs concerns the way in which judgements, beliefs, utterances, premises and conclusions are justified, or in which concepts are acquired in the course of a subject's cognitive development, our notion of apriority will remain epistemic in a trivial sense of the term.

Similarly, most radical empiricist would probably concede that there might be some contrasts between what *their opponents* call *a priori* and empirical justifications. One of these contrasts might be that the former support beliefs about non-natural states of affairs, while the latter provide warrant for claims about the natural world. No radical empiricist would ever want to claim that this contrast could not be taken as constitutive of apriority. On such construal, they would probably admit that some knowledge can be regarded as *a priori*. Nevertheless, they would insist that, from an epistemological point of view, there is no important difference between the contrasted kinds.

In the following chapters, I shall argue that there is a perfectly good sense in which *a priori* knowledge, justification and evidence can be distinguished from their empirical counterparts. Nevertheless, the distinction that I shall propose will invoke a contrast between the nature of the truth conditions of our *a priori* and empirical beliefs. By reference to this distinction, I shall argue that there must be a corresponding contrast between the nature of those epistemic mechanisms by which we can detect the obtaining or absence of these conditions in the actual world, a contrast whose specification is the task of our empirical sciences.

An inquiry into the nature of *a priori* truth, knowledge, justification and evidence, of course, presupposes that one can (at least vaguely) identify the intended subject matter of this investigation. In view of the above disagreements concerning the extension of our concept of apriority, the best strategy that we can adopt here is to examine the least contentious "paradigm" instances of the contrasted types, develop an adequate proposal of the defining characteristics of the relevant classes on this commonly accepted evidential ground, and then consider whether the more controversial instances, initially set aside, can be classified as *a priori* or as empirical by means of the newly articulated conceptual resources.

In the following chapters, I shall adopt the foregoing strategy too. First, I shall suppose that our perceptual claims about our direct spatiotemporal environment are the least contested examples of empirical representations, while our claims

in pure logic and mathematics are the clearest instances of *a priori* truthbearers. Corresponding to this classification, I shall also assume that perceivable truths, pieces of perceptual knowledge, perceptual justifications and pieces of perceptual evidence are the paradigm cases of empirical truth, knowledge, justification and evidence, while the respective counterparts of these entities in pure logic and mathematics are what we usually regard as the clearest examples of *a priori* truth, knowledge, justification and evidence.

In view of the observable specificities of these paradigms, then, I shall argue that from among the numerous contrasts that obtain between the relevant types of entities there is one that might be invoked in an adequate real definition of apriority as well. Again, the contrast in question will concern the nature of those conditions whose obtaining or absence determines the truth value of the relevant representations, and therewith the nature of those epistemic mechanisms by which we can detect the previous facts in the actual world.

In possession of the proposed analysis of apriority, at the end of chapter 7, I shall finally consider what the suggested construal implies concerning the classification of representations and epistemic grounds beyond the examined paradigms. As we shall see, the account will ratify the *a priori* status of our analytic claims in general, and it will delineate the possible content of synthetic *a priori* knowledge of the actual world as well.

2. Truth

Developing a proper understanding of truth is one of the highest priorities of present-day analytic philosophy. In this section, I shall provide a brief sketch of the account that I am to develop of this subject in greater detail in the remaining part of this work.

First, I intend to say a few words about the primary bearers of this property. In what follows, I shall take truth to be primarily a property of truth-apt mental and physical representations, such as thoughts and their various linguistic expressions, rather than a

property of objectified semantic contents, such as abstract propositions.[17] By this assumption, I do not query that our syntactically identifiable mental and physical symbols are true (or false) partly in virtue (of a certain segment) of their semantic contents (namely, their truth conditions). What I deny is that, because of this role, semantic contents also qualify as bearers of truth and falsity. Having said this, I shall further recognise that, beyond the above fundamental applications, the semantic values under scrutiny can be, derivatively, predicated of some mental attitudes and physical acts (e.g. beliefs, judgements, claims and utterances) involving the relevant primary bearers as well. A common feature of these secondary bearers is that they can be equally regarded as (purportedly correct) declarative applications of the primary bearers involved. This fact fits well with the fifth major tenet mentioned below, according to which truth is identical with the correct unembedded declarative applicability of our truth-apt representations in the context of their actual use.

Second, contrary to deflationist theories of truth, I will maintain that our actual notion of truth is not merely a logical device, but it stands for a substantive property whose nature and metaphysical status can be further characterised. My reason for adopting this view is explanatory in character, and it will be presented in detail in chapter 4, where I examine, among others, the explanatory adequacy of the most elaborated current form of deflationism, Paul Horwich's minimal theory of truth.

Third, beyond committing myself to a substantive construal of the notion of truth, I shall further assume that the property represented by this concept characterises its bearers independently of what anyone actually thinks of this particular circumstance. Note that my claim is not that the truth (or falsity) of a certain bearer can never be the result of our epistemic activities. In fact, as I shall argue in chapter 7, the truth of our *a priori* beliefs is always a product of our epistemic activities. In particular, it presupposes that we have developed certain suitably

[17] In section 6, I shall say more about the ontological and semantical assumptions underlying my claims about alternative sorts of representations.

connected representations in our heads. The current point is rather that these activities do not include anyone's actual recognition or awareness of the relevant truths (or falsities). In a specific sense, which I shall clarify below, this tenet amounts to the endorsement of a general realist construal of truth, which reinforces the view that truth is a correspondence between what is believed, on the one hand, and what actually obtains in the world, on the other.[18] The explanatory consideration I present against deflationism in chapter 4 will reveal my main reason for adopting this general realist construal of truth as well. In chapter 5, I shall examine the most influential "anti-realist" objections to this view, and show that none of them provides us with good reason for revising the previous commitment.

Fourth, I shall suppose that truth, at least when it is knowable for a natural subject, is a natural property that characterises its bearers in virtue of two separate facts: first, the fact that these bearers stand in a certain semantic relation with some conditions that may or may not obtain in the spatiotemporal world; and second, the fact that the conditions in question actually obtain in the spatiotemporal world. In the case of our beliefs about abstract states of affairs, this tenet comes to grips with the standard referentialist construals of truth, according to which the truth or falsity of a certain belief is always to be understood in terms of (the obtaining or absence of) those conditions that the belief in question purports to be about. My reason for adopting this naturalist tenet is also explanatory in character. In particular, I believe that its opponents cannot explain all observable characteristics of (what we take to be) the paradigms of *a priori* knowable truth. Deflationist and anti-realist versions of anti-naturalism will be rejected in chapters 4 and 5 as

[18] It must be noted, however, that the standard referentialist form of the correspondence theory will be rejected here together with the standard referentialist construal of truth. Instead, what will be endorsed is that truth is correspondence between what is said or believed and the obtaining truth conditions of what is said or believed, regardless of whether or not the latter conditions are identical with what the claims or beliefs in question purport to be about.

variants of deflationism or anti-realism about truth. In chapter 6, in contrast, I shall argue against those who defend a realist (i.e. platonist) form of anti-naturalism in the semantics of our paradigm *a priori* discourses. Finally, in chapter 7, I shall show that the proposed naturalistic constraint is compatible with an adequate construal of truth in the semantics of our paradigm *a priori* discourses as well.

Fifth, I shall understand truth as a property that is identical with that of the correct unembedded declarative applicability of its primary bearers in the context of their actual applications.[19] Accordingly, I will assume that the truth conditions of a given representation are identical with those conditions whose obtaining or absence is supposed to govern the correct declarative use of this bearer in the context of its actual applications. The correct declarative use conditions of truth-apt representations will be supposed to be determined by the correct declarative use conditions of their basic constituents and the mode of their composition. The correct declarative use conditions of concepts and their linguistic expressions (i.e. the basic constituents of truth-apt representations) will in turn be identified with those conditions whose obtaining or absence is supposed to govern the correct applicability of these constituents in the context of alternative truth-apt representations. My reason for adopting this strictly realist form of a broadly Wittgensteinian use-theoretic semantics is also explanatory in character, and it will be articulated in chapter 4, in the course of my discussion of Horwich's deflated version of this use-theoretic account.

Finally and (from the perspective of this work) most importantly, in what follows, I shall abandon the standard referentialist construal of truth, the idea that the truth conditions of our truth-apt representations are always to be understood in terms of those conditions that these bearers purport to be about.

[19] By invoking contexts of actual applications, I wish to ensure the explanatory resources for a suitable account of indexicality, lack of truth value and other semantic phenomena emerging from the context dependence of the correct declarative applicability of certain truthbearers.

Contrary to this referentialist understanding, and thus also to Tarski's highly influential semantic theory of truth, I shall maintain that referentiality is not encoded in our general notion of truth.[20] Moreover, if my explanatory considerations in support of the naturalistic construal of knowable truths are correct, then in the case of our paradigm *a priori* discourses about abstract domains standard referentialism cannot even be true. It is not that I query our capacity of developing truth-apt representations of abstract domains whose truth conditions are referential in character. What I shall argue for is merely that our purportedly *a priori* claims in pure logic and mathematics are not such representations. They are not, because they are all analytic, and the truth value of our analytic claims is determined by the obtaining or absence of some conditions in the domain of representations within our heads, rather than by the obtaining or absence of those states of affairs that these claims purport to be about. The truth values of those claims whose truth conditions are rightly supposed to obtain (or fail to obtain) in an abstract domain, on the other hand, will be regarded as necessarily inscrutable to human minds.

Of course, the above construal of truth (and *a priori* truth in particular) can be accepted only if it satisfies all those adequacy conditions that we can reasonably set for a proper theory of these subjects. In chapter 2, I shall say more about these conditions, and in the rest of this work, I shall show that the construal specified above is indeed capable of satisfying them all.

3. Realism

For a long time in the history of philosophy, realism has been understood as a doctrine of the metaphysical status of various entities. Although the term did not lose its original metaphysical connotations, in the late 20th century it became common to

[20] Tarski (1944).

construe realism in semantical terms. Michael Dummett, for instance, argued that the *prima facie* metaphysical disagreement between realists and anti-realists concerning the status of certain problematic domains can be best understood as a debate about the applicability of classical logic within the discourses about these domains. Hilary Putnam also described the doctrine of metaphysical realism as the adoption, in the semantics of the relevant discourses, of a correspondence theory of truth.[21] The adequacy of these and other semantical construals has been heavily debated in contemporary literature. Although a proper discussion of the arguments for and against these understandings is beyond the reach of the current section, a brief characterisation of how I shall use the term in the following chapters might be appropriate.[22]

In what follows, I shall adopt the traditional view and regard realism and anti-realism as doctrines about the metaphysical status of various thinkable entities or domains. In particular, I shall suppose that realists about certain subjects maintain that the subjects in question obtain or exist (if they actually do) independently of whether or not anyone ever does or can think or know of this circumstance. Anti-realists, in contrast, will be supposed to query either the possibility of the obtaining or existence of the relevant entities or the independence of this obtaining or existence of anyone's actual or potential thought or knowledge of it.[23]

[21] Dummett (1963) and Putnam (1981).

[22] An influential defence of the idea that realism is a purely metaphysical doctrine, the correctness of which is entirely independent of our conceptions of truth and valid inference, can be found in Devitt (1984). In chapter 5, I shall examine Dummett's and Putnam's conceptions in detail, and argue that neither of these authors endorses anti-realism about truth in the sense specified below.

[23] The construal does not render realism about conscious mental states inconsistent. There is nothing inconsistent in the claim that our conscious mental states can occur in the actual world independently of anyone's actual thought of this circumstance. Fictive objects and objects "existing" only *within* a conscious mind, on the other hand, cannot be consistently regarded as real. Depending on how one applies the term 'existence', one must either deny the possibility of these objects' existence or maintain that their existence relies on someone's actual or

The second thing that I shall assume here concerning realism is that the adoption of this doctrine concerning a certain entity implies no commitment to the actual obtaining or existence of this entity. What the realist assumes is merely that the entity can obtain or exist in the actual world, independently of anyone's actual or potential thought or knowledge of this circumstance. Thus, for instance, one might adopt a realist (i.e. platonist) construal of some abstract entities without committing herself to the actual existence of those entities, just as much as someone might adopt a realist construal of some spatiotemporal entities without committing herself to the actual existence of those entities. It is one thing that my idea of Napoleon is an idea of a real object, and another that I believe in the (past) existence of this object in the actual world. If my realist construal of Napoleon implied my belief in his actual existence, then I could not consistently entertain the thought that, despite appearances, Napoleon did not exist at all.

The third point that I wish emphasise here is that in this work the concept of reality will be distinguished from the concept of objectivity. Although the two concepts are often interchangeably applied, in what follows, I shall suppose that the question whether or not a certain entity is real concerns the metaphysical status of this entity, while objectivity is a feature of the correct declarative applicability of mental and physical symbols.[24] In particular, I shall say that the declarative applicability of a certain symbol is objective if and only if the symbol's use is subject to a certain sort of evaluation (i.e. it can be correct or incorrect), and no opinion of this issue (i.e. no one's

potential thought or awareness of this circumstance. Eliminativism about a certain domain exemplifies anti-realism stated in terms of the first vocabulary, while phenomenalism or idealism concerning a class of entities exemplifies anti-realism understood in terms of the second.

[24] In a derivative sense, the feature can be predicated of the relevant symbols or their particular applications as well. Supposing the objectivity of mathematical truths and falsities, one can also say that the mathematical representations expressing those truths and falsities, or their actual applications by us (i.e. our claims and beliefs), are objective.

actual or potential judgement of this correctness) is conceptually prevented from being mistaken. A certain claim made by a certain subject in a certain context, for instance, will be understood as a particular declarative application of a certain truth-apt representation in that context, which is subject to the relevant kind of evaluation (it can be correct, i.e. true, or incorrect, i.e. false). To say that the truth value of this claim is objective is to maintain that no one's actual or potential opinion of the correctness of this (or any other) application of the relevant representation is conceptually prevented from being mistaken.[25]

Objectivity is a relatively transparent feature of the applicability of our mental and physical symbols. In most cases, we can easily decide whether anyone's actual or potential use of a certain symbol is constitutive of how the symbol in question has to be applied (in which case the relevant subject's cognitive or linguistic practice would be conceptually prevented from being mistaken). In contrast, our view about the metaphysical status of various entities or domains depends on highly theoretical explanatory considerations. For instance, hardly any philosopher of mathematics would query today the objectivity of mathematical truths. In contrast, the same thinkers are rather divided concerning the metaphysical status of mathematical objects and properties. In what follows, I shall observe that objectivity is a general (i.e. discourse-independent) feature of the applicability of our truth-apt representations (and thus a general feature of truth and falsity), I shall regard this feature as a major *explanandum* for a theory of any sort of truth, and finally, in chapter 4, I shall argue that a proper account of this *explanandum* requires a substantive realist construal of the truth conditions of our truth-apt representations. Putting it briefly, in this work, I

[25] The fact that the use of a certain meaningful symbol is never conceptually prevented from being mistaken, of course, does not exclude the possibility of perfect applications, or in other terms, the occurrence of an ideal thinker or speaker, who never fails to know whether or not the declarative use conditions of the symbols applied by her obtain in the actual world.

shall invoke objectivity as a major *explanandum*, while reality as an indispensable *explanans*.[26]

Despite the previous endorsement of the traditional metaphysical construal of the contrast between realism and anti-realism, I shall add that there can be at least two reasons for which realism about a certain domain can also be regarded as a genuine semantical position. First, since the metaphysical distinction at hand concerns thinkable conditions, objects, properties and domains that can be identified as subject matters of particular thoughts and beliefs, it is always possible to regard realism about these entities indirectly as a conception of the subject matters or purported referents of those thoughts or beliefs. Second, in so far as the contrasted metaphysical doctrines concern a conceivable domain whose states of affairs can be the truth conditions of certain thoughts and beliefs, the doctrines can, again, be equally regarded as components of a conception of truth in the semantics of discourses involving those thoughts and beliefs. Note, however, that even if these links may explain the emergence of alternative semantical construals of realism, the possibility of these readings does not imply that the notion would lose its essentially metaphysical character. The fact that the debate between realists and anti-realists concerns the metaphysical status of some semantically relevant conditions does not lessen the metaphysical nature of this controversy.[27]

One hotly debated issue in contemporary literature on realism is whether the metaphysically thick notions of existence, reference or aboutness, correspondence, facts, objects, properties, and domains of quantification, in terms of which the contrast between realist and anti-realist positions has often been

[26] In chapter 2, I shall say more about the methodological considerations underlying this argumentative strategy.

[27] The same observation appears in Devitt (1991), 46. Despite their semantical characterisation, a number of influential examples of anti-realism (e.g. error theory, emotivism, expressivism, instrumentalism) can be understood as metaphysical doctrines concerning truth conditions and intended referential domains. Mackie (1977), Field (1980), Ayer (1946), Blackburn (1993), Gibbard (1990).

characterised, can be intelligibly distinguished from their metaphysically thin counterparts whose application in alternative discourses implies no metaphysical commitments on our parts concerning the relevant subject matters.[28] If the distinction is unintelligible, then the characterisations of the contrast between realism and anti-realism made in terms of these notions cannot be sufficiently illuminative either. This would be a rather sobering conclusion, since in absence of an unambiguous metaphysical vocabulary, it might become futile to argue for one or another metaphysical position.

An anti-realist about arithmetical objects, for instance, who denies the reality of the intended abstract domain of pure arithmetic, is supposed to maintain that there can be no numbers and no arithmetical facts outside the human mind that our numerals and arithmetic sentences could refer to. Most of them, however, do not want to deny that there are three prime numbers between 70 and 80. Clearly, if the term 'there are' is applied in the same sense in the metaphysical claim as in the arithmetical one, then the two commitments cannot be reconciled with each other. In other words, an anti-realist could not consistently believe in the truth of this apparently correct existential statement of pure arithmetic.

Some anti-realists have famously accepted this conclusion, and argued that our received mathematical theories, whose truth would require the existence of platonic objects and properties, or the obtaining of some platonic conditions, are actually false and the reason for which we still accept them is not because they are

[28] For a useful discussion of this problem, see Blackburn (1993) and Fine (2001). As may be apparent from my earlier claims about the concept of truth, contrary to Fine, Blackburn, Horwich and many others who observed the possibility of a "quasi-realist" or a deflated reading of the above metaphysical and semantical terms, I believe that the application of the predicate 'is true' does not allow such a construal. While there is a sense in which we can think about, believe in, refer to, and quantify over some facts, objects, properties or domains that are merely invented by human minds, the conditions whose obtaining or absence is supposed to determine the declarative applicability of our truth-apt mental and physical symbols always obtain (if they do) in the real world.

true, but because they have some other utility in our theorising about the world.[29] Most anti-realists about abstract mathematical objects and properties, however, do not accept this conclusion. They argue that their metaphysical view is fully compatible with the truth of their mathematical beliefs. They maintain that they have, in fact, two different concepts in mind that happen to be expressed equally by the term 'exists'. From the fact that the metaphysically neutral concept can be truly applied to a thinkable object it does not follow, they observe, that the metaphysically thick concept also applies to it on pain of contradiction. One may wonder, however, if the metaphysically heavy notion of existence can indeed be intelligibly stated (or denied) about particular thinkable objects and properties, and, once the answer to this question is positive, also whether the conditions under which the notion can be correctly applied to those subject matters can be informatively communicated or characterised.

These questions can be equally raised concerning all other concepts that used to play a central role in the specification of the metaphysical notion of realism. In particular, one may wonder whether the metaphysically heavy notions of aboutness/reference, correspondence, facts, objects, properties, and domains of quantification can indeed be intelligibly applied in various representational contexts, and whether the conditions governing these applications can be informatively characterised.

Since the re-emergence of metaphysical concerns in late-20th-century analytic philosophy, it had been standardly taken for granted that the proper response to the above questions is affirmative, and the alternative characterisations of realism in terms of the above notions can (more or less adequately) specify the content of the intended metaphysical positions. More recently, however, this opinion has been challenged again by

[29] Field (1980), for instance, argues that the utility of our accepted mathematical theories resides in their conservativeness (the fact that their addition to nominalistically interpretable theories does not extend the circle of nominalistically interpretable consequences of these theories), and the fact that they make the derivation of those consequences easier than it would be otherwise.

Simon Blackburn's quasi-realist considerations in the semantics of various discourses.[30]

According to Blackburn's original idea, a quasi-realist is a type of sophisticated anti-realist, who observes that her metaphysical commitments do not prevent her from an honest imitation of the linguistic practice of her realist opponent. In meta-ethics, for instance, she can honestly maintain that our ethical notions and beliefs refer to genuine ethical properties and facts, or that these beliefs are true if and only if they correspond to these facts. She can endorse this realist verdict, because the metaphysically significant terms of this formulation admit of a deflated, metaphysically non-committal reading as well, which is fully compatible with an anti-realist construal of the subject matter of ethical beliefs. If this quasi-realist observation is true, then it is hard to see how the contrast between realists and anti-realists could be properly characterised. The minimal conclusion that appears to follow from this result is that we must be quiet about these metaphysical issues, and give up the project of attributing metaphysical beliefs on the basis of others' linguistic utterances.[31] However, one may go on and argue that, since intelligible conceptual distinctions cannot fail to receive proper linguistic expressions, the ultimate moral that we should draw from the previous considerations is that the metaphysical contrast between realism and anti-realism is ultimately unintelligible.

The significance of this quietist challenge to the argumentation that I shall advance in the chapters that follow can hardly be overestimated. If the contrast between realism and anti-realism concerning a certain domain is indeed unintelligible or inexpressible, then the most important claims of this work cannot be properly understood either.

[30] Blackburn (1993).

[31] Similar conclusions have been drawn by Fine (1984), Putnam (1987) and Dworkin (1996). In his review of Blackburn's book on quasi-realism, on the other hand, Rosen (1998) argues that the quasi-realist observations need not contradict realism in the semantics of the discourses under consideration.

Accordingly, in defence of the metaphysical vocabulary that I shall rely on in the remaining chapters two claims have to be established: first, that we can distinguish the metaphysically thick notions of existence, correspondence, reference, objects, properties, and domains of quantification from their deflated counterparts; and second, that we can also communicate these distinctions, and express our metaphysical beliefs in a sufficiently transparent way. As regards the first of these points, one may observe that a proper understanding of Blackburn's argumentation presupposes the distinguishability of metaphysically thick concepts from their deflated counterparts. If we had no idea of what the realist (as opposed to a quasi-realist) intends to say by maintaining that a certain domain exists independently of our thoughts or knowledge of it, or that the true sentences of a given discourse refer or correspond to some facts in the actual world, then we could hardly understand the challenge that these formulations can be equally interpreted in an alternative, quasi-realist way, which makes them entirely acceptable to anti-realists as well.[32]

As regards the second point, at first sight, the observed anti-realist reinterpretability of the received characterisations of realism may seem to undermine the communicability of the relevant metaphysical positions. If all sentences that a realist would utter in order to characterise her own commitments could be truly endorsed (under a different interpretation) by anti-realists as well, then it would seem rather plausible to conclude that the linguistic expression of these, otherwise perhaps intelligible, conceptions can never be transparent enough. This conclusion, however, does not follow from the correctness of Blackburn's observations. The fact that there is a quasi-realist reading of the

[32] Less sophisticated evidence for our ability to distinguish between the thick and thin notions of existence, facts etc. is that we recognise the difference between the metaphysical status of Napoleon's and Little Red Riding Hood's grandmothers, even if we are ready to endorse that both of these individuals had a grandmother. Apparently, the simple fact that we truly quantify over something does not mean that we believe in its thick existence.

received realist manifestos does not mean that there can be no linguistic manifestation of the difference between a genuine and a quasi-realist construal of certain domains of facts, objects and properties, and their semantic relations to our thoughts and linguistic expressions.

One obvious way to decide whether someone is a realist or a quasi-realist concerning a certain domain is to ask her if there is a sense in which she would deny the existence of the facts, objects and properties within that domain. Supposing that the person addressed understands the difference between the metaphysically thick and the deflated thin notions of existence, facts etc., which has been observed to be a precondition of a proper understanding of Blackburn's argumentation, we may expect her to reply negatively if she is a genuine realist about the domain, and to say 'yes' if she is merely a quasi-realist about the relevant facts, objects and properties. If this is true, then we may conclude that our understanding of the quasi-realist challenge to the intelligibility and communicability of the metaphysical notions of realism and anti-realism actually guarantees not only the intelligibility, but also the communicability of these positions.

Supposing now that realism is, after all, an intelligible metaphysical doctrine, which concerns the metaphysical status of thinkable objects, properties, conditions and domains, and can be characterised by the claim that these thinkable entities exist (if they do) independently of our actual or potential thoughts or knowledge of this circumstance, I shall finally spell out a systematic ambiguity in which this metaphysical notion is often applied to alternative discourses, as in the case of mathematical or ethical realism. If the truth conditions of our claims within certain discourses can differ from what these claims purport to be about, then there will be two senses in which one can adopt realism (or anti-realism) with regard the discourses under consideration. First, one can subscribe to realism about those objects, properties and conditions that the beliefs in question purport to be about. Second, one can adopt realism about the truth conditions of these beliefs. I will call these metaphysical doctrines, in their generic forms, *realism about subject matters* and *realism about truth*

conditions, respectively, while in their specific forms in the context of particular discourses, for instance, *realism about logical or mathematical objects and properties* and *realism about logical or mathematical truths*, respectively.[33]

In the following chapters, I shall argue that in the case of our paradigm *a priori* claims about abstract domains (as well as in the case of our justifiable claims about causally inert entities in general), truth conditions cannot be identified with intended referents, and therefore a realist construal of the former entities is compatible with any view of the metaphysical status of the latter.

4. Reference

Reference is mostly explicated as a semantic relation between meaningful mental or physical symbols (e.g. concepts, thoughts, or their linguistic expressions), on the one hand, and what the users of these symbols think of or talk about while applying these representations in particular contexts, on the other. There is much disagreement concerning what can be truly said about the nature of this relation, but most people accept that, intuitively, it is an asymmetric "word-world" relation that is essentially associated with intentionality or aboutness.

Of course, the elaboration of a suitable account of the nature of this relation requires a relatively clear view about which symbols under what circumstances can be regarded as referring entities. Unfortunately, opinions about this issue are at least as divergent as those of the extension of the concept of apriority. It is commonly recognised that not all meaningful expressions in our language can refer to something. The key symbols of our non-indicative sentences (e.g. 'why', 'yes' and 'thanks') have arguably no referring power at all. Some words, however, are taken as referring expressions by virtually everyone. The least

[33] For a similar distinction in philosophy of mathematics, see Shapiro (2000), 24-33.

contested examples of the category are proper names, but most philosophers acknowledge the referring capacity of indexicals and natural kind terms as well. Definite descriptions are also amongst the popular examples, though Russellians famously oppose the classification underlying this opinion, arguing that the semantic relation of definite descriptions to the world is less direct than, and radically different from, the one obtaining between a logically proper name and its actual bearer. There is more disagreement about the referring capacity of those symbols which cannot be taken to stand for concrete particulars. For instance, many philosophers reject the view that adverbs and adjectives occurring in a predicate position can refer, for instance, to some properties or relations. Even less intuitive is to assume such a capacity in the case of some logical symbols, such as 'and' and 'nothing'.

Beyond delimiting the circle of symbols that are capable of referring to something, one may wonder also whether reference is conditional upon the actual existence of the intended entity. For instance, one may ask whether it can be reasonably supposed that empty names, names of fictive entities, expressions representing uninstantiated universals, or subject terms in true negative or false positive existentials refer even in the absence of the intended referent-candidates.

Disagreements about exactly which symbols under what circumstances can be claimed to refer to something may have two major sources: one is epistemological, the other semantical. If we apply the term 'reference' in the same sense and still disagree about its applicability in a given situation, then our disagreement must have some epistemological ground. Some of us have presumably failed to recognise the obtaining of the declarative use conditions of the symbol in the actual world. Otherwise, the relevant clash in use can be explained by invoking the semantic fact that we have actually different notions of reference in our mind. Clearly, the controversies indicated above cannot result merely from epistemic mistakes. While applying the term 'reference', philosophers often speak about different asymmetric "word-world" relations that involve intentionality or aboutness.

In one extreme, there are those who hold that reference is a substantive semantic link between a certain class of singular expressions and some concrete particulars that exist in the actual world. For a strict Russellian, for instance, the class of referring expressions includes merely the logically proper names of our language. On this construal, symbols like 'the tallest man in the world', 'is larger than' or 'Little Red Riding Hood' do not qualify as referring entities.

On the opposite side, there are those who construe reference in a very inclusive way. Frege, for instance, assigned referents to all those expressions whose meaning was supposed to contribute to the determination of the truth value of a sentence.[34] Deflationists, in turn, reject any substantive construal of this relation, and maintain that in the suggested deflated sense every expression applied in declarative contexts has a certain referent, which is specified by the pertinent instance of the general schema "a refers to a", whether or not the entity in question exists in the actual world.35 According to this construal, the fact that Little Red Riding Hood is a fictive individual who cannot occur in the actual world does not exclude that we can refer to her by applying her name in various declarative representational contexts.

In this work, I shall apply the terms 'reference' and 'referent' in a relatively inclusive sense.[36] On the construal that I shall adopt here, our ordinary acts of thinking of and speaking about some conditions, objects, properties or domains can all be called reference to these entities independently of whether or not the latter are real and obtaining or existing in the actual world. The term 'referent' will in the meantime be applied interchangeably with that of 'subject matter'. As a consequence, I shall suppose

[34] Frege (1893).

[35] Recent proponents of a deflationist construal of reference include Brandom (1994), Horwich (1998b), Horwich (2005), Field (1994) and Field (2001).

[36] In the following chapters, when I speak about reference, I shall always mean the semantic relation. In order to speak about what a referring symbol actually refers to, I shall apply the term 'referent'.

not merely that definite descriptions, adjectives, adverbs, and logical terms (or the corresponding mental symbols) have a referential link to what they purport to be about, but also that this link holds independently of whether or not the intended entities obtain or exist in the actual world. In other terms, in contrast to the case of truth, where I think our cognitive and linguistic practices undermine the viability of a deflationary construal of the subject, in the case of reference I shall subscribe to such a construal, allowing that a meaningful symbol can refer to entities that are not real or (if real) not existing in the actual world as well.[37]

In line with this understanding, I shall say that a construal of (a certain type of) truth is referential exactly when it specifies the truth conditions of the truth-apt representations under scrutiny in terms of the conditions, or the objects and properties, that these representations, or their basic constituents, purport to be about. For instance, in the case of mathematical truth, a construal is referential if and only if it implies that the truth value of our mathematical thoughts and sentences depends on whether the intended mathematical objects actually exist and possess the intended mathematical properties. It will be in this sense that I shall hold that the truth conditions, and thus the truth values, of our paradigm *a priori* claims about abstract domains (and our presumably justified claims about causally inert entities in general) cannot be suitably regarded as referential in character. In other terms, again, the point that in the following chapters I shall argue

[37] A deflationary understanding of reference does not exclude the reality and actual existence of some referents. More importantly, the adoption of this construal is fully compatible with a strictly realist explanation of the emergence of fixed referential relations between our meaningful symbols and their intended referents. The account may even presuppose our capacity to refer to actually existing entities in a narrower (Russellian or other realist) sense of the term, and require someone's acquaintance with various aspects of the actual world for "grounding" the most basic instances of substantive reference. For an early example of such a substantive theory of reference, see Devitt (1981). In chapters 5 and 7, I shall develop the outlines of such an explanation. If correct, the account will show how our mental and physical symbols can acquire fixed referential relations to fictive or actually non-existent entities as well.

for is that the truth conditions of these representations cannot be specified in terms of those objects and properties that the symbols occurring in them purport to be about.

A further important characteristic that I shall assume of reference in this work is 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. For instance, the fact that by applying the symbol 'Little Red Riding Hood' we intend to think of or speak about a fictive individual guarantees that no individual in the actual world can be the referent of this representation, or in other terms, that Little Red Riding Hood is essentially (and thus necessarily) fictive (i.e. not real) in character. By the same token, the fact that by applying a symbol in pure logic and mathematics we intend to think of or to speak about an abstract object or property guarantees that our claims or beliefs in pure logic and mathematics cannot refer to anything in the spatiotemporal world, or in other terms, that the subject matter of our representations in pure logic and mathematics is essentially, and thus necessarily, abstract (i.e. non-spatiotemporal) in character.

The intuitive ground of this assumption is relatively clear. At least it seems quite obvious that an appropriate account of, say, why the term 'tree' refers to trees rather than bushes must in some way invoke the circumstance that when the meaning of this term was once taught to us our attention was turned toward trees rather than bushes. Of course, the fact that by applying the symbols of pure logic and mathematics most people intend to refer to some abstract entities does not mean that everyone does or has to do the same. Referential intentions accompanying the application of a certain symbol may vary from person to person and from time to time. In many cases, however, these intentions are relatively well harmonised: by applying a certain symbol most people intend to refer to the same kind of entities most of the time. The greater is the uniformity among these referential intentions, the more compelling it becomes to regard the

commonly intended entities as the standard referent of the relevant symbols.

Some people, however, may still diverge from these common practices, and dispute what most others hold about the nature of some referents. In chapter 3, I shall call these dissidents "revisionists" about the relevant subject matters. In philosophy of mathematics, for instance, it is commonly believed that the subject matter of pure mathematics is abstract in character. Nevertheless, some thinkers have famously queried this consensus, and maintained that mathematics is in fact about the spatiotemporal world. According to the above terminology, they are revisionists about the nature of mathematical objects and properties.[38]

The main problem with this and most other versions of revisionism about various subject matters is that they run counter to an existing (and dominant) cognitive and linguistic practice.[39] Of course, sometimes this practice may turn out to be inconsistent. If by applying the term 'Madagascar' some people intend to speak about the island (or *this* territory over here), some others about the original bearer of the name on the African mainland (or *that* territory over there), and at the same time each of these people about the same territory as all the others, then it is clear that any consistent reconstruction of the defining characteristics of this subject must neglect at least some of the

[38] The classical proponent of this view is Mill (1843), while its most influential recent advocate Kitcher (1984). A further example of revisionism, which in the last few decades became rather influential in philosophy of mathematics, is structuralism. On a structuralist interpretation of arithmetic, for instance, the referent of our numerical term 'number one' is not a particular abstract object, but instead either any (abstract or spatiotemporal) object that occupies a certain position in a certain type of structure (viz. an ω-structure) or the position itself that the previous objects were meant to occupy.

[39] I suppose that our referential intentions accompanying the standard application of a certain symbol can always be communicated by well-chosen claims about the symbol's intended referent. The fact, for instance, that we reject the idea that numbers can occur in the spatiotemporal world is a clear manifestation of the inadequacy of a naturalistic construal of these arithmetic objects.

above intentions.[40] Unfortunately, in so far as our referential intentions are impregnated by some fallible beliefs about the intended referents, they can be never perfectly protected from such inconsistencies.[41] Nevertheless, if the beliefs in question are relatively fundamental, or the intentions under scrutiny are shared by many people, then it is highly unlikely that such inconsistencies remain concealed from most of us for such a long time.

For this reason, in this work, I shall largely neglect those proposals which attempt to save the received referentialist construal of truth in our paradigm *a priori* discourses by adopting a revisionist conception of the nature of the relevant subject matters. Thus, for instance, I shall not argue against those philosophers who deny that the subject matter of our beliefs and claims in pure logic and mathematics is abstract in character. The only versions of revisionism whose implications to our current topic I shall briefly examine are those (highly influential) structuralist construals which nevertheless maintain that the subject matter of our paradigm *a priori* discourses is abstract in character.[42]

[40] The historical example involving this term occurs in Evans (1973) as an illustration of the phenomenon of reference change. By invoking it in the current context, I intend to emphasise that the idea that referential intentions play an essential role in the determination of the nature and identity of referents is compatible with a broadly causal theory of this semantic relation. In chapters 5 and 7, I shall say more about how I think these conceptions can complement each other.

[41] Of course, not everything that we believe about a certain subject enters our referential intentions while we are thinking of this entity. One important difference between reference-fixing and other beliefs in the cognitive and linguistic practice of a given thinker is that the falsity of the former type of beliefs is not conceivable for this thinker.

[42] In chapter 4 and chapter 5, where I discuss the deflationist and anti-realist forms of referentialism about truth, I shall not argue separately against the structuralist versions of these doctrines. I believe that the arguments that I put forward in these chapters equally hold against the corresponding structuralist conceptions as well. In chapter 6, on the other hand, where I examine the viability of platonist (i.e. realist) forms of referentialism in the semantics of our paradigm *a priori* discourses, I shall be more explicit about why I think that the structuralist versions of these doctrines cannot save the adequacy of referentialism either.

Finally, it must be noted that reference in the suggested deflated sense will amount to a very fine-grained asymmetric "word-world" relation. In order to see this, let me briefly compare it with the relation that Frege denoted by this term. On the Fregean construal, two semantically complex symbols, such as the expressions 'the evening star' and 'the morning star', can actually refer to the same thing even if they are composed of semantically non-equivalent constituents. Unquestionably, what Frege conceptualised under the symbol 'reference' is a semantic relation that is exemplified in the actual world. In a coarse-grained sense of the term, the two complex symbols mentioned above are indeed about the same entity: they are both about the planet Venus.

The fact, however, that Frege's notion of reference is an adequate representation of a certain asymmetric "word-world" relation between our meaningful symbols, on the one hand, and what we intend to think of or talk about while applying these representations in particular contexts, on the other, does not mean that there are no other such relations in the actual world to be conceptualised. In the case of our previous example, for instance, there is clearly a sense in which the two expressions, which proved to be coreferential in the Fregean sense, are not exactly about the same aspects of the actual world. While the symbol 'the evening star' is semantically related to evenings, the symbol 'the morning star' to mornings on Earth. As is well known, Frege described this semantical contrast as a difference in sense. Note, however, that every particular exemplification of this fine-grained semantical contrast between two semantically complex meaningful expressions is a difference between the deflated referents of at least some components of the contrasted symbols under consideration. In other terms, what Frege attempted to explain by invoking a platonic semantic correlate of our meaningful expressions is nothing but a (deflated) referential

(i.e. asymmetric "word-world") link between our complex symbols and their respective fine-grained intended referents.[43]

Summing up, my main opponents in this work will be those philosophers who believe that an adequate theory of truth can state or imply that the truth conditions of our truth-apt mental and linguistic symbols are always to be understood in terms of those conditions or entities that these symbols refer to in the above deflated, essentially intentional and fine-grained sense of the term. Contrary to this overarching referentialist construal of truth, in the following chapters, I shall argue that in the semantics of our paradigm *a priori* discourses (and in the semantics of our presumably justified claims about causally inert entities in general) truth cannot be adequately defined as the actual existence or obtaining of the relevant fine-grained intended referents.

5. Natural vs. Abstract Domains

As mentioned before, one of the core tenets that I wish to argue for in the following chapters is that the truth conditions of our paradigm *a priori* claims (i.e. our claims about the abstract domains of pure logic and mathematics) obtain in the natural world, and therefore cannot be specified in terms of the relevant abstract intended referents. In order to ensure the proper understanding of this tenet, in this section, I wish to say a few words about the sense in which I will talk about natural and abstract domains.

Our concepts of nature have been central to philosophy from the beginnings of the discipline. With the development of modern natural sciences and the emergence of naturalism in philosophy, it has become popular to specify the content of our notion of natural domains by reference to the future ontology of

[43] In chapter 4, I shall show how this fine-grained construal enables us to neutralise the so-called "slingshot" arguments, lines of thoughts concluding that all true sentences refer to the same thing, and thus consistently maintain that we may have true representations of different aspects of the world.

our complete empirical sciences.[44] Those who believe that, in an ideal state, all scientific laws and domains will be reduced to those of fundamental physics might maintain that this ontology will be that of our complete physical theory. Others might hold that it will involve the ontologies of all complete empirical sciences. According to these construals, the atemporal domains of pure logic and mathematics, Kantian moral theory, or Fregean semantics, the private domains of a Cartesian theory of mind or phenomenology, and the fictive domains of literature must be equally regarded as non-natural.

One may raise several objections to these specifications of what being natural amounts to. For instance, one may notice that, for many people, the fact that a certain domain is mental or fictive does not imply that it cannot be natural in character. The question whether something is abstract or natural is, for these people, orthogonal to the question whether it is mental or physical, as well as to that whether it is real or fictive. In response to this challenge, the proponents of the previous construals may observe that the existence of alternative concepts of nature does not entail the inadequacy of their analyses, which are meant to be merely of one of these alternatives.

A stronger objection to this type of characterisation is to observe that our ideas of nature are meant to stand for certain domains independently of what our empirical sciences may ever tell us about the actual world. Accordingly, the question whether something is abstract or natural is meant to be independent of the ontology of our complete empirical sciences as well. If this is true, and we suppose that our referential intentions impose essential constraints upon the nature and identity of referents, then the previous characterisations of what it means to be natural cannot be adequate.[45] The moral to be drawn from this insight is

[44] Dewey (1925), Hatfield (1990).

[45] Of course, if the idea of completeness is defined by reference to everything that obtains in a natural domain, then the characterisations may become sound, though circular and thus uninformative.

that our concept of natural domains is better elucidated in non-epistemic terms.

The second way in which people often explicate their notion of natural domains, which clearly satisfies the previous requirement, is by invoking those properties that are supposed to be the essential features of natural entities. There are two characteristics that are often essentially associated with natural conditions or entities: spatiotemporality and causal efficacy. Accordingly, one may adopt two simple construals of what being natural amounts to. On the first construal, a certain condition, individual, property, or event is natural if and only if it is meant to obtain or exist at some spatiotemporal location or locations. On the second interpretation, a certain condition or entity is natural if and only if its obtaining or existence is meant to be causally significant.

The two characterisations are different, since they reduce naturality to different essential properties. Whether the resulting notions of nature differ with respect to their extensions is more controversial. Someone who maintains that causally efficient mental states, events or properties appear in time without possessing a determinate spatial location will certainly deny that the two properties delineate the same domain. Similarly, someone who believes in the existence of causally inert, but nonetheless spatiotemporally located moral or other normative characteristics will definitely reject the idea that the two properties specify the same domain.

In the following chapters, I shall apply the term 'natural' interchangeably with the term 'spatiotemporal'. In other terms, I shall suppose that the domain of natural entities is identical with the domain of conditions, individuals, properties, states and events obtaining, existing or occurring in space and time. Further, in accordance with our received cognitive and linguistic practice, I will also assume that spatiotemporality is a prerequisite for causal efficacy, so that only natural conditions and entities can enter causal relations. Finally, I shall suppose that mental states and events, together with their distinctive characteristics, occur in space as much as in time, and thus they may enter causal relations

with other conditions or entities obtaining or existing in the natural world.[46] With these assumptions in mind, in what follows, I shall argue that in so far as knowledge requires the obtaining of a suitable causal relation between the knowing mind, on the one hand, and the obtaining truth conditions of the relevant true beliefs, on the other, a proper construal of the latter conditions in the semantics of our paradigm *a priori* discourses must be naturalistic in character.

Alternative notions of abstractness have been central to philosophy since the earliest records of the discipline too. In the current literature, there are at least three notions that must be clearly distinguished from each other. The first can be contrasted with the notion of concreteness, the second with the idea of particularity, while the third with the concept of spatiotemporality.

In the first sense of the term, something is abstract if and only if it can exist merely as an aspect of a concrete entity. Autonomous existence, some philosophers say, is the privilege of concrete entities. On this construal, properties of concrete individuals, such as the colour of this rose in front of me, count as abstract entities, while concrete objects that can exist only outside space and time, such as the number one, do not.

In the second sense of the term, something is abstract if and only if it can be fully present at various spatiotemporal locations. On this understanding, universal properties, such as the property of redness in general, are abstract entities, while the particular instances of this property or the particular objects of a non-spatiotemporal domain are not.

Finally, in the third sense of the term, something is abstract if and only if its existence is atemporal (i.e. it exists, if at all,

[46] Note that this construal of what being natural amounts to is compatible with the idea that there might be causally inert entities in the natural world. Moral, aesthetic or epistemological values attributed to spatiotemporal objects are often regarded as examples of this category. Further, the construal does not presuppose the reality of natural entities either. Little Red Riding Hood, for instance, is a fictive individual, but the world in which she is supposed to live her life is nevertheless spatiotemporal, and thus, in the suggested sense, natural too.

outside the natural world). Adopting this construal, atemporal objects and properties, such as numbers, *ante rem* structures, the content of Kantian categorical imperatives, moral values, Fregean senses, propositions, inferential relations and truth values, are abstract entities, while universal properties instantiated in space and time, or the particular instances themselves, are not.

In this work, I shall focus on the semantics and epistemology of discourses about abstract domains in the third sense of the term. This is simply because I intend to provide an account of *a priori* truth and knowledge, and the sense in which the subject matter of our paradigm *a priori* beliefs is meant to be abstract is exactly this third one. In line with this understanding, the semantical tenet repeated at the beginning of this section is meant to claim that the truth conditions of our pure logical and mathematical beliefs about atemporal objects, properties and states of affairs obtain in the spatiotemporal world, and consequently cannot be constituted by these atemporal subject matters.[47]

6. Representations

Having argued against the most influential referentialist understandings of the examined paradigms of *a priori* truth, in chapter 7, I shall put forward and defend a particular representationist form of non-referentialism concerning these entities, according to which the truth value of our paradigm *a priori* beliefs is determined by the obtaining or absence of some conditions in the domain of representations that we develop in our heads in the course of our cognitive engagement with our direct natural environment. The last conceptual element that may call for some elucidation from the advanced characterisation of

[47] In chapters 6 and 7, I shall explain how the conflation of ideas about abstract entities in the second and the third senses specified here may result in serious misconceptions of the role of mathematics in the empirical sciences, and of the way we can justify our beliefs about mathematical entities.

the dialectical purpose of this work is the intended notion of representation. In this section, I shall specify what the suggested representationist construal of *a priori* truth and knowledge presupposes concerning the existence, the nature, and the semantic content of representations.

Representations are commonly understood as syntactically identifiable objects (symbols) that have semantic properties in virtue of their relations to each other and some non-representing conditions or entities in the world. They are mostly supposed to have semantic contents, referential relations, conditions of correct applicability, analytic links, and (if they are truth-apt) also propositional contents, truth conditions, truth values, and various logical relations. This much is acknowledged by virtually everyone who is engaged in talk about representations. There is more disagreement about the metaphysical nature and actual existence of the entities invoked in this characterisation.

At one extreme, there are those who deny the existence of any kind of representation. One might think, for instance, that the world does not include objects with substantive semantic relations, and future science will explain human behaviour without any reference to the manipulation of meaningful symbols. Alternatively, one may argue that the idea of representation presupposes that there are representable entities in the world whose identity conditions are independent of the classificatory work of the mind. The world, however, runs this argument, has no such structure in itself. Thus, by the use of what others take to be mental symbols, in fact, we create, rather than merely represent, the intended aspects of the world.

Most people do not query the existence of representations. Nevertheless, they often disagree about the nature of these entities. Representations are usually regarded as physical or mental objects. Some philosophers, however, maintain that there are abstract (atemporal) representations as well. Physical representations are thought to include various physical symbols, such as pictures, sounds, letters, words, sentences, and patterns of neural activities in human brains. Mental representations are taken to include mental symbols, such as (syntactically

understood) mental images, concepts, thoughts and theories, that can be entertained with different attitudes (toward their semantic content) by human minds.[48] Abstract representations have been postulated as intersubjective counterparts of subjective mental representations or (somewhat confusingly) public reifications of semantic contents, such as (Fregean) senses, concepts, propositions and theories, which may stand in further semantic relations with other entities, such as (Fregean) referents. Clearly, one's belief in the existence of one of these kinds does not imply that one is also committed to the existence of the others.

In this work, I shall suppose that there are both mental and physical representations in the spatiotemporal world.[49] In fact, I shall explain various cognitive and linguistic phenomena by assuming that we can develop within our heads more or less correct mental and physical representations of a number of different domains. This assumption is apparently incompatible with an eliminativist attitude to the standard ontology of folk psychology. According to this view, the generalisations of folk psychology are strictly speaking false, since the mental states and representations posited by this theory do not exist in the actual world. The internal determinants of human behaviour are different from these representational states, and will be revealed by our mature neuroscience. The fate of mental representations, accordingly, will be similar to that of any other entities, like the caloric fluid or the crystalline heavenly spheres, that were posited by the advocates of explanatorily inferior theoretical frameworks:

[48] According to the so-called Representational Theory of Mind, intentional mental states, such as thoughts, beliefs, desires, perceptions and imaginations, are relations of subjects to mental representations, and the intentionality of these states is best understood in terms of the semantic properties of the latter entities. In this work, I shall follow this terminology, and regard the previous kinds of mental states as various attitudes to mental representations. The only divergence that I wish to emphasise here is that by applying the term 'thought' I shall mostly refer to a truth-apt mental representation, rather than to the specific attitude of "entertaining" such an entity as the advocates of RTM tend to do.

[49] In chapter 6, I shall explain why I think that we can never reasonably believe in the existence of platonic entities, and thus in the existence of platonic representations either.

they will be eliminated from the ontology of a more adequate characterisation of the field.[50]

There are two major points that I intend to make in response to the previous observation. First, I admit that the formulations adopted in this work indicate a commitment to the existence of mental representations in the actual world. I chose these formulations, because I believe (with many others) that the considerations advanced in support of eliminativism rely on a false premise concerning the nature of folk psychological concepts, and these concepts (or some others of this kind) will be present in our daily cognitive practice independently of whether or not they are needed in an exact scientific explanation of human behaviour. The premise whose adequacy I question is that our ideas of mental entities are similar to our notion of electron or notion of caloric fluid, which represent unobservable theoretical entities and are part of our conceptual scheme only in so far as they are needed in our best scientific explanation and prediction of some phenomena. In contrast to this picture, I maintain that our notions of mental entities represent properties, states and events that are observable from a first-personal perspective, which means that our belief in the existence of these entities is based on evidence that is independent of our best scientific explanation of overt behaviour.[51]

Note, however, and this is the second major point to be made, that if mature neuroscience retained the idea of correct and incorrect neural representations, and preserved the distinction

[50] The most influential exposition of this view is Churchland (1981) and Churchland (1988/1992).

[51] In chapter 7, I shall argue that our pre-scientific concept of water has to be distinguished from its scientific counterpart that was introduced by the scientific reduction of water to the chemical compound H_2O. On the account that I shall defend there, today we have (at least) two mental representations of by and large the same stuff in the actual world: the new scientific and the old pre-scientific notions of water. If this account is correct, then our pre-scientific concept of water provides an excellent illustration (and thus amounts to a much more plausible paradigm than the eliminated concepts mentioned by Churchland) of what I think we should expect to happen with the vocabulary of folk psychology after the emergence of a mature neuroscientific explanation of overt behaviour.

between those neural mechanisms that underlie the emergence of correct representations within *a priori* discourses, on the one hand, and within empirical discourses, on the other, then an eliminativist could reconstruct the central topic of this work within her own theoretical framework as well. She could understand that my primary aim here is to provide a rough (i.e. scientifically further specifiable) account of the nature of what is currently regarded as correct *a priori* belief formation, something that she could construe within her own framework as correct *a priori* neural representation formation. Although the representations that she would quantify over would differ from those posited by a functionalist, reductive materialist, or a property dualist advocate of folk psychology as the neural correlates of representational mental states in the actual world, her construal of the *a priori* / *a posteriori* distinction could nevertheless still coincide with mine.[52] The coincidence would occur if she maintained that the essential feature that

[52] I suppose that the construal defended in this work is compatible with a functionalist, a reductive materialist and a property dualist position as well in philosophy of mind. The only assumption in it that may impose some constraint upon one's background theory of the nature of mind is the idea that mental properties, states and events occur in space and time and they can be causally related with each other as well as with some non-mental aspects of the spatiotemporal world. By adopting this assumption, one can no longer subscribe to an epiphenomenalist form of property dualism, and one runs counter to Davidson's influential anomalous monism in philosophy of mind too. As is well known, the latter position is based on the idea that mental properties, states and events are ascribed to a subject by others from a third-person perspective and with an interpretative stance towards the subject's overt behaviour, and the ascriptions in question are constrained by the rationality principles of correct interpretation, rather than by some evidence of the hidden causal determinants of the observed behaviour. Putting it otherwise, folk psychological ascriptions are meant to serve the rational, rather than the causal explanation, of overt behaviour. Clearly, this view about the content and function of mental concepts is just as antagonistic with the earlier endorsed belief in the first-person observability of mental properties, states and events as the eliminativist premise of the theoretical character of these entities. If that earlier belief is correct, then, having acquired the content of her mental concepts, each person becomes capable of recognising her own mental states without relying on observations of her own overt behaviour and the principles of rationality. Davidson (1970).

distinguishes our *a priori* neural representations from the empirical ones is that the correct declarative use conditions of the former obtain (if they do) in the domain of neural representations that we develop in our heads in the course of our cognitive interaction with our direct natural environment. In view of this possible agreement, one may conclude that (with some reformulations) the representationist construal of apriority suggested in this work can be reconciled with an eliminativist stance towards mental representations as well.

The opponent may object that eliminativism about the ontology of folk psychology involves the denial of the existence of mental representations, the primary bearers of truth and falsity, and consequently it is hard to see how an eliminativist could reasonably accept any particular account of the nature of *a priori* truth and knowledge. Note, however, that the contrast between *a priori* and *a posteriori* knowledge, justification and evidence is compatible with various conceptions of truth and the nature of primary truthbearers. A physicalistic construal of representations in human (and other) heads does not undermine the idea that the essential feature of *a priori* representations is that the correct applicability of these entities is a function of the obtaining or absence of some representational conditions in the relevant heads.

Beyond assuming the existence of mental and physical representations in the actual world, in what follows, I shall suppose also that the semantics of these representations (i.e. the semantics of thought and language) is compositional. In other terms, I shall suppose that the semantic content of complex representations is determined by that of their constituents and the way they are composed with each other. One may think that this assumption is incompatible with a connectionist conception of the computational architecture of the mind. According to this conception, mental representations are realised by patterns of activation in (natural or artificial) neural networks, and mental processes are the spreadings of such activation patterns through the networks in which they appear. Connectionism is often regarded as the major alternative of the classical view of the

subject, which took mental representations as quasi-linguistic symbolic structures consisting of semantically contentful constituents, while mental processes as rule-governed manipulations of these symbolic structures. The conflict between connectionism and the idea of a compositional semantics emerges, because the computational units in a connectionist network have no separable semantic contents, so their organisation into semantically contentful patterns cannot be compositional.[53]

As some connectionists observed, however, the compositionality of meaning does not imply anything about the semantic properties of the computational units of a representational system. What the adequacy of this principle implies is that the semantic content of a semantically complex representation realised in the system is determined by that of the semantically simple constituents of this representation. If some higher-level patterns of activation in a connectionist network can be identified with the semantically simple, while some others with the semantically complex, representational elements of a classical computational system, and the relation of the semantic contents of these elements corresponds to what is required by compositionality, then the attribution of a compositional semantics to our thought and language proves to be compatible with a connectionist view of the computational architecture of the mind as well.[54]

Speaking about semantic content, a further important tenet that I shall endorse in this work is that the semantic content of a symbol consists in the sum of those conditions whose obtaining or absence in various situative and representational contexts determine whether or not the symbol in those contexts can be

[53] Major advocates of the classical view include Turing (1950), Fodor (1975), Newell and Simon (1976) and Marr (1982), while the chief representatives of the connectionist tradition include McCulloch and Pitts (1943), Rumelhart and McClelland (1986), Smolensky (1988) and Rumelhart (1989). For an apt overview of the debates among classicists and connectionists, see MacDonald and MacDonald (1995) and Millican and Clark (1996).

[54] For a defence of this compatibility, see Smolensky (1989).

correctly applied. In what follows, I shall focus merely on that component of semantic content which is relevant to the correct *declarative* application of representations. After all, my primary aim here is to provide an adequate construal of *a priori* truth and knowledge, which phenomena emerge with the declarative application of our mental and physical symbols. With this specific interest in mind, the assumption just made boils down to the idea that the above component of a symbol's semantic content consists in the correct declarative use conditions of the symbol.[55]

In the theory of meaning, use theories are often contrasted with truth conditional accounts. Further, the contrast is often construed as a clash between anti-realist and realist conceptions of the subject. In chapter 4, I shall argue that these assumptions are false. First, the idea that semantic contents can be identified with conditions of correct use is compatible with the idea that (in the case of non-embedded truth-apt representations) they can be identified with conditions of truth. Second, the idea that semantic contents can be identified with conditions of correct use is compatible with the idea that the latter conditions have to be construed along realist lines. In this work, I shall explore both compatibilities. On the one hand, I shall stipulate that the truth conditions of a certain truth-apt representation are identical with the declarative use conditions of this symbol in "zero representational context" (i.e. in a non-embedded state). On the other hand, to account for the objectivity of truth and correct use in general, I shall also assume that the declarative use conditions of our meaningful symbols obtain (if they do) in the real world (i.e. independently of anyone's actual thought or knowledge of this particular circumstance).[56]

Another crucial assumption concerning the semantic content of representations that I shall rely on in this work is that the declarative use conditions of a mental or physical symbol are

[55] In chapters 5 and 7, I shall explain how I think our mental and physical symbols acquire their relations to their correct declarative use conditions.

[56] In chapter 4, I shall say more about the explanatory considerations underlying this realist construal of the declarative use conditions of meaningful symbols.

in some truth-apt representational contexts not referential. In other terms, in some contexts, they are not identical with those conditions that the symbol in question purports to be about. The general semantical picture underlying this assumption is the following. Each meaningful symbol may have two kinds of semantic relation to its environment. On the one hand, it may have referential links to some conditions that may or may not be real, and if real, that may or may not obtain in the actual world. On the other hand, it may have analytic links to other contentful symbols within the domain of representations that it belongs to. Both kinds of relations are essential to the constitution of the symbol's semantic content. While, in most cases, referential relations associate a symbol with its declarative use conditions in synthetic representational contexts, the analytic links between it and its symbolic environment constitute its declarative use conditions in analytic representational contexts.[57]

Now, there is a much discussed issue in philosophy of mind, which must be addressed by anyone who maintains, as I do, that (at least some part of) the semantic content of mental symbols is constituted of conditions that are supposed to obtain outside the

[57] Earlier in this chapter, I elucidated the sense in which I intend to speak about reference in this work. The concept is meant to stand for a fine-grained semantic relation obtaining between symbols and some entities or conditions that the user of these symbols intends to think of or speak about, independently of whether the latter *relata* are fictive or real, and whether they exist (obtain) or not in the actual world. The concept of analytic relations will be elucidated in chapter 7, where I shall provide an account of the emergence of these relations among some representations in our head. Here and now, it suffices to say that by an analytic link I shall understand a relation between two arbitrary symbols in space and time that ensures that in every synthetic representational context the correct declarative use conditions of one of these symbols are constitutive of the correct declarative use conditions of the other. The analytic link between our concept of bachelor and our concept of man, for instance, ensures that in each synthetic representational context the applicability of one of these concepts implies the applicability of the other. Since analytic links are supposed to obtain independently of anyone's actual thought or knowledge of this circumstance, they can constitute a factual ground for realistically interpreted analytic truths, and substantiate the traditional semantical distinction between analytic and synthetic beliefs.

subject's psychic realm. As some philosophers rightly observed, there is a manifest conflict between this externalist view of the semantic content of mental representations, on the one hand, and the relatively common conviction that psychological explanations are causal and content-based, where the causal factors invoked operate inside the subject's psychic realm, on the other.[58] The standard answer to this problem is to observe that psychological explanations can do without invoking the external components of the semantic content of mental representations, exploring merely the so-called narrow content of these symbols.[59] The assessment of this answer, of course, depends on how we construe, and therewith how we draw the line between, the external and the internal components of content. The looming problem is clearly that psychological explanations seem to invoke not merely the analytic (or inferential) relations, but also the subject matter (the least questionably external part of the semantic content) of their *explanans*.

So, the crucial question to be answered here is whether we can find a coherent characterisation of mental content, which is compatible with the following three observations: first, that most of our mental representations are about conditions that are supposed to obtain outside our psychic realms; second, that the referential aspects of our mental representations (i.e. the subject matters of our beliefs and desires) are highly significant to the psychological explanation of our behaviour; and third, that the obtaining or absence of the intended extra-psychic conditions is entirely irrelevant to the same explanations.[60]

Obviously, the three tenets cannot be reconciled unless there is a difference between the referential aspects of our mental

[58] Stich (1983) has argued that the conflict is best resolved by abandoning the idea that psychological explanations must be content based. Instead, he proposed a syntactic theory of the mind, according to which the semantic features of mental states have no role in the explanations of a scientific psychology.

[59] Classic construals of narrow content include Putnam (1975b), Fodor (1981), Fodor (1987), Fodor (1994), and Block (1986).

[60] It may be relevant to the explanation of the success or failure of our behaviour, but not to the explanation of what we actually do.

symbols invoked in psychological explanations and the extra-psychic conditions of which these symbols purport to be about. After all, the former are supposed to obtain (if they do) within, and the latter outside, our psychic realms. This invites the acknowledgement of the existence of narrow semantic contents, which include the former referential aspects of our mental symbols as well. Note that the existence of these narrow contents may provide us with explanatory resources for a proper account of what happens in reality when a certain subject alternates between two mental states, the entertaining of a thought about a certain condition and then another about another, without thereby bringing about any change in her extra-psychic environment of which these thoughts are supposed to be about. But how should we think about these narrow components of semantic contents?

The first thing that may jump into our mind is that these components can be conceptualised as ideas of those external conditions that the mental symbols in question purport to be about. The problem with this proposal is that it fails to register the difference between narrow contents and mental symbols syntactically understood. One may argue that, in contrast to their various public expressions, mental symbols have no syntactic identity conditions at all. Rather, they are all identified by reference to their narrow semantic contents, which in turn stand in a one-to-one correspondence with those (mostly external) conditions that the symbols in question purport to be about.

Some observations may support this construal of mental symbols. First, in contrast to the case of physical symbols, we never come across a mental representation without knowing something about its semantic content. Second, while a physical symbol can be associated with different semantic contents, the meaning of a mental representation seems, at least on one natural construal, to be devoid of any ambiguity.[61] Note, however, that a

[61] Of course, if mental symbols are identified by reference to the physical symbols that are used for their public expression (so that, for instance, the concept of water is supposed to be whatever happens to govern the use of our term 'water'),

certain semantic content can be associated with more than one mental symbol, as it happens in the case of definitions or the analysis of a concept in terms of some others. If this is true, then the identity conditions of mental symbols cannot be merely semantical in character.

Beyond illuminating the fruitfulness of a syntactic construal of mental representations, the last insight may help us articulate a suitable notion of narrow semantic content as well. Consider again the case when a certain subject alternates between the entertainings of two mental symbols, but now suppose that the symbols in questions have the same semantic content, like a defined concept (e.g. bachelor) and its *definiens* (e.g. unmarried man). Earlier we saw that by invoking narrow semantic contents we can explain how a subject can alternate among the entertainings of mental symbols about different external conditions without inducing any change in the intended environment. Now I wish to observe that the same explanatory perspective can illuminate not merely the difference between narrow semantic contents and the mental symbols that they are associated with, but also the way in which we can identify the former semantic entities as correlates of our mental representations syntactically understood. In the case of the current example, we can assume that whenever a subject alternates among the entertainings of various mental representations of the same external conditions, the narrow semantic content of the symbols entertained is also identical. The internal difference among the contrasted mental states in this case lies in the distinct syntactic characteristics of the chosen mental representations.

Before concluding this section, I wish to raise one more issue whose proper treatment should inform our idea of the nature of representations. In daily cognitive and linguistic practice, we usually distinguish between the possession and the actual entertainment of some representations. For instance, it is

then these contrasts between mental and physical representations will no longer obtain.

commonly held that a subject may possess a certain concept or have a certain belief (an attitude to a certain truth-apt representation) without actually entertaining the symbol in her conscious mind. If I have acquired the notion of unicorn or developed the belief that there are no unicorns in the actual world, then (in some sense) I possess this representation in my head even if I am actually not thinking of unicorns. Now, one may wonder how a certain representation could possibly qualify as mental if it is meant to be able to exist (and be possessed by a subject) even while it is not entertained in its possessor's conscious mind.

A brief reconstruction of this challenge runs as follows. First, it is assumed that concepts and thoughts are mental representations. Second, it is stipulated that mental representations are constituents of conscious mental states. Third, it is supposed that concepts and (various attitudes to) thoughts can be possessed by a subject even when they are not consciously entertained in the subject's mind. Finally, it is observed that the three premises just mentioned cannot be true at the same time, which means that at least one of them has to be eliminated from any acceptable conception of the field.

In what follows, I shall assume that in the actual world people can possess concepts and beliefs without entertaining them in their conscious mind. Whether or not a person possesses a concept or a belief with a certain semantic content will be regarded as a real fact, which obtains independently of anyone's actual thought or awareness of this circumstance.[62] The assumption will be significant, since the construal of *a priori* truth that I shall defend in this work suggests that the truth conditions of our *a priori* beliefs obtain (if they do) within the domain of the representations that we possess in our heads. Beyond this assumption, I shall leave it open whether or not a concept or a belief that is merely possessed by a subject should be classified as

[62] Since narrow contents are supposed to obtain within the subject's psychic realm, I suppose that the possession of a concept or a belief is always a possession of a semantically contentful representation.

a *mental* representation.[63] When it is actually entertained by its possessor, I shall regard it as mental.

Further, I shall also suppose that the entertainment of a concept or a belief is the successful realisation of a (natural) epistemic access to a particular representation or representational state that exists in the head of (i.e. is possessed by) the subject independently of the occurrence of this epistemic event.[64] In this regard, entertaining a mental symbol is similar to perceiving an external object or property. Both activities involve the occurrence of conscious states and presuppose the obtaining of some non-epistemic conditions in the actual world. While in the case of perception the object perceived must exist independently of the occurrence of these epistemic events, in the case of thinking some representations in the subject's head must so exist. An important semantical difference between the contrasted types, however, whose appreciation will play a crucial role in this work, is that, in contrast to the case of perception, the entities (i.e. the possessed representations) accessed by a subject in the course of her conscious thinking need not be identical with the subject matter of the resulting epistemic states. By developing an epistemic access to our concept of zebras, we mostly formulate ideas of the animals, not the accessed representation thereof.

With these observations, I conclude the elucidation of the last central concept of this work. Before proceeding to the next chapter, let me provide a brief summary of what has been done so far.

[63] In order to indicate this neutrality, I shall keep the phrase "the domain of representations in our heads" (as opposed to the phrases "the domain of representations in our brains" or "the domain of representations in our minds") to talk about the location of *a priori* truth conditions in the natural world.

[64] By realising this access, we come to know the semantic content (i.e. the correct declarative use conditions in various representational and situative contexts) of the symbols entertained. Note, however, that our knowledge of this content may be just as partial as our knowledge of other parts of the real world. Discovering facts about the correct declarative application conditions of our concepts in various representational and situative contexts may be just as difficult as discovering facts about the intended referents of these representations.

Summary

In this chapter, I attempted to elucidate the key notions of the construal of *a priori* truth and knowledge that I am to defend in this work.

As a point of departure, I put forward the three main tenets that I will argue for in the chapters to come. The first two concerned the nature of truth within our paradigm *a priori* discourses about abstract domains (i.e. pure logic and mathematics), while the third the nature *a priori* truth, knowledge, justification and evidence in general. The first tenet was that the truth conditions of our paradigm *a priori* beliefs obtain (if they do) within the domains of representations that we develop in our heads in the course of our cognitive engagement with our direct natural environment. The second ascertained that, accordingly, the truth conditions of our paradigm *a priori* beliefs cannot be adequately specified in terms of those conditions or entities that these representations purport to be about. Finally, the third laid down that by reference to the first of the previous two characteristics (viz. representationality), we can provide a useful and adequate real definition of *a priori* truth, knowledge, justification and evidence as well.

After presenting these tenets, in the six consecutive sections of this chapter, I reviewed seven crucial constituents of the given formulations, and elucidated what I shall mean by these terms in the remaining chapters of this work. In particular, I advanced the most important assumptions that I shall adopt, sometimes on explicitly stated considerations and sometimes without further reasoning, of apriority, truth, realism, reference, naturalness and abstractness, and, finally, representation.

I am aware that many of these assumptions are substantial and in no way uncontroversial. Unfortunately, a detailed discussion and defence of these premises would intolerably exceed the confines of this work. The most I could do here is to present them in a clear form. So far, what I was aiming at was not so much persuasion as the elimination of misunderstanding. If the reader of this chapter has managed to acquire a clear sense of

why she might not like the philosophical position that I wish to advocate in this work, then the efforts invested were arguably not in vain.

Having finished this preliminary study on the conceptual fronts of my enterprise, in the following chapter, I shall turn to the articulation of the major methodological assumptions that will inform my argumentation in this work.

CHAPTER 2

Methodological Preliminaries: Adequacy Conditions for a Construal of *A Priori* Truth

Introduction

In this chapter, I shall present the most important methodological assumptions that I wish to adopt and follow during my quest for a viable account of the nature of *a priori* truth in the subsequent five chapters of this work.

In section 1, I shall advance some general methodological principles that I think should guide our systematic belief formation concerning any particular segment of the world, and then indicate what I think the application of these principles amounts to in the context of our current investigation. My primary aim in this section is to clarify why I think that the best way to start an inquiry into the nature of *a priori* truth is to compile a relatively extended list of the most obvious and striking characteristics of truth in our paradigm *a priori* discourses, and then regard the potential to support a reasonable explanation of all these characteristics as a minimal condition of adequacy for an account of *a priori* truth in general.

In section 2, I shall put forward such a list of what I regard as the most important *explananda* for a proper theory of *a priori* truth. My subsequent claims about the inadequacy of standard referentialism and the minimal adequacy of a particular version of non-referentialism about truth in the semantics of our paradigm *a priori* discourses will rely on the observation that only the suggested non-referentialist conception supports a viable account of all the phenomena that I shall briefly characterise in this part.

1. General Methodological Principles

The primary purpose of a theory of a certain subject or field is to state explicitly or imply everything that can be truly said about the purported subject matters. A proper theory, however, cannot include arbitrary claims. Whatever it says about its purported subject matter must be based on some evidence. Some claims seem to be accepted on the basis of simple observations, while the confirmation of others requires serious theoretical considerations. Claims of the former type are usually called "observational statements", as they are supposed to state the obtaining of some "directly observable" conditions.[1] Those of the latter type are, in contrast, usually denoted as "theoretical statements", as they are supposed to state the obtaining of some conditions whose relation to what we can directly observe is itself dependent on theoretical assumptions. Putting it otherwise, while the obtaining or absence of the truth conditions of observational claims is supposed to be directly observable for us, that of the truth conditions of theoretical claims is not.[2] Our reasons for believing in the obtaining of the latter conditions are rather (more or less) theoretical: we believe in this obtaining, because we believe that it is (at least partly) responsible for the obtaining of those directly observable conditions whose actual obtaining we regard as a supporting evidence for these claims.

The first major claim, accordingly, that I should suggest here is that the relation of a certain theory to its evidential base is never one-directional. Rather, it is always justificatory and explanatory at the same time: while the occurrence of a certain piece of evidence confirms a certain theoretical belief, the truth

[1] Note that an observational claim, at least in the currently intended sense of the term, need not be empirical. Some of our paradigm *a priori* claims in pure logic and mathematics, for instance, are accepted on the basis of simple "intuitions" (i.e. without reliance on serious theoretical considerations), and their role in our purportedly *a priori* theory formation is similar to that of our basic perceptual claims in our theorising about the spatiotemporal world.

[2] The most influential "anti-realist" objections to this picture will be addressed and responded to in detail in chapter 5.

of this belief is supposed to contribute to the explanation of the occurrence of that piece. Note that this observation implies an important methodological principle as well: namely, a theory can be supported by reference to a certain piece (or pool) of evidence only if its truth can be supposed to contribute to an appropriate explanation of the occurrence of this evidential ground. In absence of this explanatory significance, what we can directly observe cannot provide any support to the theoretical assumption under scrutiny. The moral to be drawn from this principle is that if we want to assess the adequacy of a certain theory, then (beyond checking its consistency) what we have to examine first and foremost is whether its truth contributes to a reasonable explanation of what we can directly observe concerning its purported subject matter.

The correctness of the above methodological principle presupposes that there is an epistemic hierarchy among our beliefs, and our observational beliefs possess a more basic position within this hierarchy than our theoretical beliefs. Also, the principle seems to assume that beliefs of the former kind are based merely on direct observations, while beliefs of the latter kind are adopted on theoretical considerations, by drawing various types of inferences (e.g. inductions, deductions or abductions). Some may observe that the acceptance of this picture amounts to the embracement of a foundationalist programme in epistemology. Of course, in so far as our observational beliefs are not meant to be infallible, the picture implies no commitment to the traditional strong version of foundationalism. Nevertheless, assuming the existence of epistemically basic, non-inferentially confirmable beliefs is a defining characteristics of foundationalism. So, the suggested conception seems to entail at least a weak, fallibilist, version of foundationalism.

In fact, the methodological framework that I shall adopt does not presuppose the sharp separability of observational and theoretical beliefs or, in other terms, the existence of strictly observational (i.e. non-inferentially confirmable) beliefs. What it does assume is that there is an epistemic hierarchy among our

beliefs, and that in most cases it is more rational for us to assess the adequacy of our less basic beliefs in view of the more basic ones than the other way around. Clearly, this assumption is compatible with an anti-foundationalist perspective in epistemology as well. In particular, it is compatible with the anti-foundationalist claim that there are no beliefs whose justification can be regarded as purely observational. The fact that some beliefs are more fundamental than others does not mean that we cannot occasionally have reason to revise them in view of what follows from, or can be explained by means of, the others.[3]

A further methodological point to be noted is that we can never actually develop a complete picture of the evidential ground of a serious theory (of some segment) of the world. No matter how many observable characteristics we specify, whose collective explanation constitutes an adequacy condition for a theory of a certain subject, we will never be in possession of all evidence that is significant to the assessment of the theory under consideration. Accordingly, we must recognise that the appraisal of a theory which can contribute to a reasonable explanation of all observed characteristics of its subject matter can never be higher than "minimally adequate".

On the other hand, the more inclusive our evidential ground for assessing a certain theory is, the more likely it becomes that the evaluation based on this ground is correct. Accordingly, if we want to enhance the conclusiveness of our argumentation for or against a theory of a certain subject, then the best we can do is to collect as many crucial *explananda* for the theory in question as we can.

In this work, I shall follow the previous methodological principles while developing an account of the nature of *a priori*

[3] One may wonder whether this hierarchic conception of rational belief formation can play a regulative role in our cognitive practice, and if it can, then how. The idea that in a naturalistic methodological framework there is no room for the establishment of substantive and informative epistemic norms has been recently defended by Knowles (2003). My brief reaction to the sceptical arguments advanced in that book can be found in Novák (2006).

truth. First of all, I shall suppose that a certain conception of *a priori* truth cannot be adequate unless it supports a reasonable explanation of everything that we can safely establish by observation of this particular subject matter. The observations in question may be fallible, and in principle even revisable on theoretical considerations, but practically they will be very likely true, and thus naturally contribute to the suggested evidential base.

As I already mentioned in chapter 1, in order to ensure the greatest initial agreement among readers with slightly divergent notions of apriority, I shall first concentrate on the observable characteristics of truth as it is attributed to our *paradigm a priori* representations, most importantly to our logical and mathematical claims or beliefs about abstract states of affairs. I shall take it that supporting a reasonable explanation of the occurrence of all these characteristics is a minimal condition of adequacy for not merely a theory of the examined paradigms, but also for a construal of *a priori* truth in general. Once we have developed an understanding that satisfies this minimal adequacy condition, we can consider whether it supports an explanation of those specific features as well that we can observe about the less canonical instances of *a priori* truth. If this extension of our evidential base does not necessitate any change in the candidate account, then the contested instances can be simply classified as *a priori* too. If the new *explananda* undermine the adequacy of our conception, then we can either disqualify the instances or revise the account.

My list of the most important *explananda* for a theory of *a priori* truth will be relatively inclusive. On the one hand, I believe it will include sufficient material for me to show the inadequacy of what I shall call the standard referentialist conception of *a priori* truth. On the other hand, it is meant to contain all those characteristics as well, whose occurrence may prompt philosophers to query the adequacy of the alternative naturalistic construal of this subject that I shall advocate below. In accordance with the last methodological consideration mentioned above, I shall suppose that by imposing a relatively demanding constraint upon the envisaged construal of *a priori* truth, we can

substantially enhance the conclusiveness of our argumentation in the remaining chapters of this work.[4]

Having argued that the best way to start an inquiry into the nature of *a priori* truth is to compile a relatively extended list of the most obvious and striking properties of truth within our paradigm *a priori* discourses, and then assess the alternative accounts of the subject on the basis of their explanatory potential with respect to this evidential ground, in the following section, I shall specify the envisaged *explananda* (i.e. collect those observable phenomena, whose joint explanation I shall regard as a minimal condition of adequacy for a construal of *a priori* truth).

2. Major Explananda *for a Construal of* A Priori *Truth*

A proper account of *a priori* truth must explain, either in itself or as part of a larger theory, two kinds of characteristics of its intended subject matter: those which are possessed by truth within any discourse whatsoever, and those which are specific features of the *a priori* instances under scrutiny.[5] In the first part of this section, I shall briefly review the most important elements of the former class.

1. Fit with a General Construal of Truth

The first thing that a proper account of any kind of truth must clarify is what makes the particular sort of instances under

[4] Note, however, that my list will not include the (broadly Tarskian) observation that, for any sentence (or proposition) p, p is true if and only if *p* (where p is the metalinguistic name of p). I take it that this observation of the correct use of our notion of truth (or our predicate 'is true') in the above type of representational context is also obviously true, and thus also to be accounted for by an adequate construal of (*a priori*) truth. The reason for which I shall neglect this *explanandum* here is that its existence seems to provide no explanatory challenge to any account that I intend to discuss in the rest of this work.

[5] Many of these *explananda* are brought into relief in the context of philosophy of mathematics by Shapiro (2000).

scrutiny instances of truth in general. In other terms, an account of any kind of truth must fit with a general construal of truth.[6] According to the received referentialist construals, truth characterises its bearers in virtue of the obtaining of those conditions that the bearers in question are supposed to be about. In view of this background theory, an account of the specified paradigms of *a priori* truth can meet the current adequacy condition if it implies that the instances under study characterise their bearers in virtue of the obtaining of the relevant referential conditions. If the account does not imply this conclusion, then the only way to establish the required harmony is to modify the above referentialist conception of truth. In this work, I shall argue for the appropriateness of the latter strategy, and instead of adopting a referentialist construal of truth in the semantics of pure logic and mathematics, I will rather offer an alternative understanding of the nature of truth in general.

2. Objectivity of Truth

The second feature to be explained by an adequate construal of virtually any kind of truth is the objectivity of the truth value of the relevant truth-apt representation. The epistemically basic observation in this case is that the paradigms of *a priori* truth characterise their bearers independently of what anyone actually thinks about this issue.[7] In other words, no actual epistemic states

[6] Benacerraf (1973), 666, Shapiro (2000), 31.

[7] The only case in which truth may appear as an epistemic property is when it characterises a representation of a reflective epistemic state, such as the claim *that I am aware of my current state of mind*. The idea that truth is objective is often understood as an essential tenet of realism about truth. According to Shapiro, for instance, realism about truth (in his terminology: realism in truth-value) is the doctrine "that mathematical statements have objective truth-values, independent of the minds, languages, conventions, and so on of mathematicians" and "[p]art of what it is for mathematical statements to be objective is the possibility that the truth of some sentences is beyond the abilities of humans to know this truth". Shapiro (2000), 29-30. Accordingly, for him, realism about truth is *defined* in terms of objectivity, while objectivity in terms of verification transcendence (in Dummett's sense of the term). This is why he does not take objectivity as an

concerning the truth value of an *a priori* belief are constitutive of the latter's possession of that value. Putting it briefly, it seems that in pure logic and mathematics (as in most other truth-apt discourses) no epistemic agent is *conceptually* prevented from committing an epistemic mistake. The most natural way of explaining this characteristic is to maintain that the facts determining the truth value of these representations obtain in the real world (i.e. independently of our actual beliefs about these circumstances). The main reason for which I shall suggest, in chapters 4 and 5, the rejection of any deflationist or anti-realist form of mainstream referentialism about truth is that these accounts cannot explain the above characteristic of this semantic property. In chapter 7, on the other hand, I shall argue that the naturalistic account advocated in this work of the nature of *a priori* truth provides sufficient explanatory resources to satisfy this second adequacy condition as well.

3. Emergence of Semantic Content

The third general *explanandum* that I wish to invoke is the emergence and existence of those semantic relations that turn particular facts into the obtaining intended referents and/or obtaining truth conditions of particular beliefs. Without these semantic relations, our thoughts and sentences could not be about and/or made true by the obtaining of anything in the world. One received way of accounting for this *explanandum* is to

undisputed characteristic of mathematical truth. After all, some philosophers believe that there are no unprovable mathematical truths. In contrast, my notion of objectivity does not appear in the *definiens* of realism about truth, and neither of these notions is defined in terms of verification transcendence. Further, I suppose that objectivity (on the adopted construal) is a virtually undisputed characteristic of mathematical (and most other) truths and falsities, whose presence has to be observed and explained by any account of these truths. An anti-realist or a deflationist may query the realist construal of the truth conditions of our truth-apt representations, but it is very unlikely that any of these opponents would want to hold that the truth value of those representations is determined by someone's actual judgement of this value.

invoke the idea of acquaintance. In particular, one may argue that the semantic relations between representations and their intended referents or truth conditions are established by the mind while it is acquainted with those *relata* (once they obtain) in the course of its cognitive interaction with its environment. Of course, this account can be accepted only if we can legitimately assume the existence of this epistemic contact between the above *relata*. In chapter 6, I shall argue that a platonist construal of the truth conditions of our purely logical and mathematical beliefs undermines all explanatorily useful concepts of such an epistemic contact. On the other hand, in chapter 5 as well as in chapter 7, I shall also argue that acquaintance is merely the most basic, but not the only way in which our mind can establish its semantic relations to various aspects of the world. In fact, I believe that, in possession of their alternative concept-forming resources, human minds can develop representations whose truth conditions obtain in a platonic realm. Accordingly, the main reason for which I shall suggest, in chapter 6, the rejection of all platonist forms of referentialism about truth in the semantics of our paradigm *a priori* discourses is *not* that these accounts cannot explain the emergence of semantic relations between platonic facts and human minds, but instead that they cannot explain how we could know of the obtaining or absence of the relevant platonic conditions. In chapter 7, I shall devote a separate section to explaining how our paradigm *a priori* beliefs acquire their non-referentialist (i.e. non-platonic) truth conditions, largely because a proper understanding of this phenomenon provides us with key ideas for explaining some other crucial characteristics of these *a priori* truths as well.

4. Knowledge / Reliability of Evidence

In many discourses, we can reasonably suppose that we have knowledge of the target domain. A piece of propositional knowledge is always knowledge of the obtaining or absence of some truth conditions, or, in other words, the truth or falsity of some representations. Due to this conceptual link between

knowledge and truth, a suitable account of any knowable kind of truth must support a proper explanation of this potential knowledge as well.[8] The received view today is that an acceptable explanation of knowledge must include an intelligible account of why our actual evidence for a piece of knowledge is a reliable indicator of the obtaining of the relevant truth conditions, and the account in question can hardly be given without assuming a suitable link, a one-way influence or identity, between obtaining truth conditions and observed epistemic grounds. If this view is correct, then an understanding of a certain kind of truth can meet the current adequacy condition only if it construes its subject matter either in an anti-realist way, or in terms of the obtaining of some causally efficient conditions. An alternative approach to this *explanandum* is to adopt a non-causal epistemology, which is compatible with an anti-naturalist construal of truth conditions as well. While in the semantics of our broadly physicalistic discourses realist forms of the standard referentialist construal of truth are compatible with a broadly causal explanation of knowledge, in the semantics of our paradigm *a priori* discourses they are apparently not. As I mentioned above, the main reason for which I shall suggest, in chapter 6, the rejection of all realist forms of referentialism (i.e. platonism) about truth in the semantics of our paradigm *a priori* discourses is that these accounts, in my view, cannot explain how the obtaining of the alleged platonic truth conditions could in any way influence, and thus reliably inform, the development of our paradigm *a priori* beliefs. In chapter 7, on the other hand, I shall argue that the non-referentialist account advocated in this work can clearly meet this explanatory requirement as well.

[8] Together with the first *explanandum*, this requirement is at the heart of Benacerraf's famous dilemma about mathematical truth. Benacerraf (1973), 667. The case will be presented in more detail in chapter 3. Shapiro invokes this *explanandum* together with apriority as part of a complex phenomenon: the possibility of *a priori* knowledge of mathematical domains. Shapiro (2000), 22-23.

5. Intersubjectivity of Semantic Content

The fifth adequacy condition for a construal of any kind of truth is to support an account of the intersubjectivity of the semantic content of our mental and physical representations. The basic observation here is that different people can entertain thoughts with the same semantic content. One traditional account of this *explanandum* is the Fregean idea that the shared aspects of semantic contents are platonic entities (i.e. elements of an epistemically accessible intersubjective abstract realm).[9] Of course, the acceptability of this account presupposes that we can develop an epistemic access to platonic entities. In chapter 6, I shall argue that we have no reason to suppose that the latter condition obtains. Therefore, in line with what I said in the last section of chapter 1, in this work, I shall reject this Fregean view of semantic contents. Instead, I shall suppose that semantic contents are constituted of two conceptually separable semantic correlates of our mental and physical symbols: the first is constituted by the declarative use conditions, while the second by the (fine-grained) intended referents of these representations. If the declarative use conditions of a certain symbol can be specified in terms of its intended referents, then the two correlates actually coincide. In chapter 6, I shall concede that the standard referentialist and realist (i.e. platonist) construal of these correlates in the semantics of our paradigm *a priori* discourses provides a relatively simple account of the current *explanandum* as well. Nonetheless, in chapter 7, I shall show that the alternative non-referentialist construal of the paradigms of *a priori* truth that is advocated in this work can meet this adequacy condition (i.e. support a proper explanation of the intersubjectivity of the relevant semantic contents) as well.

[9] Frege's use of the term 'content' is not uniform. In his early writings, he takes content to be either the referent or the sense of a certain symbol. Later he abandons the notion, and clarifies his terminology by the consistent application of his contrast between sense/thought and reference/truth value. Cf. Frege (1879) and Frege (1892).

6. Observable Convergence of Beliefs

The next feature that a proper account of any kind of truth must, within a larger theory, explain is the observable measure of convergence among the relevant kind of beliefs. In the case of logic and mathematics this convergence is strong, while in the case of other discourses, such as ethics or aesthetics, it seems considerably weaker. For those who maintain a cognitivist construal of these discourses, both sorts of phenomenon call for an explanation. Since convergence and divergence are supposed to characterise beliefs with shared semantic contents, any account of the current *explanandum* must be brought into harmony with the earlier explanation of how we can entertain thoughts with shared semantic contents. Further, since the most natural way of accounting for the (more or less strong) convergence of semantically equivalent beliefs is by reference to the (more or less strong) reliability of those epistemic capacities that underlie the formation of these beliefs, a proper account of (the observable measure of) convergence within our paradigm *a priori* discourses must fit with the former explanation of knowledge acquisition about these domains as well. If semantical platonists had an acceptable account of our alleged epistemic access to intersubjective platonic facts, their construal of the paradigms of *a priori* truth would easily satisfy the current adequacy condition too. The inadequacy of platonist theories of knowledge, which I shall argue for in chapter 6, however, undermines the corresponding platonist accounts for convergence as well. In chapter 7, therefore, I shall provide an alternative account of this *explanandum*, and show how a non-referentialist, representationist construal of the paradigms of *a priori* truth can fit into that account, and thus satisfy the current adequacy condition as well.

7. Infinity of Semantically Non-Equivalent Truth-Apt Representations

The last general feature of truth that I shall include in the current list of major *explananda* for a suitable construal of *priori* truth is the infinity of semantically non-equivalent truth-apt

representations. In our paradigm *a priori* discourses, as well as in our empirical discourses about the natural world, there seems to be no upper limit for the formation of semantically distinct true (or false) beliefs. Accordingly, a proper account of *a priori* truth must ensure that these semantic properties can, in principle, characterise infinitely many different beliefs. In chapter 6, I shall argue that the standard platonist construal of the examined paradigms of *a priori* truth can meet this explanatory requirement by endorsing a simple referentialist construal of the semantic content of our paradigm *a priori* claims and observing that the intended referential domain of these truth-apt representations is infinite in character. In chapter 7, on the other hand, I shall show that the infinity of semantically distinct *a priori* truths does not require the infinity of the domain of obtaining *a priori* truth conditions, and that a non-referentialist, representationist construal of these conditions supports a suitable account of the current *explanandum* despite the presumable finiteness of the crucial representational domain in our heads.

Beyond these general features of truth, whose proper explanation is an adequacy condition for a construal of virtually any species of this semantic property, some further characteristics pertain specifically to *a priori* truths, or at least the paradigm instances of them, and thus constitute an *explanandum* only for an account of this particular type of truth. In the remaining part of this section, I shall briefly review the most important examples of these specific *explananda* as well.

8. Apriority of Evidence

A defining characteristic of any *a priori* knowable truth is that we can establish its presence or absence without reliance on sensory experience.[10] Our claims about abstract states of affairs are the

[10] As it has been emphasised in the first section of chapter 1, the *a priori* knowability of certain truths and falsities does not imply the *a priori* acquirability

primary examples of truthbearers with *a priori* knowable truth values. So, the first specific adequacy condition for an account of *a priori* truth is that it must support a proper explanation of how the distribution of this semantic property can be discovered without reliance on sensory experience. The requirement can be understood as a specification of the former condition of explaining knowledge acquisition. Accordingly, a suitable account of the apriority of evidence in pure logic and mathematics can be taken as a specification of a suitable account of the reliability of that evidence. An account, for instance, which stipulates the existence of a specific *a priori* source of evidence can meet this adequacy condition only if it supports a suitable understanding of why the deliverances of this source are reliable indicators of the relevant *a priori* truth conditions. This is why I believe that the arguments to be advanced in chapter 6 against semantical platonist accounts of knowledge acquisition can undermine the standard platonist conceptions of apriority as well. On the other hand, the *explanandum* imposes a substantial constraint upon the suggested naturalistic construal of *a priori* truth too. In particular, as we shall see in chapter 7, it will provide the main motivation for adopting a representationist construal of *a priori* truth (i.e. an account according to which the truth conditions of our *a priori* claims in general obtain in the domain of representations in human heads).

9. Necessity of (the Paradigms of) A Priori *Truth*

The second specific feature of the paradigms of *a priori* truth that a proper construal of them must, within a larger theory, explain is that the relation of this semantic value (or its opposite) to its bearers is necessary in character.[11] In the standard referentialist framework, this requirement is usually taken to be met by stipulating that the truth conditions of paradigm *a priori* beliefs

of the conceptual elements of the relevant truthbearers. The *explanandum* currently at issue is merely the former phenomenon. Shapiro (2000), 22-23.

[11] Shapiro (2000), 21-22.

obtain (or not) in an abstract realm, whose facts are all necessary in character. A representationist construal of *a priori* truth can, of course, hardly rely on such stipulations, since the natural facts that it takes to be the factual ground of *a priori* truths are supposed to exist contingently in the actual world. In chapter 7, I shall show that the construal can nevertheless meet this adequacy condition as well, since the modal contrast between the paradigms of *a priori* truth, on the one hand, and the paradigms of empirical truth, on the other, can be explained without reliance on the distinction between contingently and necessarily obtaining states of affairs.

10. Applicability of A Priori *Knowledge in the Empirical Sciences*

The third often cited specific feature of the paradigms of *a priori* truth to be explained by a proper construal of the subject is the applicability of our knowledge of these paradigms in the development of our empirical theories of the natural world.[12] How is it that our knowledge of abstract objects and properties, and thus of the distribution of the corresponding paradigm instances of *a priori* truth, can help us discover the truth value of claims about the spatiotemporal world? In a referentialist semantical framework, one may try to explain this phenomenon by supposing that the abstract facts that determine the truth value of our logical and mathematical claims stand in a suitable metaphysical relation with the spatiotemporal facts that determine the truth value of our empirical claims about the natural world. This explanation, however, is rather contestable in the light of the alleged metaphysical gap between the relevant abstract and natural realms. In fact, I believe that beyond the problem of explaining how we could gain knowledge of platonic objects and properties, this *explanandum* provides the most serious challenge to a realist (i.e. platonist) version of referentialism concerning the paradigms of *a priori* truth. The two explanatory

[12] Shapiro (2000), 23, 33-39. Steiner (1998) provides a more sophisticated characterisation of this *explanandum*.

puzzles are clearly independent of each other. Even if there were an acceptable platonist theory of *a priori* knowledge acquisition, in absence of a suitable characterisation of the metaphysical link between the intended abstract and natural domains (in a referentialist semantical framework), the applicability of logical and mathematical knowledge in the empirical sciences would still appear as a miraculous coincidence. Although my arguments against a platonist construal of *a priori* truth will not draw on the former observation, nevertheless I suppose that a full case against that doctrine could also include the objection that a platonist construal of *a priori* truth cannot account for the applicability of logical and mathematical knowledge in the empirical sciences either. In chapter 7, on the other hand, I shall show that a naturalistic explanation of the emergence of semantic content within our paradigm *a priori* discourses provides us with a suitable account of this *explanandum* as well.

11. Abstractness and Infinity of Intended Domains

The last specific feature of the paradigms of *a priori* truth that I wish to include to the current list of major *explananda* is that the bearers of these paradigm instances are about an abstract and often infinite domain.[13] One may wonder how the instances under scrutiny actually relate to this specific sort of subject matter, and what makes it the case that they can characterise bearers with such subject matters. In fact, this *explanandum* can be understood as the specification of the third presented above. It imposes specific constraints upon a proper account of the emergence of semantic relations between representations and aspects of the world. It requires that the account in question explain the emergence of semantic relations between our paradigm *a priori* beliefs and their abstract and often infinite subject matter. As mentioned before, the received view today is that truth in general is to be understood along referentialist lines.

[13] Shapiro (2000), 28-29.

If this referentialist construal is correct, then a proper account of the emergence of referential relations between our paradigm *a priori* representations and their abstract and infinite subject matters will *eo ipso* explain how these representations acquire their truth conditions as well. In chapters 4, 5 and 6, however, I shall argue that the paradigms of *a priori* truth do not admit of a referentialist construal. Accordingly, an adequate construal of these truths must be compatible not merely with an account of the emergence of semantic relations between the relevant beliefs and their truth conditions, but also with an account of the emergence of determinate referential relations between them and their intended abstract and infinite subject matter. In chapter 7, I shall show that the naturalistic construal of the paradigms of *a priori* truth advocated in this work satisfies this adequacy condition as well, since it supports a suitable explanation of how we acquire the capacity of thinking truly about abstract and infinite domains.

In the following chapters, I shall regard the eleven characteristics specified above as the primary *explananda* for a proper construal of *a priori* truth. As I emphasised in section 1, the list is far from complete. It merely includes those features whose explanation usually occurs as a striking *desideratum* in the literature on the relevant *a priori* truths. There are definitely a number of other observable facts, whose proper explanation could be taken as a further condition of adequacy for an account of the current subject.[14] Nevertheless, I believe that the constraints imposed by the above explanatory requirements will be sufficient for us to realise that a viable construal of what we take to be the paradigms of *a priori* truth must be non-referentialist and representationist in character.

[14] One may observe, for instance, that a proper account of truth in the semantics of our paradigm *a priori* discourses has to support an explanation of the epistemological significance of proofs in pure mathematics, and the successful applicability of thought experiments in the development of our scientific theories of the spatiotemporal world as well.

Summary

In this chapter, I presented the most important methodological assumptions that I wish to adopt in this work.

In section 1, I advanced those general methodological principles that are supposed to inform my argumentation in the following chapters. First, I argued that the relation of a certain theory to its evidential base is justificatory and explanatory at the same time, and therefore a theory can be supported by reference to a certain piece (or pool) of evidence only if its truth can be supposed to contribute to an appropriate explanation of the occurrence of this evidential ground. Second, I highlighted that my argumentative strategy will not presuppose the sharp separability of observational and theoretical beliefs, but will assume that there is an epistemic hierarchy among our beliefs, so in most cases we can assess the adequacy of our less basic beliefs by considering whether they are in harmony with the more basic ones. Third, I conceded that our actual evidential ground for a serious theory of (some aspect of) the world can never be complete, and therefore the most that we can claim about a theory that can contribute to a reasonable explanation of all striking characteristics of its subject matter is that it is minimally adequate. Finally, I observed that the more inclusive our evidential ground at the assessment of a certain theory is, the more likely it becomes that the assessment relying on this ground is correct, so that if we want to enhance the conclusiveness of our argumentation for or against some theories of a certain subject, then the best we can do is to collect as many crucial *explananda* for these theories as we can. Applying these general principles to our current subject, I concluded that the best way to start an inquiry into the nature of *a priori* truth is to compile a relatively extended list of the most obvious and striking characteristics of truth in our paradigm *a priori* discourses, and then regard the potential to support a reasonable explanation of all these characteristics as a minimal condition of adequacy for a construal of *a priori* truth in general.

In section 2, I provided a brief characterisation of the envisaged major *explananda* for a theory of *a priori* truth. In view of this evaluative ground, in the remaining part of this work I shall argue for two major claims: first, that the standard referentialist construals of *a priori* truth are equally inadequate, since neither of them can support an appropriate account of all characteristics specified above; and second, that a non-referentialist, representationist construal of the examined paradigms of our subject can satisfy all advanced explanatory requirements, and thus amounts to a minimally adequate account of *a priori* truth.

CHAPTER 3

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

[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.

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

[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 semantical 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 semantical 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 semantical 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 semantical assumption, also explicitly touched upon by Benacerraf, is that the subject matters of mathematical

[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.

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*

[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

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

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

[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

[31] Ramsey (1927), Strawson (1950), Quine (1970), Grover, Camp and Belnap (1975), and Horwich (1998b).
[32] Blackburn (1993).

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.

[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 semantical 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

[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.

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

[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 semantical 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 semantical 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

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

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 semantical 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 semantical responses that they can give are that our epistemologically problematic beliefs about causally inert domains

[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

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

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.

CHAPTER 4

Referentialist Responses to Benacerraf's Dilemma I: Deflationist Construals of Truth

Introduction

Deflationist theories of truth can be characterised by their allegiance to the claim that truth is not a substantive property whose nature and relation to our epistemic capacities could be specified by a proper theory of the subject. In this, deflationism can be contrasted with all realist and anti-realist theories of truth. Alternative versions of deflationism endorse various further claims about the cognitive function of our concept of truth.[1] Most deflationists today believe that the concept is merely a logical device, which does not stand for a property whose nature could be further specified. When we apply the predicate 'is true' to certain thoughts or their linguistic expressions, we do not attribute a real property to these representations, but instead we assert something that could as well have been asserted by endorsing the relevant thoughts or sentences.

By adopting a deflationist theory of truth, an advocate of standard referentialism can provide a *prima facie* acceptable response to Benacerraf's epistemological challenge in the

---

[1] There are several classifications of deflationist accounts in the literature. The main versions that are usually distinguished include Ramsey's Redundancy Theory (Ramsey (1927)), Tarski's Semantic Theory (Tarski (1944)), Strawson's Performative Theory (Strawson (1950)), Quine's Disquotationist Theory (Quine (1970)), Grover's, Camp's and Belnap's Prosentential Theory (Grover, Camp and Belnap (1975)), and Horwich's Minimalist Theory (Horwich (1998b), Horwich (2005)). Further proponents of deflationism include Ayer (1935), Wittgenstein (1953), Leeds (1978), Fine (1984), Soames (1984), Soames (1997), Field (1986), Field (1994), Williams (1986), Loar (1987), Brandom (1988), Brandom (1994).

semantics of our paradigm *a priori* discourses. As it has been shown in chapter 3, in its most generic form, the epistemological argument questions 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. By abandoning the realist tenet, a deflationist can subscribe to the standard referentialist construal of truth without falling prey to Benacerraf's argument. Clearly, if truth is not a substantive property, then referentialism about truth does not imply the (thick) existence of the subject matter of true beliefs either. In a deflationist semantical framework, referentialism can be maintained independently of what can be truly said in metaphysics and epistemology. In the semantics of our paradigm *a priori* discourses, this means that our beliefs can be regarded as true or false in the standard referentialist sense independently of what can be said about the metaphysical status of the relevant abstract subject matters, and the ways in which we acquire knowledge or develop reliable beliefs about them. A deflationist may admit that she has no suitable explanation of how we acquire knowledge of causally inert domains. Nevertheless, in view of the alleged autonomy of semantics from metaphysics and epistemology, this failure is not supposed to interfere with her belief that our notion of truth is best specified in the standard referentialist way.

In this chapter, I shall examine the viability of this deflationist version of standard referentialism, and argue that the indicated advantages of the theory are cancelled out by the fact that a deflationist can provide no suitable explanation of the objectivity of truth. This is because in absence of a real distinction between the situation in which truth does characterise a particular representation and that other in which it does not there seems to be no way to understand how the obtaining of either of these options could be conceptually independent of what anyone ever believes about this issue. In other terms, a deflationist conception of truth does not tell us why none of us is ever conceptually prevented from committing an epistemic mistake concerning the applicability of representations. A realist

construal of truth provides a natural explanation of this phenomenon: the conceptual possibility of these epistemic mistakes is a consequence of the reality of truth. If so, then by applying the predicate 'is true' we attribute a real property to our representations. Moreover, this property seems to have an underlying relational nature as well, which can be characterised as a sort of correspondence. According to this construal, truth is the property of having suitable semantic relations to some conditions that obtain in the actual world, or having a semantic content that corresponds to some fact in the actual world.

Since the arguments that I shall advance attack a general feature of deflationism, there is no need to discuss the various formulations of this programme. Instead, in section 1, I shall focus on the best elaborated deflationist conception in the current literature, Paul Horwich's minimalist account of the concept of truth. After a brief presentation of his position, I shall argue that although many of his core insights about our concept of truth seems to be well-considered and adequate, nevertheless his negative tenet that by applying this concept we do not attribute a substantive property whose nature and relation to our epistemic capacities can be specified by a proper theory of the subject is clearly incompatible with one of the key distinctions of any plausible theory of meaning and truth (including his own "*use-regularity* conception of meaning"). The distinction in question is that between the actual and the correct use, or between our belief in the correctness and the objective correctness of the use, of various sorts of representations. The incompatibility obtains, because the distinction presupposes what Horwich's minimalism, together with other forms of deflationism, denies: namely, that truth is a real property, which characterises our truth-apt representations independently of whether anyone ever knows of the obtaining of these particular states of affairs.

Having developed my argument against the deflationist conception of truth, in section 2, I shall defend the received realist (i.e. correspondence) theory of this semantic property against the so-called "slingshot argument", which is supposed to

demonstrate that all true beliefs refer to the same thing, and therefore cannot be made true by the obtaining of their own *separate* truth conditions. Although the argument's conclusion is compatible with a realist construal of truth, and also with the identification of truth with the property of having suitable semantic relations to some (namely: all) conditions that obtain in the actual world, nevertheless it undermines the received realist assumption that our true beliefs may refer to, and be made true by, different facts in the world. After presenting the argument in its simplest and most transparent form, I shall show that two of its background assumptions rely on a highly coarse-grained (Fregean) notion of reference, which is different from the one that we use when specifying the semantic content of our representations along the standard referentialist lines. If we replace the former collapse-generating concept with the latter fine-grained notion in these assumptions, then the argument breaks down, and no longer challenges the adequacy of the received realist (i.e. correspondence) theory of truth.

1. Minimalism and the Explanation of the Objectivity of Truth

Paul Horwich's minimal theory of truth is the most fully elaborated and widely discussed form of deflationism today. In the second edition of his book on the concept of truth, he introduces his account as a reaction to two misconceptions entrenched in present-day philosophy: "first, that truth *has* some hidden structure awaiting our discovery; and secondly, that hinging on this discovery is our ability to explain central philosophical principles […], and thereby to solve a host of problems in logic, semantics, and epistemology".[2] The principles whose explanation seems to call, according to many philosophers, for a substantive notion of truth include, for instance, that truth

[2] Horwich (1998b), 2.

is the aim of science, that true beliefs tend to facilitate the achievement of practical goals, that truth is preserved in valid reasoning, or that to understand a sentence is to know which circumstances would make it true. According to Horwich, the presented misconceptions are rooted in our belief that the predicate 'is true', like other familiar predicates, such as 'is magnetic' or 'is diabetic', also designates a certain complex feature of the world whose underlying structure could be revealed by a succinct philosophical or scientific analysis.

In contrast to this belief, Horwich maintains that "the truth predicate exists solely for the sake of a certain logical need".[3] Occasionally, we may want to endorse propositions that we cannot directly identify by using their name, either because we are ignorant of what those propositions are, or because they are so many that we cannot grasp them all in our mind. In such cases, we can express our attitude by applying the truth predicate to the available descriptions of the relevant propositions (e.g. "the proposition that Peter argued for yesterday is true", "every proposition of the form *p or not p* is true"). Beyond this logical function, there is no role for our concept of truth to play, and a proper account of this concept should say no more than what is necessary for explaining this role. It is not that the concept *cannot* figure in the explanations of the earlier cited principles. It can also appear in accounts of such truth-related phenomena as knowledge or the nature of various types of fact. Horwich's point is rather that the theoretical role that our concept of truth plays in the latter sorts of explanation can be derived from the former logical function, and therefore a theory which can account for that function must be able to carry out these further explanatory duties as well.

The minimalist conception advocated by Horwich is that our concept of truth can be fully characterised by the statement that "the law governing its use is that we are prepared to provisionally accept any instance of the schema, '<p> is true ↔

[3] Horwich (1998b), 2.

p'".[4] Accordingly, a theory of truth should contain nothing more than the uncontroversial instances of the equivalences schema 'It is true *that p* if and only if p'.[5] With this minimal theory in mind, his main purpose is to demonstrate that virtually all phenomena whose explanation was traditionally meant to call for a substantive account of truth can be explained on the basis of the above minimal assumption alone.[6]

How does this minimalist project relate to the naturalistic construal that I wish to advocate in this work? On a short reflection, it may become clear that they are plainly incompatible with each other. First, while on the naturalistic construal to be proposed, truth and falsity are substantive real properties of thoughts and their linguistic expressions that characterise these bearers in virtue of what the latter mean what actually obtains in the spatiotemporal world, the minimalist conception denies that our concepts of truth and falsity stand for such properties in the world. Second, while the naturalistic construal is motivated by the conviction that its substantive assumptions about truth are necessary for explaining all observable phenomena involving this semantic property, the minimalist conception is based on the belief that the explanatory duties in question can be carried out even if we reject those assumptions.

In what follows, I shall argue that the minimalist tenets mentioned in the previous paragraph are equally false. First, I shall show that if we endorse Horwich's positive claim about the notion of truth, namely that it can be truly characterised by the instances of the equivalence schema, and we accept his invitation to adopt a "use-regularity" conception of meaning, which

[4] Horwich (2005), 26.

[5] Horwich (1998b), 7, 19-20.

[6] As Horwich observes, in the case of those *explananda* which concern a relation between truth and an other thing, such as knowledge acquisition, "it is perfectly proper to make use of theories about these other matters, and not to expect all the explanatory work to be done by the theory of truth in isolation. Horwich (1998b), 7. So, on his view, the virtue of minimalism is "that *it provides a theory of truth that is a theory of nothing else, but which is sufficient, in combination with theories of other phenomena, to explain all the facts about truth*". Horwich (1998b), 24-25.

nevertheless preserves the conceptual difference between the actual and the correct declarative use of various sorts of representation, then we can derive the realist idea that truth is the property of having suitable semantic relations to conditions that obtain in the actual world. What this demonstration will show us is that Horwich's negative, deflationist claim about the notion of truth cannot be reconciled with the two positive commitments mentioned above.

As we have seen, according to the minimalist conception, a theory of truth should contain nothing more than the uncontroversial instances of the equivalence schema 'It is true *that p* if and only if p'. As Horwich also recognises, in order to be informative, the account must presuppose that we can understand the right-hand side of these instances without relying on the notion of truth. In other terms, the theory can provide an informative characterisation of truth only if our understanding of propositional contents is not truth-conditional.[7] The way in which Horwich ensures the informativeness of his minimalist conception of truth is by replacing the truth-conditional with a use-theoretic account of meaning.[8] According to this account, the meaning of a term derives from its use, and the acquisition of meaning requires merely the recognition of some basic regularities in use, which can be achieved without reliance on the concept of truth.

Of course, meaning cannot simply be meant to reduce to actual use, since this would, among other things, exclude the conceptual possibility of incorrect use. A further fact about meaning to be taken into consideration is that, once it has been established, no one is conceptually prevented from committing an epistemic mistake while deliberating over the actual applicability of a certain representation, since no one's judgement about this issue is constitutive of the *objective* applicability of that representation. Accordingly, a proper theory of meaning must

[7] Horwich (1998b), 68-71.

[8] For a detailed presentation and defence of this theory, see Horwich (1998a) and Horwich (2005).

construe this semantic property in a way that accounts for the conceptual possibility of incorrect application in the case of each epistemic agent. If the factual ground of this possibility is, indeed, that the correctness of a certain use is independent of our actual beliefs about this issue, then the position at which the previous constraint leaves us is that a proper theory of meaning must construe the conditions of correct use in a realist way.

On Horwich's view, the meaning of a term reduces to a (non-semantic) use or acceptance property which characterises the term due to those *basic regularities* that underlie its application. Accordingly, two terms are supposed to have the same meaning if and only if they instantiate the same use property (i.e. their use displays the same basic regularities). Now, if this "use-regularity" account is supposed to meet the above adequacy condition for a theory of meaning, then it must affirm that the conditions under which the use of a certain representation displays a certain basic regularity obtain independently of our actual beliefs about this issue (i.e. they are to be construed in a realist way).[9]

But how do we arrive from this realist constraint upon Horwich's theory of meaning at a realist challenge to his minimal theory of truth? To see this, we have to return to his conception of how the two theories relate to each other. As we have seen, his view is that our concept of truth can be fully characterised by the statement that we are prepared to apply the predicate 'is true' in the context of the proposition *that p* if and only if *p*. Now, supposing that this conception is correct, and also that the

[9] In *Truth* Horwich addresses the charge that a deflationist perspective leads inevitably to relativism: to the idea that there is no such thing as objective correctness. Horwich (1998b), 52-53. His response, in my view correct, to this challenge is that the denial of a substantive realist account of truth should not be confused with the endorsement of a substantive anti-realist construal of this property, which may, indeed, give rise to "an extreme form of relativism in which it is supposed that truth is 'radically perspectival' or 'contextual' or something of the sort". Horwich (1998b), 53. The current challenge, however, is not that a deflationist semantics denies the existence of objective correctness and entails radical relativism, but instead that it cannot account for the objectivity of correct use.

conditions under which the use of 'is true' in the context of the proposition *that p* can be regarded as objectively regular must be interpreted in a realist manner, we can derive that the right-hand side of the instances of the equivalence schema must be understood also in a realist way (i.e. specifying conditions whose obtaining or absence is independent of what anyone ever believes about this circumstance).[10] According to this reading, what Horwich's minimal theory specifies (again, not necessarily by referring to them) are the objective declarative use conditions of our predicate 'is true' in the context of any conceivable proposition.[11] In possession of this theory, we will hold, indeed, the same capacity as in possession of our concept of truth: namely, for each *understood* proposition, we will be able to entertain an idea of those (non-epistemic) conditions whose obtaining in the world is necessary and sufficient for the correct declarative applicability of our predicate 'is true' in the context of that proposition.

The first important claim that I wish to derive from the previous conclusion is that even if we suppose, I think correctly, that the acquisition of meaning does not require the possession of the concept of truth (for instance, because we can recognise and follow the basic regularities characterising a sentence's use

[10] As it was emphasised earlier in this work, a realist construal of the declarative use conditions of certain representations need not commit us to the realist construal of what these representations purport to be about. This is because truth need not be understood in referentialist terms. Consequently, the current claim that the right-hand side of the instances of the equivalence schema have to be read in a realist way has no implication of the metaphysical status of the subject matter of these propositions either. Of course, if one adopts a referentialist construal of truth, then the commitments in question will follow.

[11] The clause "not necessarily by referring to them" is meant to emphasise the idea that the equivalences stated by the instances of the equivalence schema hold only because the right-hand sides of these instances need not be read in a referentialist way. In the case of propositions about abstract domains, for instance, one can endorse, as I would certainly wish to do, that 'the proposition *that two plus two equals four* is true if and only if two plus two equals four' without thereby subscribing to the referentialist claim that the truth conditions of this arithmetical proposition are to be specified in terms of the abstract subject matter of the right-hand side of this instance of the equivalence schema.

without understanding that the conditions whose obtaining or absence informs these regularities may also be regarded as the conditions of the sentence's truth), since the conditions specified by the right-hand side of the instances of the equivalence schema are the correct declarative application conditions of our predicate 'is true' in the context of the relevant propositions, they can be equally understood as the truth conditions of these truth-apt representations.[12] Having said that, with the former claim that these conditions should obtain independently of our actual beliefs about this issue, we may conclude that, contrary to the deflationist tenet, the truth conditions of our truth-apt representations are to be interpreted in a substantive realist way (i.e. that truth must be regarded as a real property after all). Moreover, if the truth conditions of our thoughts are indeed identical with the correct declarative use conditions of these representations, then truth must be identical with the property of having suitable semantic relations to conditions that obtain in the actual world.[13]

Second, we may notice that the derivation of the above realist construal of truth was based on three major premises. First, we granted that our notion of truth can be truly characterised by the instances of the equivalence schema. Second, we accepted that the meaning of a term derives from its use, and

[12] It may be worth noting that Horwich does not query the correctness of this observation. What he denies is merely that understanding derives from knowledge of these conditions *as* conditions of truth. Horwich (1998b), 69-70. An early recognition of this consequence of the acceptance of the disquotation or equivalence schema can be found in McDowell (1981): "There is a truistic connection between the notion of the content of an assertion and a familiar notion of truth … the connection guarantees, as the merest platitude, that a correct specification of what can be asserted, by the assertoric utterance of a sentence, cannot but be a specification of a condition under which the sentence is true." McDowell (1981), 229. Further endorsements of this point can be found in Wright (1993), 18-19, and Miller (2002), 364-366.

[13] Since, on this realist construal, the truth conditions of our beliefs and utterances are supposed to obtain in reality independently of the occurrence of these representations, truth may also be characterised as a substantive relation ("correspondence") between the latter bearers and the former aspects of reality.

the acquisition of meaning requires merely the recognition of some basic regularities in use, which can be achieved without reliance on the concept of truth. Finally, we maintained that, in order to account for the objectivity of correct use, a proper theory of meaning must construe the conditions of regular applications in a realist way. Since Horwich's minimalist conception contradicts the realist conclusions presented above, at least one of the former three premises must be incompatible with his position. Given that the first two assumptions are explicitly endorsed by him, what he presumably rejects is that a theory of meaning can be adequate only if the conditions of correct use are to be understood in a substantive realist way. In this, however, he could be right only if either a proper theory of meaning did not have to account for the objectivity of regular use, or it could do this without the adoption of the suggested realist tenet. Since, on my view, neither of these conditions obtains, I conclude that Horwich's presumable rejection of the above realist tenet is a mistake, which undermines the explanatory adequacy of his theory of meaning. Moreover, since his minimal theory of truth implies that truth can be understood in terms of regular use, the previous explanatory failure queries the adequacy not merely of his theory of meaning, but also of his theory of truth. If the argumentation that I advanced in this section is correct, then we can establish that Horwich's minimal theory cannot explain all observable facts about truth. In particular, it cannot account for the objectivity of truth.

Finally, we may also note that if we adopt the above realist construal of truth, then Horwich's claim about the explanatory autonomy of our theory of truth with respect to our metaphysical, epistemological and other sorts of beliefs proves to be untenable too. Someone, for instance, who, on epistemological considerations, refuses to believe in the existence of platonic entities may have to implement some modifications in her theory of truth as well. As we have seen in chapter 3, in the semantics of discourses about abstract domains, the received realist, referentialist and non-revisionist construal of truth entails that in absence of platonic entities our widely shared beliefs about these

domains cannot be true. So, if an anti-platonist accepts that a proper theory of truth must be realist in character, and she is also convinced that a large part of our widely shared beliefs in pure logic and mathematics is true, then she must concede that at least one of her remaining two assumptions in the semantics of these discourses has to be modified: she must either give up the idea that the subject matter of pure logic and mathematics is abstract (i.e. non-spatiotemporal) in character, or replace the standard referentialist construal of these truths with a non-referentialist alternative. Either way, her original theory has to be modified, because it cannot account for at least some beliefs in metaphysics and epistemology.

Summing up, in this section, I examined the most fully elaborated version of deflationism in present-day philosophy, Paul Horwich's minimalist conception of truth, and argued that if we adopt, as I think we should, a use theory of meaning that guarantees the objectivity of correct use, then we can derive that Horwich's positive tenet about the notion of truth, namely that it can be adequately characterised by the instances of the equivalence schema, undermines his negative, deflationist claim that truth is not a substantive property whose nature and relation to our epistemic capacities could be further specified. In other terms, what I have shown is that if we adopt the above premises, then the positive tenets of the minimal theory entail that truth must be understood as a substantive real property, the property of having suitable semantic relations to some conditions that obtain in the actual world. Since Horwich explicitly rejects this realist construal of truth, he must also reject the corresponding realist construal of correct use. In absence of these realist tenets, however, his minimal theory will not be able to explain all facts about truth (and correct declarative use in general). In particular, it will not be able to account for the objectivity of this semantic property, the second *explanandum* specified in chapter 2. Finally, we observed that the realist perspective advocated here undermines the explanatory autonomy of the resulting theory of truth. If truth is understood as a real feature of our representations, then our beliefs about, among others, what there

is and how we discover what there is will impose substantial constraints upon our conception of the nature of this semantic property as well.

The core insight behind the above case against Horwich's minimalism, namely that without a substantive realist construal of the right-hand side of the instances of the equivalence/disquotation schema a theory of truth consisting of these instances cannot explain the objectivity of this semantic property, can be invoked in an argument against the explanatory adequacy of any form of deflationism. The specific feature of Horwich's minimalism is that it nicely illuminates the conceptual relation between our theory of truth and theory of meaning, and thus provides us with the necessary premises for deriving our case from basic intuitions about correct use in general, rather than merely about truth in particular.

A common feature of Horwich's minimalist and the suggested realist conception of truth is that they both understand truth in terms of the obtaining of some semantically significant conditions. The difference between the two conceptions is in the interpretation they provide of these conditions (or, rather, of the metaphysical/explanatory impact of their obtaining). According to Horwich, there are no explanatory considerations that would force us to adopt a substantive realist interpretation of the truth conditions whose obtaining is implied by the endorsement of the right-hand side of the instances of the equivalence schema. Thus he can maintain a referentialist construal of these conditions without thereby subscribing to controversial doctrines in metaphysics and epistemology.[14] In contrast, his realist opponent maintains that without a realist construal of these conditions the

[14] Reading Horwich, I have the feeling that one major motivation behind his efforts to develop a minimal theory of truth is the desire to escape the anti-naturalist consequences of traditional (i.e. referentialist) realism about truth in the semantics of our discourses about causally inert domains, a concern fully supported by the current work as well. What Horwich may intend to achieve by the denial of substantive realism about truth, I suggest achieving by the adoption of a non-referentialist form of realism in the semantics of our discourses about causally inert domains.

account cannot be explanatorily adequate, so the controversial doctrines in question must be avoided in some other way: either by insisting that, contrary to appearances, the conditions referred to on the right-hand side of the instances of the equivalence schema obtain within an unproblematic domain, or by conceding that the conditions whose (thick) obtaining is implied by the endorsement of this side are not necessarily the ones we refer to while making this endorsement.[15]

The idea that our semantically distinguishable beliefs have (mostly) different truth conditions enables us to develop a rather fine-grained explanatory apparatus to account for various truth-related phenomena in the world. Suppose, for instance, that I successfully found my glasses on the fridge, and also successfully made a 7 no-trump contract in a bridge game I played this evening. A relatively coarse-grained explanation of these achievements would refer, in both cases, to the fact that my beliefs on which my successful behaviour was based were true. A realist might add, what the minimalist would deny, that the same fact could have been identified by saying that the beliefs in question had suitable semantic relations to the actual world. Nevertheless, this supplement would not alter the fact that the two *explananda* have been explained by reference to the same *explanans*. Now, if our semantically distinguishable beliefs have, indeed, different truth conditions, then my achievements can be

[15] If the right-hand side of the instances of the equivalence schema can specify the truth conditions of the propositions in the left-hand side in a non-referentialist way (i.e. without being about these conditions), as I think it must in the case of knowable propositions about abstract domains, then there must be pairs of beliefs with different semantic contents but identical truth conditions. This is because for each belief with non-referential truth conditions we will find a semantically distinguishable belief with referential truth conditions *about* the obtaining of those conditions. So, the fact that our belief *that p* and our belief *that q* have different semantic contents does not *guarantee* that the truth conditions of *p* and *q* are different too. Semantic contents are more fine-grained than truth conditions. In the case of an arbitrary truth-apt representation *p*, semantic contents include the declarative use conditions of *p* in larger (i.e. embedding) representational contexts as well, while the truth conditions of *p* are the declarative use conditions of *p* in an unembedded state.

explained in a more fine-grained manner as well, which corresponds to the intuition that the explanatory grounds of my success in the search, on the one hand, and in the game, on the other, are not entirely the same. According to these fine-grained accounts, the fact that my glasses were lying on the fridge (partly) explains my success in the search, but not in the game, while the fact that the dummy provided me with sufficiently high cards (partly) explains my success in the game, but not in the search. Clearly, if our semantically distinguishable representations did not have different truth conditions, then such sophistications in our understanding of these truth-related phenomena could not be achieved.

After presenting my argument against the deflationist construals of the notion of truth, in the following section, I shall turn to a case which purports to show that the received realist intuition that our semantically distinguishable true beliefs can refer to, and be made true by, different facts in the world is mistaken. Although the conclusion of the argument is compatible with a degenerated form of the suggested realist construal of truth, nonetheless it provides a serious challenge to the explanatory aspirations of any substantive theory of this semantic property, and thus can be read as an indirect case for the adequacy of deflationism.

2. Slingshot: A Case against the Plurality of Facts

A seemingly powerful objection to the received realist assumption that our true beliefs can refer to, and made true by, different facts in the world is to argue that (i) facts cannot be individuated independently of those representations to which they are supposed to correspond, and (ii) if they are individuated as the referents or truthmakers of particular representations, then we have no means to distinguish them from each other. Clearly, this result, if accepted, is devastating to any substantive (realist or anti-realist) theory of truth which purports to specify the nature of truth in terms of facts, the obtaining of some semantically

significant conditions. An influential argument, commonly known as the "slingshot", has been advanced by a number of philosophers to establish from a few plausible tenets about the representational features of our thought and language the surprising conclusion that all true beliefs or sentences refer to the same thing.[16] If this is so, however, then the "thing" that a true belief is supposed to refer to cannot fulfil the explanatory role that the advocates of substantive theories of truth have attributed to it. Among other things, it cannot explain the intuitive difference between the (referential) truth conditions of various true beliefs. If all true beliefs are made true by the same Great Fact, then it is hard to see why it seems to us that these mental representations have different semantic contents, and their truth can be explained by reference to different aspects of the world.

For the purpose of the current discussion, in what follows, I shall briefly present the Gödelian form of the argument, whose assumptions are less contentious than those employed by the advocates of other formulations. Having reconstructed the case, I shall examine its major presuppositions and explain why I think that it cannot be taken as a convincing point against the realist (or anti-realist) belief in the existence of distinct referential truth conditions. For greater clarity, I shall present the argument using concrete examples. Suppose that the following three sentences are true in the actual world:

1. Peter is not identical with Thomas.
2. Peter is tall.
3. Thomas is short.

[16] It is sometimes assumed that a version of this argument was behind Frege's famous decision to take all true sentences as referring to the value TRUE, while the false ones to the value FALSE. Frege (1892). Alternative formulations of the argument appear, among others, in Gödel (1944), Church (1943), Church (1956), Quine (1953b), Quine (1960b) and Davidson (1969). For a detailed presentation and analysis of these formulations, see Neale (1995), Neale (2001).

Beyond these factual presuppositions, the argument also assumes that the following three semantic principles hold:

I. *Fa* and $a = (\iota x)(x = a \,\&\, Fx)$ stand for the same fact.[17]
II. Any sentence that stands for a fact can be put into a predicate argument form.[18]
III. The referent of a composite expression, containing constituents which themselves have a referent, depends only on the referents of these constituents, and not on the manner in which this referent is expressed.[19]

From these presuppositions, Gödel's slingshot establishes that, contrary to appearances, (2) and (3) stand for the same fact. To see this, consider that, due to the specified assumptions, the consecutive members of the following series of sentences are all coreferential:

i. Peter is tall.
ii. [Peter] is identical with [the object that is (tall and identical with Peter)].
iii. [The object that is (tall and identical with Peter)] is identical with [the object that is (not identical with Thomas and identical with Peter)].
iv. [The object that is (not identical with Thomas and identical with Peter)] is identical with [Peter].
v. [Peter] is not identical with [Thomas]. // [Thomas] is not identical with [Peter].

[17] The definite description '$(\iota x)\varphi$' is supposed to stand for the unique thing which satisfies φ.
[18] Among others, this premise guarantees that an identity claim involving two singular terms ('*a = b*') can be equivalently regarded as the predication of identity of any of the denoted objects with the other (*IdB(a)* or *IdA(b)*). According to this premise all these representations are supposed to stand for the same fact.
[19] In other terms, the substitution of coreferential component expressions is not supposed to change the referent of the composite phrase.

vi. [The object that is (not identical with Peter and identical with Thomas)] is identical with [Thomas].
vii. [The object that is (short and identical with Thomas)] is identical with [the object that is (not identical with Peter and identical with Thomas)].
viii. [Thomas] is identical with [the object that is (short and identical with Thomas)].
ix. Thomas is short.

In particular, (i) and (ii), as well as (viii) and (ix), are coreferential because of (I); (iv), (v) and (vi) are coreferential because of (II); (ii), (iii) and (iv), as well as (vi), (vii) and (viii), are coreferential because of (III) and the coreferentiality of the relevant names and definite descriptions, which is supposed to hold because of (ii) and (iv), on the one hand, and (vi) and (viii), on the other.

The most popular way, among the friends of facts, to avoid the above collapsing conclusion is to adopt Russell's Theory of Definite Descriptions, according to which definite descriptions do not stand for objects as proper names do.[20] By this move, the derivation can be blocked, since the arguments of the main connective 'is identical with' in (ii), (iii), (iv), (v), (vi), (vii) and (viii) will no longer be coreferential, and therefore (III) will not guarantee the coreferentiality of the corresponding sentences either. From a semantical point of view, however, Russell's idea that definite descriptions cannot have (Fregean) referents in themselves, and thus cannot be coreferential with proper names, is at least as counter-intuitive as the anti-factualist conclusion that the move purported to neutralise.[21]

Another response to the argument could draw on the point suggested earlier that claims about relations among truth

[20] Russell (1905).

[21] In a world in which Peter is tall, the definite description 'the object that is tall and identical with Peter' can refer to a particular aspect of the world just as much as any proper name. As I shall argue, the problem with the slingshot is not that it presupposes this referential relation, but rather that it also assumes that the aspect associated with this definite description in the specified world cannot be distinguished from that associated with the proper name 'Peter'.

conditions should not be conflated with claims about relations among referents. If the epistemological case against the platonist construal of truth in the semantics of discourses about abstract domains is correct, then we have good reason to suppose that the truth conditions of these purportedly *a priori* commitments are different from the states of affairs that they intend to be about. Consequently, the conclusion of the slingshot should not be taken to establish anything about the truth conditions of these beliefs. One problem with this reaction is that it neutralises the slingshot only in the case of those beliefs whose truth conditions are non-referential in character. In the case of our broadly physicalistic claims, however, the adequacy of standard referentialism about truth has not been ruled out by the rejection of semantical platonism. Consequently, one might insist that the argument still demonstrates that at least in the case of our discourses *about the spatiotemporal world* all true beliefs are made true by the same thing. The second problem with the above reaction is that the slingshot's conclusion is counter-intuitive independently of what we believe about the relation of truth and reference. The surprising thing in this conclusion is that it contradicts our fundamental conviction that our true beliefs can have referential relations to different aspects of the world.

In my view, the real problem with the slingshot is that its presuppositions are informed by the Fregean coarse-grained idea of reference, which does not grasp those fine-grained referential aspects of our language which lend support to a referentialist understanding of truth conditions. The point will be clearer after examining the semantical assumptions (I) and (III) from the suggested referentialist point of view. Let me start with the latter. According to this assumption, for instance, the sentences (ii), (iii) and (iv) are coreferential. Why? Because all three are meant to refer to, and *prima facie* even made true by, the self-identity of the very same object, which is alternatively denoted by 'Peter', 'the object that is tall and identical with Peter', and 'the object that is not identical with Thomas and identical with Peter'. Although we can maintain that the coreferentiality of the three sentences relies on the obtaining of two further facts, namely Peter's being tall,

on the one hand, and his not being identical with Thomas, on the other, nevertheless, the three identity claims do not *refer to* those further facts, at least not in the received Fregean sense of the term. Frege's notion is intuitive enough. Of course, there is a sense in which the three identity claims are, indeed, *about* the same thing, namely the self-identity of the denoted object. This sense, however, is too coarse-grained to support a referentialist construal of truth conditions. This is manifest in the linguistic fact that while we may accept the claim that the three identity statements refer to the same fact, and at first glance maybe also that they are made true by the same fact, nonetheless, after realising the significance of those two further facts to the truth of (ii), (iii) and (iv), we become certainly more reluctant to subscribe to the second claim, independently of our view about the first. We will probably move toward the opinion that the truth conditions of (ii) involve the state of affairs that Peter is tall, those of (iv) that Peter is not identical with Thomas, while those of (iii) both of these states of affairs. Again, what this may show us is that our idea of truth conditions is more fine-grained than the Fregean concept of reference.

This result would already be enough to defend the classical realist view that our true beliefs possess their truth value in virtue of the obtaining of various states of affairs from the challenge of the argument. On the other hand, departing the field with this conclusion would leave our concept of reference, together with our notion of facts that we are supposed to think of or speak about, prey to the slingshot. The situation, however, does not seem to be even that bad. For there is clearly a sense in which (ii), (iii) and (iv) are not about the same aspect of the world. While (ii) is about Peter's being identical with the object that is tall and identical with him, (iv) is about Peter's being identical with the object that is not identical with Thomas and is identical with him. To be sure, in this fine-grained sense of the term, reference would embrace everything that Frege wanted to separate under the notion of sense. The fact that his terminology had well-known motivations in linguistic phenomena does not imply, however, that his choice optimally observed all theoretical

concerns. If we want to use the notion of reference to denote a substantive and sufficiently fine-grained semantic relation between our beliefs or sentences, on the one hand, and some particular aspects of the world, on the other, then we may well have to reconsider our received Fregean terminological heritage. Adopting the suggested fine-grained concept of reference, our principle of compositionality would still hold in the following, slightly modified form:

> III' The referent of a composite expression, containing constituents which themselves have a referent, depends on the referents of these constituents.

After abandoning the distinction between sense and reference, the clause 'and not on the manner in which this referent is expressed' at the end of (III) will lose its significance.

Turning to the first semantical assumption of the argument, we may draw more critical conclusions. Here again, we might admit that, in the Fregean coarse sense of 'reference', the sentences 'Peter is tall' and 'Peter is identical with the object that is tall and identical with Peter' refer to the same thing. Why? Because both are supposed to be made true by the same fact, namely Peter's being tall. In this case, to reply that 'because both are about Peter's being tall' is even less intuitive than the corresponding answer to explain the coreferentiality of (ii), (iii) and (iv) above. Be this as it may, we can certainly identify a sense in which (i) and (ii) are not about the same aspect of the world either. While (i) is about Peter's being tall, (ii) is about Peter's being identical with the object that is tall and identical with him. Adopting, again, this fine-grained concept of reference, we should conclude that the first assumption of the argument is simply false. Moreover, this construal would not rule out the referentialist understanding of truth conditions either. For, arguably enough, the truth conditions of (i) and (ii) are just as different as they were in the case of (ii), (iii) and (iv). The clearest evidence of this difference is the fact that if Peter were not tall, then (i) would be false (on *any* interpretation), while (ii) would be

either false or neither true nor false (depending on the semantics of the connective 'is identical with' and the definite description 'the object that is tall and is identical with Peter').

Summing up, in this section, I have argued that the case presented against the idea of separate facts, subject matters and truth conditions, and thus against any substantive theory that purports to specify the nature of truth in terms of distinct subject matters, relies on an intuitive, but rather coarse-grained idea of reference, which does not grasp those fine-grained referential aspects of our language that may lend support to a referentialist understanding of truth conditions in the semantics of broadly physicalistic discourses. If we realise that the standard referentialist conception of truth is based on a fine-grained notion of subject matter, which preserves the intuitive difference between the referential aspects of, say, (i) and (ii) or (ii) and (iii) and (iv), then we also understand why the slingshot provides us with such a surprising conclusion. It does so because it invites us to ignore a substantial part of those compositionally determined semantic relations that our truth-apt mental and physical representations have to various aspects of the world. If we refuse this invitation, then the slingshot collapses, and the idea of correspondence between truthbearers and particular aspects of reality can be maintained without inconsistency.

Summary

In this chapter, I examined the viability of the deflationist version of standard referentialism in the semantics of our cognitive discourses in general, and argued that the advantages the theory has (including the capacity to neutralise Benacerraf's challenge without abandoning referentialism in the semantics or inventing dubious explanations in the epistemology of pure logic and mathematics) are cancelled out by the fact that a deflationist can provide no suitable explanation of the objectivity of truth.

In section 1, I attempted to demonstrate this point by examining the most fully elaborated form of deflationism, Paul

Horwich's minimal theory of truth. Beyond providing a case for the claim that a proper account of this *explanandum* requires a substantive realist construal of truth, I have also shown that Horwich's positive tenets about meaning and truth entail the adequacy of a version of the standard realist correspondence theory of truth.

In section 2, I presented an argument which has been used to challenge an important presupposition of this realist account, namely that our semantically distinguishable true beliefs can refer to, and be made true by, different facts in the world. Having examined the major assumptions of the argument, I observed that at least two of them rely on a rather coarse-grained notion of reference, which cannot grasp those highly differentiated referential aspects of our thought and language that may lend support to a referentialist understanding of truth conditions in the semantics of broadly physicalistic discourses. The conclusion I drew was that with the adoption of a sufficiently fine-grained notion, we can easily block the slingshot and save the integrity of the suggested realist position.

In the following chapter, I shall turn to the second referentialist response to Benacerraf's dilemma left standing in chapter 3, which opposes the above realist account on epistemological considerations from the perspective of a substantive anti-realist theory of truth.

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

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

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

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]

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

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

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

[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).

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

[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

[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.

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

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

[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.

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

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]

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

[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.

[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

[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).

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

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]

[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.

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:

[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

[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).

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-

[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

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.

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

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

[52] Quine (1960b).

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

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

[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.

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:

[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

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

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

[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

[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.

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.

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

[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.

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

[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]

[65] Putnam (1981), 46.
[66] The point was stressed by Lewis (1984).

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.

[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,

[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.

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

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

[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

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

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

[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*.

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

[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

[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.

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

[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

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

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.

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

[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.

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

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.

CHAPTER 6

Referentialist Responses to Benacerraf's Dilemma III: Platonist Construals of A Priori Truth and Knowledge

Introduction

In the previous two chapters, I argued that in order to explain the objectivity of our apparently legitimate knowledge claims, we need a substantive theory of truth whose subject matter is understood along realist lines. Since our purportedly *a priori* judgements about abstract domains seem to be as objective as our ordinary empirical beliefs about the spatiotemporal world, the above result is supposed to hold across the board in the semantics of all cognitive discourses, whether *a priori* or empirical. In chapter 3, however, we saw that in the standard referentialist framework a realist construal of truth in the semantics of claims about causally inert subject matters is incompatible with a causal theory of how we can acquire knowledge or reliable beliefs about such domains. So, if the explanatory considerations in support of realism about truth advanced in chapter 4 are correct, and consequently neither the deflationist nor the anti-realist forms of referentialism can provide a suitable response to Benacerraf's updated and generalised challenge in the philosophy of discourses about causally inert domains, then we must either abandon the standard referentialist framework and develop a non-referentialist construal of the problematic truths, or come up with a new account of how we are supposed to acquire knowledge or reliable beliefs about the obtaining or absence of causally inert conditions in the world.

In this chapter, I shall examine the prospects of the latter strategy in the semantics of our paradigm *a priori* discourses, such as logic and mathematics. The primary purpose of this

investigation is to show that the advocates of platonism about the relevant truths have no suitable answer to Benacerraf's epistemological challenge presented in chapter 3, and that the responses they have actually given to this challenge are either *ad hoc* and uninformative, undermining all constructive methods for examining the nature and shortcomings of the relevant forms of belief formation, or insufficient, leaving us without any positive reason for supposing that the conditions whose obtaining they assume to be necessary and sufficient for the truth of our logical and mathematical beliefs indeed obtain in the intended platonic realms.

In section 1, I shall briefly review the most important explanatory considerations that may be raised for and against the platonist construal of our paradigm *a priori* truths. In section 2, I shall first examine the available platonist responses to what I regard as the most fundamental objection to this doctrine, namely Benacerraf's challenge that a platonist referentialist has no suitable account of how we could develop knowledge or reliable beliefs about platonic objects and properties, and then explain why I think that none of these responses can save the adequacy of the platonist construal of truth in the semantics of our paradigm *a priori* discourses about abstract domains.

1. The Explanatory Virtues and Vices of Platonism about A Priori *Truth*

As a point of departure, let me recall the most important motivations behind the platonist understanding of our knowledge claims about abstract states of affairs. According to Benacerraf, as we saw in chapter 3, one virtue of this construal is its *homogeneity* with our standard referentialist account of truth in the semantics of broadly physicalistic discourses. The essential tenet of this account is that the truth conditions of our legitimate knowledge claims are those states of affairs that these claims purport to be about. Since in the case of our paradigm *a priori* discourses the intended referents of our sentences are arguably abstract states of

affairs, the standard realist construal of these truths is certainly that provided by the advocates of semantical platonism. So, the first advantage of the platonist construal is that it meets the first adequacy condition set for such an account in the second part of chapter 2.

Beyond its homogeneity with our standard referentialist semantics, the construal provides a realist picture of the truth conditions of our paradigm *a priori* beliefs, and thus possesses the necessary explanatory resources to account for the objectivity of these truths. In this manner it satisfies the second adequacy condition listed in chapter 2 as well.

In chapter 5, we saw that the two most influential "anti-realist" attacks on the idea that our thoughts and sentences may have determinate semantic relations with those (often verification-transcendent) states of affairs that they seem to be about are equally unsound. In view of this result, we may assume that the platonist construal of the truth conditions of our paradigm *a priori* beliefs can be supplemented with a suitable account of the emergence of determinate semantic relations between the relevant *a priori* beliefs, on the one hand, and their intended abstract subject matter or truth conditions, on the other. At the very end of chapter 5, I provided an outline of a (moderate) causal account of reference determination, which, if true, also shows how the platonist construal satisfies the third adequacy condition on our list in chapter 2.

In possession of this account, a platonist can also explain how our paradigm *a priori* beliefs can be about an abstract and infinite domain (cf. the eleventh *explanandum* in chapter 2), and how we can develop, in principle, infinitely many semantically different representations of these domains (cf. the seventh *explanandum* in chapter 2). Her accounts can draw on the compositional character of semantic content. As to the former *explanandum*, one can think of an abstract domain by thinking of a domain that is not spatiotemporal, and one can think of an infinite domain by thinking of a domain that is not finite in character. The elements of this domain can be distinguished by invoking those (infinitely many) essential properties that can be

recursively composed of the atomic properties of the domain. As to the latter *explanandum*, since the platonist understands the semantic content of our claims about abstract domains along the referentialist lines, she can explain the infinity of semantically different truth-apt representations within our paradigm *a priori* discourses by invoking the infinity of the available abstract truth conditions within the intended abstract domains.

Supposing that the abstract conditions thus related to our paradigm *a priori* beliefs are identical with the truth conditions of these truth-apt representations, semantical platonists can considerably enhance the apparent explanatory adequacy of their theory. By stipulating the necessity of the obtaining (or the absence) of the intended abstract conditions, for instance, they seem to be able to explain the necessary truth or necessary falsity of our paradigm *a priori* beliefs (cf. the ninth *explanandum* in chapter 2). Further, supposing that we have some sort of epistemic access to the intended non-spatiotemporal realms, the advocates of the platonist construal may offer a simple and natural account of the apriority of the evidential grounds or ways of justification that we rely on during the formation of these beliefs (cf. the eighth *explanandum* in chapter 2). Finally, by assuming the previous access, they can also explain the intersubjectivity of the relevant semantic contents and the observable convergence of the relevant beliefs (cf. the fifth and sixth *explananda*, respectively, in chapter 2).

In view of this remarkable explanatory potential, it is no wonder that platonist construals of the truth conditions of our paradigm *a priori* beliefs still preserve their appeal in present-day analytic philosophy. Nevertheless, as it is often emphasised, there are at least two major difficulties with this theory in the semantics of discourses about abstract states of affairs. The first is that the categorical separation of the domain of pure mathematical and logical beliefs from that of our empirical claims about the spatiotemporal world makes it hard to explain the applicability of the former types of knowledge claims in the empirical sciences

(cf. the tenth *explanandum* in chapter 2).[1] The second, explicated so aptly by Benacerraf and recast later in slightly different terms by Field, is that if the truth conditions of our logical and mathematical beliefs have no causal interaction with, or no influence on, the natural world, then the obtaining or absence of these conditions cannot be detected by spatiotemporally located human minds, which means that the platonist theory undermines the possibility, and the proper explanation, of mathematical and logical knowledge or reliable belief formation (cf. the third *explanandum* in chapter 2). In absence of a coherent notion of epistemic access to platonic realms, one may query the adequacy of the advanced platonist explanations of apriority, intersubjectivity and convergence as well, which considerably reduces the explanatory power of the platonist construal.

Of course, many platonists are aware of the significance of the previous explanatory difficulties, and they have various proposals of how these problems can be properly dealt with in the suggested platonist theoretical framework. In the following section, I shall examine the most influential platonist theories of how we can acquire knowledge or reliable beliefs about the posited platonic domains. I shall grant that the viability of any of these accounts would provide a decisive case for the adequacy of platonism about truth in the semantics of our paradigm *a priori* discourses, and thus support the cogency of standard referentialism as a conception of truth in general. The main message of the section, however, will be that the explanatory problem specified by Benacerraf's updated and generalised challenge presented in chapter 3 cannot be suitably resolved by any of the advanced platonist epistemologies. In view of this result, I shall conclude that the correct realist response to Benacerraf's challenge in the semantics of our paradigm *a priori* discourses (and our cognitive discourses about causally inert conditions in general) is to abandon the received referentialist construal of these truths and adopt a naturalistic understanding of

[1] For a detailed study of the problem of applicability of mathematics in our empirical theories of the world see Steiner (1998).

them that is compatible with a suitable account of all the *explananda* listed in chapter 2, and thus satisfies all adequacy conditions identified there for a proper theory of *a priori* truth.[2]

2. Platonist Accounts of Knowledge of Abstract Domains

Adopting Mark Balaguer's classification, platonist replies to Benacerraf's challenge can be grouped into two major categories.[3] One group maintains that, contrary to Benacerraf's assumption, human minds are capable of developing an epistemic access to platonic entities, while the other believes that knowledge of abstract domains does not require the mind's interaction with the obtaining platonic truth conditions. Following Balaguer's terminology, I shall call the two sorts of accounts, respectively, "contact theories" and "no-contact theories" of our knowledge of platonic states of affairs.[4]

[2] Since I believe that the epistemological challenge cannot be properly answered in the suggested platonist framework, I shall take it that the advocates of this construal have no suitable account for the apriority, intersubjectivity and observable convergence of the relevant beliefs either. Further, I believe that a platonist conception of truth in the semantics of pure logic and mathematics also undermines the proper understanding of how these purportedly *a priori* theories can be applied in our empirical accounts of the spatiotemporal world. In this chapter, however, I shall not argue for these further negative claims. I take it that the case provided here against the available platonist epistemologies is sufficient to show the inadequacy of the platonist construal. On the other hand, in chapter 7, I shall show that the suggested non-referentialist conception resolves all explanatory puzzles surrounding the relevant truths, including those left unexplained in the platonist semantical framework.

[3] Balaguer (1998), 24-25.

[4] In the philosophy of mathematics, Gödel's sporadic remarks on mathematical intuition are examples of the first theoretical alternative (Gödel (1944), 449, Gödel (1951), 310-312, Gödel (1964), 483-484). His views have been recently adopted by Brown (1991). BonJour's account of *a priori* knowledge provides a further, though slightly different, example of the first category. BonJour (1998). Authors falling into this contact theorist camp attempt to block Benacerraf's epistemological challenge by rejecting the first premise of his original argument reconstructed in chapter 3 (viz. that human beings exist entirely within space-time). Balaguer

The main problem with the contact theorist solutions is that their notion of a specific, non-causal information conveying channel between our minds and the alleged platonic realm is *ad hoc* and exotic in character. It is *ad hoc*, because by positing this epistemic link we can gain an account merely of the problematic *explananda* (knowledge acquisition and therewith apriority, intersubjectivity and convergence), and it is exotic, because we have no idea (neither *a priori* nor empirical) of the nature and working of this epistemic link in the actual world.

To appreciate this point, consider our reasons for adopting a contact theory of knowledge acquisition of the spatiotemporal world. By positing an epistemic contact (an information-conveying causal mechanism) between human minds and their natural environment, we can explain not merely the possibility of human knowledge of this domain (with the aposteriority, intersubjectivity and observable convergence of the relevant beliefs), but also our actual experience of the posited contact or mechanism.[5] In view of this extra explanatory impact, we can clearly reject the charge that our belief in this contact is *ad hoc* in character. Further, our theory of perception is an articulated, evidence-governed account of how the obtaining of natural states of affairs may influence our experience. It provides us with a

argues that the account developed by Maddy (1980) and Maddy (1990), according to which mathematical objects are spatiotemporal and human beings can acquire knowledge of them via sense perception, is also a contact theorist response to Benacerraf's case, although it queries the second, rather than the first, premise of the argument (viz. that if there exist any abstract mathematical objects and properties, then they exist outside space-time). In this chapter, I shall ignore the latter response, first, because it does not understand platonism in the traditional sense of the term, and second, because Maddy (1997) has abandoned this conception after all. Major examples of no-contact theories include Quine (1951), Steiner (1975), Parsons (1980), Parsons (1994), Katz (1981), Katz (1995), Resnik (1982), Resnik (1997), Wright (1983), Lewis (1986), Hale (1987), Shapiro (1989), Shapiro (1997) and Balaguer (1998).

[5] Evaluating Putnam's just more theory argument against causal theories of reference determination, in chapter 5, I set forth which aspects of our experience can be taken as caused by the obtaining of the posited contact between our minds and their natural environment.

detailed picture of the nature of knowledge acquisition, and informs us about how we could eliminate our epistemic mistakes and improve the accuracy of our belief formation. To put it shortly, our idea of this contact is not exotic either. We can say that the observability of the posited epistemic contact is fairly expectable and also highly significant from the perspective of our cognitive purposes. On the one hand, it seems quite natural to suppose that if we can detect what obtains in some part of the mind-independent world, then in principle we must be able to detect the actual exercising of this epistemic capacity (in the same part of the world) as well. On the other hand, the observability of this contact is vital for both the improvement of our cognitive performance and our capacity to distinguish genuine knowledge from those Gettier cases in which our beliefs happen to be true without being properly informed by their obtaining truth conditions.

In contrast, the nature of our epistemic contact with the allegedly obtaining platonic truth conditions of our pure logical and mathematical beliefs seems fully inscrutable. We know that it cannot be causal in character. We are supposed to know this on *a priori* grounds: platonic entities cannot enter into causal relations. A platonist may add that the acquisition of this knowledge is also due to the existence of some epistemic contact between our minds and the obtaining truth conditions of this claim.[6] The assumption that our knowledge of the relevant contact is also due to the existence of such a contact does not create vicious circularity in the argumentation. Obviously, our knowledge of the nature of human perception (i.e. the epistemic contact between human minds and their natural environment) also presupposes the existence of the described perceptual links between our minds and the relevant perceptual relations in the world. The real problem with the platonist conception is rather that we can detect neither the actual obtaining (or absence) nor the existing

[6] One may wonder what a platonist would say about the nature of these truth conditions, as they involve the existence of a certain type of relation *between* the distinguished platonic and spatiotemporal realms.

characteristics of the posited epistemic relation in the actual world.

The lack of (*a priori* or *a posteriori*) observational evidence of the suggested epistemic contact between us and various platonic domains has also important theoretical and practical consequences. First, in absence of this observational ground we have no more reason for believing in the existence of the posited contact than a religious fundamentalist would have for her belief in the existence of a corresponding epistemic link between her mind and the allegedly obtaining truth conditions of her religious beliefs.[7] Second, in absence of that ground we cannot tell apart genuine knowledge from luckily acquired true beliefs either (i.e. we cannot tell whether a certain piece of *a priori* evidence in someone's mind is part of the posited epistemic link between that mind and the obtaining truth conditions of the relevant *a priori* beliefs, or rather it is created by some natural mechanisms that are entirely independent of what obtains in the posited platonic domains).[8] Finally, in absence of the relevant observations, we cannot develop an articulated theory of the nature of *a priori* knowledge acquisition, and we cannot learn how we could eliminate our epistemic mistakes and improve the accuracy of our belief formation in our discourses about abstract domains.

Summing up, contact theorist responses to Benacerraf's original or modified and generalised challenge to platonism about truth in the semantics of our discourses about causally inert

[7] The fact that the congruent belief-systems of various religious communities (or individuals) are incompatible with each other need not undermine the appeal of the suggested contact theory of religious knowledge acquisition. For someone who believes in that contact, it merely demonstrates that some of the conflicting alternatives must be mistaken and consequently the stipulated link does not provide us with absolutely reliable beliefs.

[8] One may think that the latter difficulty can be avoided by denying that beliefs about abstract domains are based on spatiotemporal evidence, and maintaining that knowledge within the relevant discourses consists in a direct grasp of the obtaining platonic truth conditions. Note, however, that this solution would undermine the distinction between justified and unjustified true beliefs about abstract domains, and thus contradict some of our basic intuitions in the epistemology of pure logic and mathematics.

domains are all inadequate, in so far as the explanation they provide of our knowledge or reliable belief formation about these domains is *ad hoc* and exotic in character. The adoption of this platonist epistemology would discourage any further inquiry into the nature of this type of knowledge acquisition, and it would open the door for parallel stipulations in the case of any other knowledge claims, no matter how the beliefs in question were causally produced in the subjects' minds.

The majority of contemporary platonists in philosophy of mathematics prefer the second type of account of how knowledge of abstract states of affairs is possible. The common feature of these accounts is the denial of the *prima facie* plausible claim that mathematical knowledge requires epistemic contact with the obtaining truth conditions of our correct mathematical beliefs. Instead, the advocates of these no-contact theories argue that some specific properties of these platonic truth conditions and/or the way we develop our beliefs about them in the spatiotemporal world guarantee and explain the possibility of mathematical knowledge or the reliability of mathematical belief formation.[9] Of course, the crucial question in this case is whether the invoked characteristics are indeed sufficient for ensuring the envisaged result. The bare stipulation of the possibility of mathematical knowledge acquisition or reliable belief formation in space and time would hardly satisfy those who share Benacerraf's reservations about the platonist construal of mathematical truth.[10]

[9] Authors falling into this no-contact theorist camp attempt to block Benacerraf's epistemological challenge by rejecting the third premise of his original argument reconstructed in chapter 3 (viz. that if there exist any abstract mathematical objects and properties, then human beings cannot have knowledge of them), while maintaining the first and the second about the nature of human beings and mathematical entities, respectively.

[10] As Field formulated, "special 'reliability relations' between the mathematical realm and the belief states of mathematicians seem altogether too much to swallow. It is rather as if someone claimed that his or her belief states about the daily happenings in a remote village in Nepal were nearly all disquotationally true, despite the absence of any mechanism to explain the correlation between those belief states and the happenings in the village". Field (1989), 26-27. Balaguer (1998) formulates the same point against Parsons's theory of mathematical knowledge, which purports to explain the phenomenon by invoking the epistemic capacity of intuiting "quasi-concrete" objects (i.e. types of perceivable tokens) without explaining why the deliverances of this capacity would provide reliable information of "purely abstract" (i.e. platonic) states of affairs. Balaguer (1998), 38, esp. footnote 46, and Parsons (1980).

One influential strategy to account for the possibility of mathematical knowledge, which may be regarded as a no-contact theory of knowledge of platonic facts, is to adopt Quine's holistic theory of confirmation, and argue that our mathematical beliefs constitute a (maybe indispensable) part of our overarching theory of the world, which as a whole is confirmed by the (naturalistically construable) deliverances of our external and internal senses, in other words by our experience.[11] Of course, this account can be taken as a no-contact theory of knowledge of platonic facts only if we suppose that the truth conditions of our mathematical beliefs obtain in a platonic realm. Whether Quine himself maintained a platonist construal of mathematical truth is rather questionable. Those who believe that he did may rely on his famous verdict on ontological commitment. According to this proposal, our prevailing idea of what there is is determined by our best overall theory of the world: in particular, we are committed to the existence of those entities that this theory happens to quantify over.[12] Since, in Quine's view, mathematics is an integral part of our best overall theory of the world, one may conclude that Quine must have believed in the existence of mathematical objects.[13] Note, however, that Quine is also famous

[11] Balaguer (1998), 40-41. Advocates of this account of mathematical knowledge include Steiner (1975), Resnik (1997) and Colyvan (2001), Colyvan (2007). We may note also that Field's reading of Benacerraf's epistemological challenge, according to which the problem with the platonist construal is that it cannot explain the reliability of our actual mathematical beliefs, was largely put forward also as a reaction to this Quinean response to Benacerraf's original argument. Field (1989), 25.

[12] Quine (1948), Quine (1951).

[13] The same reconstruction of Quine's platonist reading appears in Hellman (1989), 3, fn. 1. In so far as the theory can be formulated, as Quine believed, in first-order language, the commitments in question will merely extend to the domain of the first-order variables.

for his deflationist (disquotational) theory of truth, which is hardly compatible with the substantive realist construal of the obtaining truth conditions of true mathematical beliefs implicated by the platonist construal.[14] Further, Quine's empiricist theory of confirmation was meant to provide an account of mathematical knowledge only in so far as mathematics is (indispensably) applied in our best overall scientific theory of the spatiotemporal world.[15] Since the truth conditions of this theory are supposed to be spatiotemporal, it is far from obvious how Quine's epistemology could qualify as a no-contact theory of knowledge of platonic domains.

Now, of course, independently of these interpretative questions, one may adopt the above empiricist strategy and maintain that our external and internal senses (i.e. our causal-epistemic contacts with the natural world) provide us with knowledge not merely of the spatiotemporal world, but also of the obtaining platonic truth conditions of our claims about abstract domains. In the philosophy of the relevant discourses, such as pure logic and mathematics, this account clearly qualifies as a no-contact theory of knowledge of platonic domains, since it preserves the idea that the truth conditions of these beliefs obtain in a platonic realm, and it does not presuppose the existence of any contact between the posited platonic objects and properties and our knowing minds. Instead of stipulating such a contact, the account rather explains the way we acquire knowledge of platonic entities by emphasising that this knowledge is inseparable from (and maybe indispensable to) our knowledge of the

[14] Quine (1970).

[15] Although Quine was quite hesitant about how much of our mathematical theories can be legitimately justified on holistic empirical considerations, he made it explicit several times that those parts that are demonstrably independent of the applicable pieces must be regarded as results of mathematical recreation, whose acceptance does not presuppose specific ontological commitments on our part. Quine (1986), 400, Quine (1995), 56-57. For a defence of this Quinean semantical division within mathematics see Colyvan (2007).

spatiotemporal world.[16] In fact, we can justify our beliefs in the existence of the relevant platonic objects and properties in the same way as we justify our beliefs in the existence of theoretical entities posited by our best overall theory of the spatiotemporal world: first, we justify our overall theory holistically, in view of its predictive success, on empirical considerations, and then, we understand its truth in referentialist terms (i.e. in terms of the existence of those entities and the obtaining of those conditions that the theory purports to be about). The question, of course, is whether we can indeed legitimately suppose such a parallelism between the two types of knowledge acquisition.

Opponents may query the correctness of this assumption on various considerations. They may observe, for instance, that platonic objects and properties, unlike electrons and their particular features in space and time, cannot be invoked in causal explanations of phenomena, so their existence cannot have the same explanatory role in our theory of knowledge as that of entities posited by our best overall theory of the spatiotemporal world. One may think that the contrast mentioned by these opponents wrongly presupposes that the causally relevant properties of electrons and other theoretical entities can be represented without reliance on concepts of abstract objects and properties.[17] But to think that they cannot seems to rest on the relatively entrenched and arguably mistaken view that our concepts of mathematical and logical properties appearing in space and time are dependent on (or posterior to) our concepts

[16] Today, it is often granted that the only serious consideration in support of the platonist belief in the existence of (non-spatiotemporal) mathematical objects is Quine and Putnam's indispensability argument. Classical formulations of the argument occur in Quine (1948), Quine (1960a), Putnam (1971), and Putnam (1975d).

[17] The point appears in Field (1989). His conclusion on the matter is that "unless a very substantial amount of explanation involving electrons can be given in a mathematical entity-free fashion, the prospects for maintaining realism about electrons without maintaining platonism are dim". Field (1989), 19-20. Earlier Field made an attempt to demonstrate that our scientific theory of the world does not rely essentially on ideas of platonic entities. Field (1980).

of platonic objects and properties in pure logic and mathematics. In chapter 7, I shall argue that the dependence between these types of concepts indeed obtains, but it holds the other way round: our ideas of abstract (i.e. non-spatiotemporal) entities in pure logic and mathematics are dependent on (or posterior to) our concepts of mathematical and logical properties appearing in space and time.[18] If so, however, then the previous observation of the categorical difference between the explanatory role of platonic objects and properties, on the one hand, and those of the theoretical entities of our empirical sciences, on the other, may be fully adequate. In view of the causal inertness of platonic objects and properties, our belief in their existence cannot be based on the same explanatory considerations as our belief in the existence of the theoretical entities of the spatiotemporal world.

Realising that our reasons for believing in the existence of theoretical entities (together with their logical and mathematical properties) in space and time cannot support our belief in the existence of those abstract objects and properties that constitute the subject matter of pure logic and mathematics may lead the (realist) opponents of the empiricist account in one of the following two directions: they may either accept confirmation holism (i.e. empiricism) in the epistemology of our discourses about abstract domains and reject the referentialist construal of truth (i.e. platonism) in the semantics of these sorts of claims, or they may reject confirmation holism and check whether there is a better (presumably apriorist) no-contact epistemology which can account for our knowledge or reliable belief formation in discourses about abstract domains in a referentialist semantical framework.[19] If they choose the second option and their answer to the previous question is negative, then they must conclude that a correct theory of knowledge of abstract domains requires the adoption of a non-referentialist construal of the relevant truths.

[18] For a proper explication of the often conflated notions of abstractness, see section 5 in chapter 1.

[19] I ignore here the realist reactions disqualified earlier in chapters 3, such as scepticism and revisionism about subject matter.

Either way, the idea of an empiricist no-contact theory of our knowledge of platonic domains is abandoned for the sake of some alternative.

Although the adoption of the first alternative (i.e. empiricism with non-referentialism) is outside the scope of the current chapter, it may be worth briefly reviewing the two main reasons for which this strategy is seen by many philosophers as a non-starter.

First, one may observe that our beliefs in pure logic and mathematics do not rely on the applicability of these theories in our best overall theory of the spatiotemporal world. The fact, for instance, that Euclidean geometry is strictly speaking no longer applied in our scientific theory of the physical world does not influence our beliefs about the objects and properties posited by this geometry. Apparently, once we develop our concepts of the relevant abstract subject matters, our reasons for adopting or rejecting a claim that involves some of these concepts will have nothing to do with our empirical findings about the spatiotemporal world. Of course, this is merely a restatement of the eighth major *explanandum* put forward in chapter 2: namely, that our beliefs about abstract domains are based on *a priori* evidence.[20] So, the first major objection to the empiricist account under discussion is that it runs counter the apparent apriority of our knowledge of the relevant abstract domains.

The second objection is strongly related to the first. It starts with the observation that the truth or falsity of our beliefs about abstract domains is necessary in character (cf. the ninth *explanandum* in chapter 2), and then it lays down that our experience can never provide us with reasons for believing that something is necessarily true or necessarily false. In other terms, the second major objection to the empiricist theory is that it fails

[20] To avoid empty terminological objections to the idea that our beliefs about abstract domains are *a priori* in character, see my clarificatory notes on apriority in the first section of chapter 1.

to account for our knowledge of the modal character of the truth value of our beliefs about abstract domains.[21]

In view of these explanatory problems with the empiricist account of our knowledge of abstract states of affairs, we may conclude that the more plausible move after the realisation of the inadequacy of the earlier "Quinean" strategy is to proceed in the second direction (i.e. to reject radical confirmation holism and check whether there is an apriorist no-contact epistemology which could account for our knowledge of abstract domains in a referentialist semantical framework).

One influential example of such no-contact theorist replies to Benacerraf's challenge has been advanced by Bob Hale and Crispin Wright.[22] According to their view, our ability to develop thoughts and acquire knowledge of platonic states of affairs cannot be problematic, as Benacerraf implies, since the truth conditions of these beliefs are, in fact, identical with the truth conditions of some other, semantically and epistemologically unproblematic, claims. The core idea behind this suggestion is that we acquire our concepts and knowledge of platonic entities by Fregean abstraction and deductive inference without ever being acquainted with (or influenced by) the entities themselves.

Consider, for instance, the case of our concepts and knowledge of natural numbers. According to Frege, the acquisition of numerical concepts requires the acquisition of the truth conditions of all identity statements involving these concepts. In Hale and Wright's reconstruction, the latter can be done by means of what has come to be called *Hume's Principle*:

(HP) The number of *F*s = the number of *G*s ↔ there is a one-one correspondence between the *F*s and the *G*s.[23]

[21] The objection plays a central role in Katz's argumentation against empiricism in the epistemology of discourses about abstract domains. Katz (1981), 208.

[22] Hale and Wright (2002).

[23] As Hale and Wright rightly observe, if the principle is meant to inform us about the truth conditions of the identity statements on the left hand side of the sentential connective '↔', then the connective must be taken to indicate that the truth conditions of the connected expressions are identical, rather than merely that the connected statements are materially equivalent. Hale and Wright (2002), 117. Note also that Frege did not think that we can contextually define our numerical concepts by means of Hume's Principle, since the latter does not appear to enable us to settle the truth value of identity claims that link denoting expressions involving numerical concepts of the form 'the number of …' with others not doing so, such as the claim "the number of Jupiter's moons = Julius Caesar". Frege (1884), 67-68. In view of this problem, Frege famously decided to identify cardinal numbers with extensions, which move rendered his system of arithmetic inconsistent, entailing Russell's paradox. Frege (1884), 79-80. For a detailed discussion and treatment of the *Caesar Problem* (in response to Dummett's criticism of the neo-Fregean account in Dummett (1991)), see Hale and Wright (2001b).

A similar principle, the so-called *Direction Equivalence*, is supposed to underlie our acquisition of the geometrical concept of abstract direction:

(DE) The direction of line *a* = the direction of line *b* ↔ lines *a* and *b* are parallel

As Hale and Wright observe, (HP) and (DE) are two instances of the same general *Abstraction Principle*:

(AP) $\forall\alpha\forall\beta\ (\Sigma(\alpha) = \Sigma(\beta) \leftrightarrow \alpha \approx \beta)$

where '≈' stands for an equivalence relation on entities of the type of α, β, and Σ is a function from entities of that type to objects.[24] According to Hale and Wright, (AP) provides us with a general tool to formulate thoughts and acquire knowledge of platonic

[24] Hale and Wright (2002), 118. One may argue that the adoption of this principle in its full generality is certainly not admissible, since some of its instances, most notably Frege's Principle of Extensional Abstraction (Basic Law V of *Grundgesetze*), lead to a contradiction. In view of this problem, Dummett (1991) warned that in absence of an explicit specification of what distinguishes the harmless instances of (AP) from the harmful ones, the neo-Fregean reliance on (HP) and (DE) is cannot be justified. For Hale and Wright's response to this "bad company argument", see Hale (1994), and Wright (1998). For an in depth discussion of abstraction principles in general, see Fine (2002).

objects without having an epistemic contact with them. On the one hand, with the help of (AP)'s instances, we can implicitly define concepts referring to abstract objects, and with the ensuing conceptual apparatus we can form beliefs about platonic realms. On the other hand, the above derivation of mathematical concepts is a key to a no-contact theorist explanation of mathematical knowledge as well. According to this conception, mathematical knowledge is grounded on knowledge of the identity conditions of mathematical objects, which merely requires that we can discern whether the truth conditions of identity claims about those objects (i.e. instances of '$\Sigma(\alpha) = \Sigma(\beta)$' in (AP)) obtain. These conditions, however, are stipulated to be identical, by the instances of (AP), with the truth conditions of some epistemologically unproblematic claims of equivalence relations (i.e. instances of '$\alpha \approx \beta$' in (AP)). Clearly, if our knowledge of the instances of (AP) and the relevant equivalence relations is unproblematic, then our knowledge of the corresponding abstract states of affairs cannot be problematic either.[25]

As it might be expected from what has been said so far, there are a couple of things on which I agree with Hale and Wright. First, I agree that we can think of and speak about platonic entities, and that the intended referents of our pure mathematical beliefs are abstract in the relevant sense of the term. Second, I also agree that we can acquire knowledge of such abstract entities, and that mathematics is the collection of such knowledge. Finally, I agree even that our knowledge of

[25] As Hale and Wright puts it: "So long as we can ascertain that lines are parallel, or that concepts [in the Fregean sense, Zs. N.] are one-one correspondent, there need be no *further* problem about our knowledge of certain basic kinds of truths about directions and numbers, for all their abstractness. For provided that the concepts of *direction* and *number* can be implicitly defined by Fregean abstraction, we can know statements of direction- and numerical-identity to be true just by knowing the truth of the appropriate statements of parallelism among lines and one-one correspondence among concepts. We can do so for the unremarkable reason that the truth-conditions of the former are fixed by stipulation to coincide with those of the latter." Hale and Wright (2002), 119.

mathematical and other abstract entities does not require an epistemic contact between us and the constituents of a platonic realm. What I do not agree with is that the previous commitments entail the endorsement of a platonist construal of mathematical truth (or truths about abstract domains in general) combined with a no-contact theorist response to Benacerraf's epistemological challenge to this semantical position.

There are at least two reasons for querying the correctness of the neo-Fregean transition from the former premises to the latter conclusion. First, the subject matter of a thought or sentence can be abstract without being platonic in character. Directions, for instance, as universal properties can characterise fictive and real spatiotemporal objects as well.[26] An account of our ability to refer to and acquire knowledge of abstract entities may therefore amount to a platonist response to Benacerraf's challenge only if the subject matter of this knowledge is abstract in the required sense of the term. In contrast to the numerical terms of pure mathematics, however, the denoting expressions appearing on the left hand side of the instances of Frege's Abstraction Principle do not necessarily stand for abstract entities in the required sense of the term (i.e. they do not necessarily stand for strictly non-spatiotemporal entities). The expression 'the number of apples in front of me', for instance, primarily stands for a property of a group of objects in the spatiotemporal world, rather than for an object of a platonic realm.[27] Other

[26] In the fifth section of chapter 1, I mentioned that a proper understanding of the epistemological problem with platonist theories of truth requires the careful disambiguation of our notion of abstractness, and separated the abstractness of spatiotemporal properties from the abstractness of entities that are supposed to exist in a platonic realm.

[27] In fact, the expression can be used to refer to both a numerical property of a group of objects in the spatiotemporal world and a mathematical object outside space and time. Due to this ambiguity, the sentence 'the number of apples on this table is the same as the number of spoons' can be interpreted also in at least two ways. On the first reading, it expresses a synthetic proposition about the numerical properties of the apples and the spoons on the table, while on the second, it expresses an analytic proposition about some mathematical objects outside space and time. On the account of semantic content advocated in this

denoting phrases, such as the expression 'the number of primes between 70 and 80', refer to a property of non-spatiotemporal entities, but the contextual definitions provided by Hume's Principle in these examples presuppose that we already acquired some concepts and knowledge of a non-spatiotemporal domain. Putting it briefly, what the neo-Fregean account seems to explain is how we can develop new concepts of certain fields from some earlier acquired ones of the very same fields. What it fails to explain is how we develop our notions of entities that cannot appear in space and time in the first place, maybe relying on our notions acquired earlier of entities appearing in space and time.[28]

work, the ambiguity is a consequence of those referential intentions that lie behind the two sorts of applications. While in the first case we intend to make an empirical claim about the contingent numerical properties of two groups of objects in space and time, in the second scenario we intend to advance an *a priori* statement of the necessary self-identity of a mathematical object within the non-spatiotemporal domain of pure mathematics. (Thanks to András Simonyi for reminding me of this ambiguity.)

[28] I admit that by stipulating that Σ is a function from entities of the type of α, β etc. to objects (i.e. individuals) Hale and Wright ensure that the concepts resulting from Fregean abstraction stand for entities of the right kind. Note, however, that the stipulation entails that the expressions 'the number of apples in front of me' and 'the direction I am actually looking at' are about strictly non-spatiotemporal mathematical objects, rather than about some properties that may characterise entities in space and time. As the observation made in the previous footnote may show, the problem with these readings is not that they are inadequate in the light of our actual linguistic and cognitive practice. By the application of these expressions we can definitely talk about the strictly non-spatiotemporal individuals of pure mathematics as well. The problem with the stipulation is rather that it does not provide us with those readings which explain the plausibility of the neo-Fregean claim that the truth conditions of the expressions on the two sides of the sentential connective '↔' in (AP) are identical. The reason for which we accept this claim is, to adopt Frege's own formulation, invoked by Hale and Wright as well, that we believe that the expressions on the two sides of the instances of (AP) "carve up" the same content in different ways. Frege (1884), 75, Hale and Wright (2002), 117. What this metaphor suggests is that the expressions on the two sides of the instances of (AP) have the same truth conditions, because they state the obtaining of the same conditions in slightly different ways. Supposing that the unproblematic expressions on the right-hand side of the instances of (AP) are not about the strictly non-spatiotemporal domain of pure mathematics, we can derive that the representations on the left-hand side are not meant to be about that

So, the first reason for querying the neo-Fregean view that we can acquire concepts and knowledge of platonic entities by means of Fregean abstraction is that the notions defined contextually by the instances of Frege's Abstraction Principle are generated by a sort of abstraction that does not guarantee the atemporal character of the resulting intended referents.

Second, even if we grant, as I think we should, that by a certain sort of abstraction we can develop concepts of strictly non-spatiotemporal entities as well, an account of knowledge of such entities can amount to a platonist response to Benacerraf's challenge only if the truth conditions of the thoughts composed of these concepts are to be construed in a realist and referentialist way. Now, the question that I propose to consider is whether the neo-Fregean epistemology advocated by Hale and Wright is compatible with such a realist and referentialist construal of mathematical truth. In what follows, I shall show that the proper response to this question is also negative.

Let us forget for a moment the previous observation that the denoting expressions on the left hand side of the instances of Frege's Abstraction Principle do not necessarily stand for abstract entities in the required sense of the term. According to Hale and Wright, our knowledge of non-spatiotemporal objects, such as numbers and (platonic) directions, must be unproblematic, since this knowledge is logically derivable from those identity claims which stand on the left hand side of the instances of Frege's

domain either. If we take the intuitive ground of the neo-Fregean claim about the identity of the relevant truth conditions seriously, then we cannot suppose that Σ in (AP) is a function from entities of the type of α, β etc. to individuals within a strictly non-spatiotemporal domain. Putting it briefly, the way in which Hale and Wright invoke (AP) in their reasoning against their anti-platonist opponents raises serious doubts about what they really mean when they stipulate that the concepts generated by Fregean abstraction denote *objects* of a certain kind. In chapter 7, I shall suggest that our concepts of strictly non-spatiotemporal entities are developed by a cognitive operation that might be formally characterised along the neo-Fregean lines, but will not support the related neo-Fregean response to Benacerraf's epistemological challenge to platonism about truth in the semantics of our paradigm *a priori* discourses about non-spatiotemporal domains.

Abstraction Principle, whose truth conditions in turn are identical with the truth conditions of those, epistemologically unproblematic, beliefs, which stand on the right hand side of the relevant instances of Frege's Abstraction Principle. Supposing that to be non-problematic, in the context of Benacerraf's epistemological challenge to platonism about mathematical truth, means being non-platonic in character, what the instances of Frege's principle suggest to us is, among others, that the obtaining truth conditions of our mathematical beliefs are not platonic in character. If this is true, however, then the referentialist construal of these conditions must be abandoned: if our mathematical beliefs are about strictly non-spatiotemporal entities, while their truth conditions are supposed to obtain in an epistemologically unproblematic (presumably spatiotemporal) realm, then the latter conditions cannot be construed in terms of the former subject matters. If they could, then, again, according to the neo-Fregean reasoning, the obtaining truth conditions of our beliefs about the relevant right-hand equivalences would be platonic as well, and our knowledge of the latter would be no less problematic from Benacerraf's perspective than our knowledge of other platonic objects and properties in general. So, the second reason for querying the neo-Fregean view that we can acquire knowledge of platonic entities by means of logic and Fregean abstraction is that a charitable interpretation of the neo-Fregean account is incompatible with a realist and referentialist (i.e. platonist) construal of mathematical truth.

Apparently, Hale and Wright do not realise this incompatibility. They believe that the (realistically construed) truth of a certain kind of belief (here, a belief about non-spatiotemporal objects and properties) implies the real existence of the intended subject matter. Consider the following summary of their position:

> [I]n order to establish an intelligible use for singular terms purporting reference to numbers, or other abstract objects – that is, objects which are not 'external' (located in space), and of which we can have no 'idea' or 'intuition', but which are, in Frege's view, nonetheless objective – it suffices merely to explain the truth-conditions of statements incorporating such terms. No precondition involving prior engagement with or attention to the referents of such terms is soundly imposed. Moreover if, under a suitable explanation of the truth-conditions of an appropriate range of such statements, suitable such statements are – or may warrantedly be claimed to be – true, then those of their ingredient terms which purport reference to numbers or other abstract objects will in fact refer – or may warrantedly be claimed to succeed in referring – to such objects; and the intelligent contemplation of such a statement will constitute thought directed upon the objects concerned.[29]

Is there a legitimate motivation behind the maintenance of the referentialist construal underlying these formulations? As a first reaction, one may argue that if Hale and Wright's conviction were true (i.e. the realistically construed truth of a certain belief indeed implied the existence of the intended subject matter of the belief), then we could not form such true beliefs about fictive entities. Creating a fiction consists of the stipulation of a number of facts in an invented (i.e. paradigmatically non-real) universe. So far as we are told, for instance, in the fictive universe of Little Red Riding Hood it is a stipulated fact that the girl has got a grandmother. The stipulation, as opposed to the girl and the grandmother, is a fact of the real world, which guarantees that the sentence 'Little Red Riding Hood has got a grandmother' is objectively true, independently of whether anyone ever recognises this circumstance.[30] So, we may assume that the above sentence

[29] Hale and Wright (2002), 115.

[30] One may, of course, start or continue a story and maybe even change the truth value of some claims within the narrative without thereby changing the intended subject matter of the applied component expressions. A claim during the actual creation of a fictive story has no realistically construable truth conditions. Its truth

expresses a (realistically) true belief about Little Red Riding Hood. Applying Hale and Wright's referentialist tenet to this case, one may argue we must conclude that our concept of Little Red Riding Hood "may warrantedly be claimed to succeed in referring" (i.e. that Little Red Riding Hood exists just as much as the abstract referents of our mathematical claims). This conclusion, however, would be obviously false, so Hale and Wright's conviction cannot be generally true either.

To this objection a neo-Fregean may answer that successful reference is not meant to imply real existence. It is merely meant to imply "intended sort of existence": real and abstract in the case of mathematical objects and properties, while fictive and mostly spatiotemporal in the case of fictive entities, such as Little Red Riding Hood. Clearly, there is an important difference between the semantic content of the contrasted types of beliefs. While our ideas of mathematical objects and properties do not exclude the metaphysically thick or "real" existence of these subject matters, our referential intentions in the course of thinking of a fictive domain imply that the intended subject matters "exist" only thinly, in the relevant fictive universe. Little Red Riding Hood cannot be part of the real world, mathematical objects and properties in principle can.[31]

The contrast is apparently correct. Contrary to the case of our thoughts of fictive entities, our referential intentions in pure logic and mathematics do not exclude the existence of the subject matter of our true beliefs. Allowing the existence of these subject matters, however, is not the same as guaranteeing or requiring

value is determined by the authors who stipulatively characterise the relevant fictive universe. Nevertheless, once the story acquires its canonic form, the stipulations made provide a factual base (in the actual spatiotemporal world) for the evaluation of any further claims about that universe.

[31] In case the neo-Fregean accepts that the truth conditions of our claims about fictive entities can be construed along realist lines, the current response would amount to the adoption of a non-referentialist account of truth in the semantics of discourses about fictive domains.

(for truth) their existence.[32] Hale and Wright's position, that the existence of mathematical knowledge of non-spatiotemporal entities presupposes the existence of these entities would be correct if our cognitive practice in pure mathematics involved the referentialist idea that our beliefs cannot be (realistically) true unless the state of affairs they are about obtain in the real world, rather than merely in a world invented and projected by human minds. Note, however, that this idea is not inherent in daily mathematical practice. It is merely the manifestation of a substantive metaphysical and semantical position in the philosophy of mathematics. We saw that this position has problematic implications in the epistemology of mathematics: its adoption undermines the possibility of mathematical knowledge. No-contact theorist accounts of knowledge of platonic domains, in particular, do not attribute a substantive explanatory role to the posited platonic objects and properties, and thus, as we saw it in the empiricist case, undermine the idea that we can have an epistemic ground for believing in the existence of these entities and the truths that this existence allegedly constitutes.33 Further, I argued that the platonist position is incompatible, under a charitable interpretation, with the suggested neo-Fregean account of mathematical knowledge and truth. Finally, we may add that the unrestricted approval of the referentialist principle, which sanctions the inference from (realistically construed) truth to the

[32] In chapter 7, I shall argue that no analytic truth requires the reality of the relevant intended referential domains. What the realist construal of these truths requires is merely the reality of the crucial link between the relevant conceptual constituents of the analytic claims under scrutiny.

[33] Here I am relying on the naturalistic conception of evidence advanced in chapter 2, according to which theories can be justified by reference to a certain pool of evidence if and only if, and because, by reference to their obtaining truth conditions we can explain the actual occurrence of this evidential ground. Since by referring to platonic entities we cannot explain the actual occurrence of anything in the spatiotemporal world, nothing in this world can be legitimately taken as an epistemic ground for adopting or rejecting a belief that has platonic truth conditions.

existence of subject matter, opens the gate for stipulating or defining entities into existence.[34]

So, why should a neo-Fregean insist on this referentialist (and thus platonist construal) of mathematical truth? If in the semantics of our discourses about abstract domains we abandon that referentialist conception, and adopt an epistemologically unproblematic naturalist construal of the truth conditions of our beliefs, then we can maintain the idea of objective knowledge of abstract entities while remaining agnostic about the actual existence of the intended abstract subject matters. The resulting position could embrace platonism about referents or intended subject matters, but it would imply anti-platonism about the relevant truths.[35] Anti-platonism does not imply anti-realism

[34] The same charge appears in Field (1984), Dummett (1991) and more recently in Potter and Smiley (2001). Hale and Wright's response to the charge is that the instances of Frege's Abstraction Principle do not stipulate into existence any objects. What they do create is merely a certain sortal concept. As Hale observes, "[w]hether that concept is instantiated is, always, a matter settled by the truth or falsity – which is of course not itself a matter for stipulation – of instances of its [the principle's - Zs. N.] right hand side. […] All that is stipulated is the truth of a (universally quantified) biconditional". Hale (2001), 347, Hale and Wright (2002), 121. As far as I can see, however, this answer misses the point. The charge holds even if the existence of the intended abstract referents of the numerical terms of the identity statements on the left hand side of the instances of Frege's Abstraction Principle is stated conditionally, if and only if the equivalence statements on the right hand side of those instances are true. The concern is, obviously, why should a non-spatiotemporal entity exist if and only if a certain equivalence among certain spatiotemporal entities, such as lines and extensions of concepts, actually obtains? Without Frege's Abstraction Principle, and the referentialist construal of truth sanctioning the neo-Fregean inference from truth to the existence of subject matter, no such existence-claim could be derived.

[35] The separation of platonism about mathematical objects and properties from platonism about mathematical truth may also help Hale and Wright understanding the sense in which some platonists have found neo-Fregean platonism "insufficiently robust" in the philosophy of mathematics. As they say, "[t]he other accusation, that abstractionist platonism falls so far short of the genuine article as to be unworthy of the name 'platonism' at all, is harder to come to grips with, partly because it is hard to find a clear, articulate and non-metaphorical account of what 'genuine' platonism is supposed to involve and partly because any of several distinct things may lie behind it". Hale and Wright (2002), 121. According to my understanding, the robust platonist's problem with the neo-Fregean account is

however: the truth conditions of our beliefs about abstract domains can acquire a realist construal in a naturalistic framework as well. A naturalist conception of mathematical truth would, indeed, explain one of the core intuitions behind the neo-Fregean account as well: it would illuminate why our knowledge of abstract entities (in both of the earlier contrasted senses of the term) is no more problematic from Benacerraf's perspective than our knowledge of some spatiotemporal entities, such as those referred to on the right hand side of the instances of Frege's Abstraction Principle.

Summing up, although Hale and Wright present their case as a platonist, no-contact theorist response to Benacerraf's challenge in the philosophy of mathematics, their neo-Fregean account does not really explain how we can develop concepts of non-spatiotemporal entities in the first place, and (on a charitable reading) it cannot be reconciled with a platonist (i.e. realist and referentialist) construal of mathematical truth either. The adoption of a non-referentialist, naturalist construal of truth in the semantics of discourses about non-spatiotemporal domains, moreover, fits well with the neo-Fregean intuition that our knowledge of abstract objects and properties is no more problematic than our knowledge of the obtaining of some spatiotemporal states of affairs, but only in so far as it leaves open the possibility of developing a (naturalistic) contact theorist account of the relevant types of knowledge.

Three other platonist accounts of mathematical knowledge, which seem to provide a no-contact theorist answer to Benacerraf's challenge, involve a straight commitment to the platonist construal of the truth conditions, as opposed to merely the intended referents, of mathematical beliefs. According to the first, advocated, among others, by Jerrold Katz and David Lewis, our knowledge of abstract mathematical facts requires no epistemic contact with the relevant obtaining truth conditions,

that it does not support the idea that there are abstract objects and properties in the world and that our knowledge about these entities involves knowledge of their actual existence as well.

because this obtaining is necessary in character.[36] The reason for which we need to contact the intended referents of our empirical knowledge claims is that these entities could have been different. If a certain state of affairs necessarily obtains, then there is no point to check whether it in fact does so.

One may wonder, however, why the necessary character of mathematical facts would guarantee that our actual mathematical claims tend to be true, rather than false. The alleged fact that mathematical objects exist and are necessarily as they are, in itself, does not seem to guarantee or explain the reliability of our actual mathematical belief formation. If there is no information-conveying link between the obtaining truth conditions of our mathematical beliefs, on the one hand, and our actual evidence for these beliefs, on the other, then it seems that we have still no reason to suppose that our mathematical claims correctly represent the facts of the intended abstract realm. Consequently, the appeal to the necessary character of mathematical and other abstract facts does not seem to resolve the epistemological problem with the platonist construal of the relevant truths either.[37]

According to the second account, formulated, among others, by Michael Resnik and Stewart Shapiro, mathematical knowledge is knowledge of certain structures, which can be exemplified by various systems of abstract or spatiotemporal objects, and we can acquire this knowledge by constructing consistent axiom systems, because such systems provide implicit definitions of the structures characterised.[38] As Shapiro puts it, in relation to arithmetic:

[36] Katz (1981), Katz (1995), Lewis (1986).

[37] Field (1989), 238, Balaguer (1998), 41-45.

[38] Resnik (1997), Shapiro (1997), Shapiro (2000). Note that *in re* versions of structuralism with a physicalist background ontology (i.e. versions maintaining that the intended subject matters of mathematics are structures to be exemplified by systems of spatiotemporal entities) could *prima facie* provide a referentialist response to Benacerraf's dilemma, since referentialism, according to these versions, would imply naturalism or physicalism, rather than platonism about mathematical truth. As it has been mentioned in chapter 3, however, the idea that the subject matter of pure mathematics (and other discourses about abstract domains) is spatiotemporal in character is incompatible with the intentionalist construal of reference briefly put forward in section 4 of chapter 1. Furthermore, these versions of structuralism have troubles with the explanation of the apriority and necessity of the relevant truths, and the infinity of the relevant intended domains. Other structuralists, such as Benacerraf (1965), Putnam (1967) and Hellman (1989), embrace realism about mathematical truth without presupposing the existence of (non-spatiotemporal) mathematical entities. These positions may be in line with the non-referentialist framework advocated here, but they must be adjusted by a suitably articulated account of the factual basis of mathematical truths in the actual world.

> The structuralist vigorously rejects any sort of ontological independence among the natural numbers. The essence of a natural number is its *relations* to other natural numbers. The subject-matter of arithmetic is a single abstract structure, the pattern common to any infinite collection of objects that has a successor relation, a unique initial object, and satisfies the induction principle.[39]

The core constituents of this structuralist response to Benacerraf's challenge correspond to those of Hale and Wright's neo-Fregean account: consistent implicit definitions can provide us with both concepts and (*a priori*) knowledge of abstract entities (in this case, structures of various systems of objects) that exist in the world independently of our actual thoughts and knowledge of them; and the development of such definitions is something that we can explain without invoking an epistemic contact between our minds and the postulated abstract (non-spatiotemporal) entities.

The main problem with these referentialist (*ante rem* or *in re*) versions of structuralism is that, similarly to Hale and Wright's neo-Fregean platonism, they cannot account for our knowledge of the existence of the intended abstract subject matters, and thus for our knowledge of mathematical truths either. Knowledge of consistent implicit definitions cannot amount to, or imply, any knowledge of the existence of what the concepts defined purport

[39] Shapiro (2000), 258.

to stand for, whether the purported entities are particular objects with intrinsic properties, or merely positions in an abstract structure that lack any individuating properties beyond their stipulated relations to other positions in the structure. The apriorist epistemology underlying these accounts can work only if the (realistically understood) truth of the relevant beliefs does not require the existence of the intended abstract subject matters, or in other terms, if truth in the semantics of these discourses is construed along non-referentialist lines. Again, such a construal would support the suggested apriorist account of knowledge by allowing an epistemologically unproblematic naturalist account of the truth conditions of the relevant beliefs. This would let us maintain the idea of objective knowledge about the intended non-spatiotemporal entities, and remain agnostic about the actual existence of these subject matters.[40]

Beyond failing as accounts of mathematical truth, structuralist forms of referentialism in philosophy of mathematics can be queried as construals of the subject matter of mathematics as well. If reference were to be understood in pure model-

[40] The "modal-structural interpretation" put forward in Hellman (1989) may be an example of this non-referentialist strategy in the philosophy of mathematics. According to this account, mathematics is about structures that could but actually may not be exemplified by particular systems of objects in the world. The standard objection to this account is that since mathematical truths are necessary in character, if some mathematical entities may exist, then they must exist, even in the actual world. Resnik (1992), 117, Shapiro (2000), 274, Isaacson (forthcoming). Note, however, that in a non-referentialist framework, the necessity of mathematical truths need not imply the necessity of the (metaphysically thick) obtaining of mathematical states of affairs. If the truth conditions of mathematics are not referential in character, then the idea that mathematics is about possibly existing entities that may not exist in the actual world becomes compatible with the view that mathematics consists of necessary truths. Of course, the objection remains valid, even in a non-referentialist framework, if our referential intentions in mathematics imply that the intended abstract subject matters in question must not exist unless they exist in the actual world. But why should one who has already abandoned referentialism about mathematical truth maintain anything about the mode of existence of mathematical objects and their properties? Such stipulations would be fully arbitrary in so far as they would no longer contribute to the explanation of any phenomenon in the actual world.

theoretic terms (i.e. in terms of "satisfaction"), then we would have, indeed, good reasons for construing mathematical objects and properties along the structuralist lines: a mathematical theory can be satisfied by various systems of objects (arithmetic, for instance, by both a system of Zermelo and a system of von Neumann ordinals), which could equally serve as the intended subject matter of the (syntactically understood) theory in question.[41] In the previous chapter, however, I argued that Putnam's permutation argument successfully demonstrates that our notion of reference or subject matter cannot be reduced to the pure model-theoretic notion of satisfaction. The fact that Peano's arithmetic is satisfied by various pluralities does not mean that our numerical concepts are about the structurally identified members of these pluralities, or about the positions themselves occupied by these members. Our numerical concepts are about numbers, which are meant to be different from both the set-theoretic entities invoked by Zermelo and von Neumann, and the structural positions that these entities occupy in the systems they are meant to constitute. The difference is manifest in our cognitive and linguistic practice as well: on the one hand, we do not think that the natural number 2 is identical with the set {{0}} in the system of Zermelo ordinals or with the set {0,{0}} in the system of von Neumann ordinals; on the other hand, we also deny that it is identical with a structural position that can be filled by various entities. In view of these cognitive and linguistic facts, it is hard to see what could justify the replacement of the traditional "object-platonist" construal of the subject matter of mathematics (and other discourses about non-spatiotemporal domains) with one of the structuralist alternatives.[42]

[41] The connection between this model-theoretic notion of reference and the appeal of a structuralist construal of mathematical entities is manifest in almost every representative paper of the structuralist tradition in the philosophy of mathematics.

[42] Structuralist construals of intended referents seem to be more adequate when we turn to our concepts of roles within a game or a system of institutions. The subject matter of our concepts of the white king's bishop or the President of the United States, for instance, is meant to be identical, in each context of application,

The third no-contact theorist account of mathematical knowledge that clearly endorses platonism about mathematical truth relies on a specific conception of the platonic realm. Mark Balaguer has baptised this view *plenitudinous* or *full-blooded platonism* (*FBP* for short), and it consists in the idea that all logically possible mathematical objects exist.[43] To see how this version of platonism is meant to supply an answer to Benacerraf's epistemological challenge, let me recall Balaguer's main objection to the "implicit definitionist" accounts of mathematical knowledge and reference presented above. The essential problem, on Balaguer's view, with Katz's and Lewis's necessity-based or Resnik's and Shapiro's structuralist replies is that none of them can explain how we could know that our implicitly defined mathematical concepts or internally consistent mathematical descriptions pick out an object in the alleged mathematical realm. As Balaguer puts it:

> Platonists can claim that the term '4' is just an abbreviation for the term 'successor of 3', but what anti-platonists will demand is an explanation of how we could know that there is an object in the mathematical realm that *answers* to this description. In other words: it's very easy to give definitions of mathematical singular terms like '4', but it's not so easy to see how we could know which terms and definitions actually refer to something.[44]

Or a few pages later:

---

either with a particular entity that occupies a certain role, respectively, in a certain game or in a certain system of institutions, or with the structural position itself occupied by the previous objects or individuals.

[43] Balaguer (1998). We must add that Balaguer's purpose in his book is not so much to defend platonism about mathematical truth and reference, but instead to show that in the philosophy of mathematics there are no conclusive arguments against either of the opposite doctrines of platonism and anti-platonism.

[44] Balaguer (1998), 42-43.

> Putting this response into the lingo that Resnik and Shapiro use, the problem is that prima facie, it seems that platonists cannot claim that we can acquire knowledge of abstract mathematical structures by merely formulating axiom systems that implicitly define such structures, because in making this claim, nothing is said about how we can know which of the various axiom systems that we might formulate actually pick out structures that exist in the mathematical realm.[45]

As Balaguer observes, however, if *FBP* is true, then there is a trivial platonist reply to this objection:

> For if all the mathematical objects that possibly *could* exist actually *do* exist, as FBP dictates, then all (consistent) mathematical descriptions and singular terms will refer, and any (consistent) representation of a mathematical object that someone could construct will be an *accurate* representation of an actually existing mathematical object.[46]

One may raise various objections to *FBP* and this reply to Benacerraf's challenge. Many of them are properly discussed and answered by Balaguer.[47] The most important problem, however, that he does not seem to appreciate is that what advocates of *FBP* can at most explain is how a reliable method of establishing logical relations (e.g. the consistency of alternative axiom systems) could also count as a reliable method of discovering platonic facts, *in case* the states of affairs in question obtain, indeed, in the real world. Still, they cannot tell us anything about how we could know that the latter condition holds after all. According to Balaguer, this is not a serious failure though, since:

---

[45] Balaguer (1998), 45.

[46] Balaguer (1998), 43.

[47] Balaguer (1998), ch. 3, esp. 58-69. For a sympathetic critical review of Balaguer's ideas, see Colyvan and Zalta (1999).

> [a]nti-platonists are not demanding here an account of how human beings could know that there exist any mathematical objects at all. That, I think, would be an illegitimate skeptical demand. […] All we can demand from platonists is an account of how human beings could know the *nature* of mathematical objects, *given* that such objects exist.[48]

I think that Balaguer is right when he distinguishes the task of showing how human beings could gain knowledge, or acquire reliable beliefs, about the nature of mathematical objects, given that such objects exist, from that of providing reasons for beliefs in the very existence of these entities. I also understand that there is a weak reading of Benacerraf's challenge, according to which the problem with platonism about truth is that it undermines the explanation of how our spatiotemporal belief-forming mechanisms *could* provide us with reliable information of platonic entities *if* the latter existed as platonists suppose. Finally, I grant that Balaguer's *FBP* amounts to an acceptable response to this reading of the challenge: it shows that under suitable circumstances our actual belief-forming mechanisms could provide us with largely true beliefs about platonic states of affairs, supposing that those indeed obtain. On the other hand, I believe that the intended reading of Benacerraf's challenge is stronger than the one suggested above. On this reading, what Benacerraf queries is the platonist's ability to explain how our spatiotemporal belief-forming mechanisms could provide us with reliable beliefs about what actually obtains in a platonic realm, including the issue of whether the truth conditions of *FBP* itself obtain.

In line with this reading, of course, I also query Balaguer's claim that the call for an account of how human beings could know that there are non-spatiotemporal entities in the metaphysically thick sense of the term is an illegitimate sceptical demand. In chapter 5, I argued that in the case of our beliefs

[48] Balaguer (1998), 43.

about the spatiotemporal world we can actually put forward an account of how human beings can know of the (thick) obtaining of the truth conditions or the existence of the subject matter of their beliefs. Further, earlier in this chapter, I argued also that in the case of our beliefs about fictive entities our referential intentions undermine the conceivability of such an account if we suppose that the truth conditions of these beliefs are referential in character, but the demand can be met if we adopt a non-referentialist, naturalistic construal of fictive truths. So, why should we abandon this demand in the philosophy of our discourses about abstract domains?

Balaguer is right: it would be a position on a par with radical scepticism if we did not allow the advocates of *FBP* (or other forms of platonism about mathematical truth) to assume at the beginning of their explanation that the relevant abstract entities exist. The charge, however, that anti-platonists raise against their opponents is not that they make initial assumptions of the intended non-spatiotemporal domains, but instead that the assumptions in question do not help us to understand why we should believe in their correctness at the end of the day.[49]

In response to this charge, advocates of *FBP* may invoke the explanatory virtues of platonism about the paradigms of *a priori* truth reviewed in section 1. In particular, they may argue that those virtues provide us with sufficient epistemic ground for believing in the correctness of a platonist's initial metaphysical assumptions. This reasoning, however, presupposes that the phenomena explained by reference to the suggested platonic facts cannot be properly explained otherwise (i.e. without invoking the existence of the relevant platonic objects and properties). In the following chapter, I shall show that this presupposition is false: by adopting a naturalistic, non-referentialist construal of the truth conditions of our beliefs about non-spatiotemporal domains, we

[49] As I argued in chapter 5, our standard assumption that there is an external, spatiotemporal world beyond the veil of our narrowly understood experience is one in terms of which we can explain why that experience can be reasonably regarded as a sign of the correctness of this assumption.

can explain every phenomenon that platonists intend to explain by invoking a non-spatiotemporal ontology. In view of these results, we can conclude that the metaphysical assumptions of an advocate of *FBP* cannot be based on the suggested explanatory considerations.

With this conclusion we have completed the survey of the most influential apriorist no-contact epistemologies in the philosophy of mathematics. As we saw, similarly to the empiricist form of this general strategy, none of these accounts can suitably explain how we could acquire knowledge of what obtains in a platonic realm. In a referentialist semantical framework, this failure undermines the accounts' adequacy as a theory of knowledge or reliable belief formation of non-spatiotemporal domains. The general problem with these attempts is that by denying the existence of an information-conveying link between our minds, on the one hand, and the (referentially construed) truth conditions of our beliefs about abstract domains, on the other, they deprive the latter conditions from any explanatory power *vis-à-vis* our mental life, so that nothing in our mind can be reasonably regarded as a distinctive evidence of the obtaining of the relevant truths.[50]

In view of this result, we shall conclude that Benacerraf's epistemological challenge cannot be properly answered by reference to a no-contact epistemology either. Together with our previous verdict on platonist contact epistemologies, this conclusion implies that the advocates of a referentialist (i.e. platonist) version of realism about truth in the semantics of our paradigm *a priori* discourses cannot account for the possibility of knowledge or reliable belief formation of the intended abstract

[50] In the conclusion of his book Balaguer also recognises this point. Balaguer (1998), 157. Nevertheless, instead of adopting a substantive realist and non-referentialist construal of mathematical truth, he rather endorses the somewhat innovative position that "*we* could never settle the dispute between platonists and anti-platonists", since "there is *no fact of the matter* as to whether platonism or anti-platonism is true, that is, whether there exist any abstract objects". Balaguer (1998), 152.

domains (i.e. their conception does not meet the fourth adequacy condition set for a theory of *a priori* truth in chapter 2).

Summary

In this chapter, I examined the prospects of the most influential platonist replies to Benacerraf's updated and generalised dilemma presented in chapter 3. The common feature of these responses is that they attempt to solve Benacerraf's problem by querying one of the epistemological premises of his case.

In section 1, I provided a brief overview of the major explanatory virtues and vices of platonism about truth in the semantics of our discourses about abstract domains, and argued that in absence of a viable account of knowledge acquisition or reliable belief formation about platonic realms one may query the adequacy of the advanced platonist explanations of apriority, intersubjectivity and observable convergence as well, which considerably reduces the explanatory power of the platonist construal.

In section 2, I turned to the discussion of the major platonist epistemologies. Following Mark Balaguer's useful distinction and terminology, I divided these accounts into two mutually exclusive and jointly exhaustive categories, and called them contact theories and no-contact theories of knowledge of platonic domains, respectively. First, I examined the contact theorist responses to Benacerraf's epistemological challenge, and argued that they are equally inadequate, since the explanation they provide of our knowledge or reliable belief formation about platonic domains is *ad hoc* and exotic in character. On the one hand, the adoption of such an epistemology would discourage any further inquiry into the nature of this type of knowledge acquisition, while on the other, it would open the door for parallel stipulations in the case of any other knowledge claims, no matter how the beliefs in question were causally produced in the subjects' minds. After this, I turned to the second group of platonist epistemologies, which query that our knowledge of

abstract domains requires some information-conveying contact between our minds and the obtaining intended abstract truth conditions. My primary objection to these theories was that by denying the existence of the above contact, they deprive the suggested truth conditions from their explanatory significance *vis-à-vis* the actual constatations of human minds, so that nothing that we are aware of remains there to be reasonably regarded as a distinctive evidence of the obtaining of the relevant abstract truths.

It is important to emphasise that we found nothing objectionable in the platonist construal of the subject matter of our paradigm *a priori* beliefs about abstract domains. As we observed, our referential intentions in pure logic and mathematics allow for such a construal. According to these intentions, numbers and propositions are non-spatiotemporal entities, which may exist in the actual world, and if they exist there, then they exist in a platonic realm. What most platonists do not seem to realise is that platonism about subject matter does not imply platonism about truth. In particular, they fail to see that the conditions whose obtaining we take to be necessary and sufficient for certain logical and mathematical truths, and whose obtaining in the world we are supposed to know when we possess logical and mathematical knowledge, are not necessarily those that these beliefs purport to be about. Once we abandon the received referentialist construal of truth in the semantics of discourses about abstract domains, our platonism about the intended non-spatiotemporal referents will no longer stand in the way of a naturalist construal of the relevant truths, and a corresponding causal, contact theorist account of *a priori* knowledge or belief formation.

In chapter 3, I argued that there are four *prima facie* admissible theoretical options for a referentialist to escape Benacerraf's original or modified and generalised challenge in the philosophy of those discourses, in which we are supposed to acquire knowledge or reliable beliefs about causally inert domains. In the last three chapters, I examined these options, and showed that none of them can save the adequacy of standard

referentialism as a universal conception of truth. The two semantical responses (deflationism and anti-realism about truth) prove to be inadequate, since they cannot account for the objectivity of our knowledge claims, while the two epistemological responses (non-causal contact theories and no-contact theories of knowledge of realistically construed abstract domains) fail, because the account they provide is either *ad hoc* and exotic (non-causal contact theories) or insufficient (no-contact theories). In view of these results, we can conclude that Benacerraf's updated and generalised dilemma cannot be answered in a referentialist semantical framework.

With the fall of the above referentialist responses, there remains only one theoretical option for us to answer Benacerraf's dilemma: adopting a non-referentialist construal of truth in the semantics of our discourses about causally inert domains. In a non-referentialist framework, we may endorse a naturalistic conception of truth conditions while maintaining the received non-revisionist (i.e. platonist) construal of the intended subject matters. The resulting conception would be realist about truth, which means that it could explain the objectivity of this semantic property within these problematic discourses as well. On the other hand, by the naturalistic construal of the relevant truth conditions it would also support a causal, contact theorist account of how we can acquire knowledge or reliable beliefs about causally inert domains. Putting it briefly, a suitable non-referentialist response to Benacerraf's dilemma seems to satisfy those adequacy conditions that its referentialist alternatives failed to satisfy. The main question, of course, is whether the advocates of this response can also account for the other major *explananda* advanced in chapter 2. My central claim is that the answer to this question is positive. And it is this that I intend to show in the last chapter of this work.

CHAPTER 7

The Non-Referentialist Alternative: A Representationist Construal of *A Priori* Truth

Introduction

In the previous three chapters, I argued that a proper construal of truth in the semantics of our purportedly *a priori* discourses about abstract domains must be realist but not referentialist (i.e. platonist) in character. The chief objection to the alternative, *prima facie* viable, non-realist and/or referentialist construals was that they cannot explain some relatively obvious features of our cognitive and linguistic practice in discourses about abstract domains.

A reaction to be expected from the advocates of the rejected deflationist, anti-realist or platonist positions would be to argue that the semantical framework proposed by their realist and non-referentialist opponents cannot meet the explanatory requirements set for an account of the relevant truths either. If this charge turned out to be true, then there would be no more reason for adopting the latter perspective, than any other of those examined before.[1]

In this chapter, I shall show that this charge is unsubstantiated. In particular, I shall argue that a specific naturalist version of non-referentialist realism about truth in the

---

[1] In chapter 2, I argued that an adequate theory must provide, either in itself or in conjunction with other theories, an explanation of all "observable" (i.e. commonly recognised) characteristics of its subject matter. What this methodological principle implies is that a theory cannot be regarded as superior over its alternatives unless it can account for all observable phenomena of its subject matter.

semantics of our purportedly *a priori* discourses about abstract domains satisfies all major adequacy conditions set for such a construal in chapter 2. I shall call the version in question a *representationist* account of the relevant truths, as its core tenet is that the conditions whose (thick) obtaining or absence determines the truth value of our claims about abstract domains obtain in the realm of representations within our head, rather than in the domain of the represented abstract states of affairs.

In section 1, I shall argue that, despite the wide consensus among present-day philosophers about the correctness of a general referentialist construal of truth, in view of our actual cognitive and linguistic practice, we have no reason to suppose that there is a conceptual link between our notion of truth conditions and our notion of subject matters or intended referents, which would make the idea of a non-referentialist construal of certain truths conceptually objectionable.

In section 2, I shall outline an ontologically naturalist account of how we can develop truth-apt representations about abstract domains with non-referential truth conditions. The account is meant to illuminate some details of the representationist construal advocated in this work, and demonstrate that the conception satisfies the third adequacy condition set for an account of truth in the semantics of discourses about abstract domains in chapter 2.

In section 3, I shall confront the representationist construal with the other explanatory challenges specified in chapter 2, and show that, in contrast with the referentialist accounts discussed earlier, this version of non-referentialism provides us not merely with a suitable response to Benacerraf's updated and generalised dilemma in the philosophy of discourses about abstract domains, but also with acceptable accounts of all those phenomena whose joint explanation we agreed to regard as a minimal condition of adequacy for a construal of the relevant truths.

Finally, in section 4, I shall briefly recall the original broader perspective of our investigation and suggest that the representationist conception advocated here qualifies as an adequate characterisation of the nature of *a priori* truth, and

therewith the nature of *a priori* knowledge and justification, in general.

1. The Consistency of the Non-Referentialist Alternative

Misgivings at the classical rationalist doctrines of *a priori* knowledge of abstract domains, according to which human beings can discover facts of platonic realms by the competent use of reason, resulted in various anti-realist or naturalist-reductionist reactions in the history of modern philosophy. The proponents of these reactions either queried the reality of the intended abstract subject matters and with them the realist construal of the relevant *a priori* truths, or adopted a revisionist, naturalistic construal of those domains and argued that our purportedly *a priori* claims are in fact empirical in character. On the other hand, as we saw in chapter 6, those who maintain that we can acquire knowledge about platonic entities take it for granted that this knowledge is knowledge of thickly obtaining platonic conditions. A common conviction of the advocates of these alternatives is that a construal of the truth conditions of a certain class of claims amounts to a construal of the corresponding subject matters and *vice versa*. Accordingly, we cannot believe in the reality of abstract truths without believing in the thick obtaining of the intended abstract conditions, and a platonist construal of the latter amounts to a platonist construal of the former.

Consider, however, the mathematical claim that there are exactly three primes between 70 and 80. On the one hand, it seems to be clear that the subject matter of this claim is not spatiotemporal in character. It says something about some numbers, and numbers do not exist in space and time. On the other hand, it seems also clear that the claim would be true even if no one ever believed that there are exactly three primes between 70 and 80. In other words, its truth conditions seem to obtain independently of what anyone would ever think of this particular issue. So, apparently, neither the naturalist-reductionist, nor the anti-realist positions can provide us with an intuitively

plausible semantics for this simple mathematical claim. Of course, platonists gladly approve this conclusion, since they maintain that the claim is true in virtue of the thick obtaining of the platonic condition that there are three prime numbers between 70 and 80. As we have seen in chapter 6, however, platonists about abstract truths have no suitable account of how we could know that the above condition actually obtains in the suggested platonic realm.

At this point, one may wonder whether the previous strategies exhaust the conceivable doctrinal alternatives in the semantics of discourses about abstract domains. The suggestion that I shall argue for in this section is that they do not. They do not, because the shared referentialist assumption underlying these strategies is not necessarily true. It would be so if our notion of truth conditions and our notion of subject matters or intended referents were related in a way that would guarantee the identity of the notions' intended referents. Our actual cognitive and linguistic practice, however, does not seem to support the idea of the obtaining of such a relation. Truth conditions are meant to be thinkable conditions whose obtaining is necessary and sufficient for the truth of a certain truth-apt representation. Further, they are meant to be those conditions whose obtaining (or absence) we must discover in order to determine the truth value of that representation. In contrast, subject matters are particular or general states of affairs that we can think of or speak about by entertaining a thought or uttering a sentence. The former conditions may actually coincide with the latter. But determining truth values and being occasionally detected by knowing minds is not the same thing as being thought or stated to obtain within a fictive or real domain. Accordingly, there is nothing inconsistent in the idea that the conditions whose obtaining we actually detect while acquiring a piece of knowledge may differ from the subject matter of the claim expressing this knowledge.

That the non-referentialist scenario is not merely a theoretical possibility can be illustrated by the case of our knowledge of fictive states of affairs. As has been observed in chapter 6, the subject matter of the claim that Little Red Riding Hood has a grandmother is clearly non-existent. Neither the girl

nor the old lady can appear in the actual world.[2] If the truth conditions of this claim were identical with the intended fictive state of affairs, then it would be hard to understand how these conditions could obtain and how we could detect their obtaining by our cognitive faculties in the actual world. Apparently, if we want to maintain that there is something whose actual obtaining (or absence) determines the truth value of the above claim and is reliably detected by knowing minds, then we must abandon the idea that this condition is identical with the intended fictive subject matter.

If truth conditions need not be identical with subject matters or intended referents, then the opponents of classical rationalism can adopt a non-referentialist, naturalist version of realism about truth in the semantics of discourses about abstract domains without subscribing to a revisionist, naturalistic construal of the apparently abstract subject matters of these claims. They can either deny or become agnostic about the existence of platonic entities and still provide a naturalist account of our purportedly *a priori* knowledge of the intended abstract domains. In the following two sections, I shall show that the non-referentialist account under scrutiny is not merely a consistent but also a well-motivated conception of truth, at least in the semantics of our purportedly *a priori* discourses about abstract domains.

2. The Emergence of Non-Referential Truth Conditions

Having abandoned the standard referentialist construal of truth in the case of our paradigm *a priori* claims about abstract domains, one may wonder what fixes the semantic relations of such representations to their non-referential truth conditions, on the one hand, and their purportedly abstract intended referents, on the other. The development of a suitable response to these

[2] As has been emphasised in chapter 6, contrary to the case of pure logical or mathematical beliefs, our referential intentions in discourses about fictive domains guarantee the non-existence of the intended fictive subject matters.

questions about semantic content-determination is fundamental in so far as it sets the framework for subsequent accounts of other observable characteristics of the relevant sorts of representations, including those which have been listed as major *explananda* in chapter 2.

In the standard referentialist semantical framework, where truth conditions are understood in terms of intended referents, the two questions formulated above coincide: explaining what fixes the semantic relations between our truth-apt mental and physical symbols, on the one hand, and their truth conditions, on the other, is nothing else than explaining what fixes the semantic relations between those symbols and their intended referents. Moreover, since the conditions constituting the relevant referential contents are entities that we are supposed to think of or speak about, the account being sought can invoke as an *explanans* the selective work of our conscious attention as well.

In possession of such explanatory resources, the referentialist story could run briefly as follows: first, we become acquainted with particular features of the world; second, our attention singles out from among these features the most striking or practically relevant ones; third, recollectable traces of these particulars are developed and kept in our memory; fourth, relying on our recollections and recognitional abilities, we identify some contrasts and similarities among the perceived particulars, and develop conceptual representations of properties as reoccurring types or universals; fifth, we observe various temporal continuities among the particular occurrences of these types; sixth, our attention singles out the most striking or practically relevant of these continuities; seventh, recollectable traces of these particular continuities are developed and kept in our memory; eighth, relying on our recollections and recognitional abilities, we identify some of these continuities as parts of single uninterruptedly existing wholes, and develop conceptual representations of individuals as possessors of these continuously existing features as essential characteristics; finally, in declarative contexts, we try to use the acquired concepts and their linguistic expressions in line with our actual evidence as to whether or not

their identified referential declarative use conditions actually obtain.[3]

Unfortunately, in a non-referentialist semantical framework, the previous account can at most serve as an explanation of how determinate semantic relations emerge between our mental and physical symbols and their intended referents. It can not explain the emergence of such relations between those symbols and their non-referential truth conditions. Worse, it appears that an account of the latter phenomenon cannot invoke the selective work of our conscious attention as a major factor in non-referential content-determination. If truth conditions are not understood in terms of intended referents, then they can hardly be singled out from a perceived or otherwise accessed domain by our conscious attention. So, those who reject the standard referentialist construal of truth in the semantics of our paradigm *a priori* discourses about abstract domains must develop an account of how our representations about these domains become associated with their non-referential truth conditions without assuming that we ever consciously attend to or think of these conditions in the course of content-determination. My primary aim in this section is to show how an advocate of non-referentialism can meet this explanatory challenge, and account for the emergence of determinate semantic relations between our representations about abstract domains and their arguably non-referential truth conditions.[4]

---

[3] The proper elaboration and confirmation of this account must, of course, emerge from a painstaking empirical inquiry into the nature of human cognition. The present outline is meant to be merely a highly abstract and rough characterisation of what scientists might once establish of the emergence of referential contents in the natural world.

[4] As we shall see, the account will involve an explanation of the emergence of non-referential truth conditions in the case of our analytic claims in general, including those whose intended referents are to be found in the spatiotemporal world.

General Theoretical Constraints

As a point of departure, let me briefly review the most important theoretical constraints that, in line with our earlier considerations, the envisaged account of the emergence of non-referential truth conditions must observe.

In the previous chapters I endorsed the commonly accepted principle of the compositionality of semantic content. According to this principle, the semantic content of a complex representation is determined by that of the representation's components and the mode of their composition. Applied to truth conditions, the principle declares that the truth conditions of a complex representation are determined by those of the representation's basic truth-apt components and the mode of their composition. The truth conditions of the latter entities, on the other hand, are supposed to be fixed by *some* appropriate semantic correlates of their semantically most basic (conceptual or subsentential) constituents and the mode of their composition. In the standard referentialist framework, the semantic correlates in question are, of course, the intended referents of these basic representations. In a non-referentialist semantical framework, however, where truth conditions are supposed to differ from intended referents, these correlates must also differ from the subject matters of the applied concepts or expressions.

In chapter 4, I argued that truth in general can be understood as correct declarative applicability in the context of the judgement or utterance under scrutiny. As we have seen, one major advantage of this construal is that in a compositionalist semantical framework it renders our theory of truth part of our general theory of correct (declarative) symbol-application. Adopting this theoretical framework, we can say that the semantic correlates of our concepts or subsentential expressions that contribute to the determination of the truth conditions of our truth-apt thoughts and sentences are what have been called the *declarative use conditions* of these atomic representations. They are conditions whose (thick) obtaining is individually necessary and jointly sufficient for the correct applicability of the relevant

concept or expression in a certain judgement or declarative sentence in the contexts of the latter's making or utterance. In the case of claims whose truth conditions can be construed in the standard referentialist way, the declarative use conditions of the key concepts and their linguistic expressions are to be identified with the intended referents of these representations. In the case of claims, on the other hand, whose truth conditions are better understood along the suggested non-referentialist lines, the declarative use conditions of the key concepts and their linguistic expressions must not be identified with the subject matter of these contentful entities. In general terms, we can say that the truth conditions of our truth-apt representations are determined by the correct declarative use conditions of their atomic conceptual or subsentential constituents and the mode of their composition.

Of course, this result, in itself, does not resolve the explanatory puzzle concerning the emergence of non-referential truth conditions. The observation that the declarative use conditions of our concepts and subsentential expressions together with their mode of composition determine the truth conditions of our truth-apt representations does not tell us anything about how those declarative use conditions become associated with the relevant atomic representations in those cases in which they are not referential in character. In the following section, I shall address and answer this fundamental question, but before doing so, let me briefly recall a few further characteristics that, on some explanatory considerations, we can reasonably assume hold of the relevant non-referential truth conditions.

First, as has been argued in chapters 4 and 5, a metaphysically neutral (deflationist) or anti-realist theory of semantic content cannot properly explain what makes the declarative use of our contentful mental and physical symbols in the actual world objectively correct or incorrect, or our truth-apt utterances objectively true or false. Since we supposed that in the absence of this explanation no theory of the subject can be adequate, we can assume that the non-referential declarative use conditions of our paradigm *a priori* representations admit of a

realist interpretation (i.e. they obtain, if they do, independently of anyone's actual thought or knowledge of this particular circumstance).

Second, as has been shown in chapter 6, a platonist construal of these conditions would undermine the possibility of any sensible explanation of how we can know of, or reliably detect, the obtaining or absence of these conditions, and therewith the declarative applicability of the relevant symbols. So, if we want to account for the possibility of knowledge of, or reliable belief formation about, a certain domain, then we must suppose also that the declarative use conditions of our representations about that domain obtain (or not) in the epistemically accessible spatiotemporal world.

Finally, our thoughts or sentences about abstract domains are the paradigms of those representations whose endorsement or rejection is supposed to be based on *a priori* considerations or evidence. Recalling the characterisation of apriority that we adopted in chapter 1, we can also say that our knowledge, or the justification of our judgements, of the obtaining or absence of the declarative use conditions of our representations about abstract domains is supposed to be independent of our experience (i.e. the deliverances of our perception of the external world and our introspection of our own bodily states). So, if we want to account for this specific feature of our representations about abstract domains, then we must suppose that the declarative use conditions of these symbols obtain (or not) in a very specific segment of the spatiotemporal world: it cannot be external to our body, since our knowledge of this part of the world is based on perceptual evidence; and it cannot be just any part of our own body, since our knowledge of most such parts is based on introspective evidence.[5] The only part of the natural world of

[5] One could claim that any first-personal knowledge of our own bodily states is by definition introspective in character. If we adopted this terminology, then the naturalistic construal of the declarative use conditions of our claims about abstract domains suggested here would render all such claims empirical, whose justification is based on experience. As I have emphasised in the first section of chapter 1, the account of *a priori* truth and knowledge advocated in this work is fully compatible with this kind of radical empiricism. The conflict between the two positions is arguably terminological. The question is, of course, whether the empiricist acknowledges the reality of an epistemologically significant substantive contrast between what is traditionally distinguished as the paradigms of *a priori* and the paradigms of *a posteriori* truth and knowledge. What I wish to defend here is that the contrast in question is real, and that (together with some further facts about the world) it can explain all observable characteristics of the distinguished entities (viz. the *a priori* and *a posteriori* pieces of truth, knowledge, justification and evidence).

which our first-personal knowledge is conventionally not taken to be introspective, and thus empirical, is the realm of our representations within our own head. Accordingly, what I intend to suggest here is that if we want to account for the apparent apriority of our knowledge of abstract domains, then we must suppose that the declarative use conditions of our representations about these domains are some specific relations among these symbols, which may or may not obtain in our head depending on how we developed them in the course of our cognitive engagement with our natural environment.

If the previous claims about the declarative use conditions of our mental and physical symbols with abstract intended referents are correct, then what we must explain in the following section is how these particular conditions in our head become associated as declarative use conditions with those symbols without ever being singled out by our conscious attention. The explanatory task in question will be accomplished in two major steps: first, I shall show how the suggested sorts of non-referential declarative use conditions appear and become associated with some of our representations in discourses about the spatiotemporal world; second, I shall show how these conditions become subsequently associated with our representations about abstract subject matters as well.

Analyticity: The Emergence of Non-Referential Truth Conditions in Discourses about Natural Domains

An important consequence of the suggested naturalistic account of the truth conditions of our claims about abstract domains is that the determination of these non-referential contents will presuppose the development of some representations in our head about the spatiotemporal world. The representations in question are, of course, also part of the spatiotemporal world, and our epistemic access to them is realised just as much by some fallible but reliable natural mechanisms as our epistemic access to the circumstances that they purport to represent.[6] The access is realised when we successfully recall those circumstances under which the relevant symbols can be correctly applied in the declarative representational contexts under consideration. Note that this realisation does not require actual thoughts *about* these representations. When we recall those circumstances under which our concept of tree can be correctly applied in the declarative representational context 'there is a … in front of me', the intentional object of our thought is not our representation of the relevant entities, but instead the entities themselves. What this reveals is that the realm of our own representations can be epistemically accessible to us even when we are actually not thinking of this realm. It is exactly this characteristic of that realm that makes it capable of providing declarative use conditions, for some symbols in some representational contexts, that are not referential in character.

Once we have developed some representations of some aspects of the natural world, we become able to recognise whether or not the declarative use conditions of these symbols (in particular declarative representational contexts) obtain in that world. In the case of synthetic declarative representational contexts, the process is realised by two consecutive epistemic

[6] Concerning the metaphysical commitments underlying this talk of representations, see section 6 of chapter 1.

mechanisms. First, we develop access to our representations, and recall those circumstances under which they can be correctly applied in the particular declarative representational context under scrutiny. Second, we examine whether or not the recalled circumstances actually obtain in the spatiotemporal world. Since the conditions to be checked in these cases are identical with what we, by applying these representations, intend to think of or speak about, the semantics of our synthetic claims about the natural world can be described as fully referentialist in character.

In the case of our analytic claims about spatiotemporal entities, in contrast, we can establish the obtaining of the relevant declarative use conditions by means of a single epistemic mechanism. By developing access to the constituents of these complex representations and recalling the circumstances under which they can be correctly applied in each other's declarative representational contexts, we can already know whether or not the truth conditions of the analytic claims in question actually obtain. We do not need to check what obtains in the represented part of the world, because the declarative use conditions of the relevant constituents in each other's declarative representational contexts are not referential in character.

As an illustration, consider the case of the paradigm analytic claim that bachelors are unmarried men. Note that in synthetic representational contexts the declarative use conditions of the semantically basic constituents of this claim are referential in character and they are supposed to obtain (or not) in the spatiotemporal world. As a result of our cognitive activity, however, some of these basic constituents become suitably linked to each other. Due to this natural relation, when I successfully access these representations in my head, I recall that they can be correctly used in each other's declarative representational contexts independently of what obtains (beyond the detected link) in the spatiotemporal world. Since the relation whose obtaining I detect when I establish that bachelors are unmarried men is not what this analytic claim purports to be about, we can ascertain that the declarative use (or truth) conditions of this truth-apt representation are also non-referential in character.

A more dubious, but in my view correct and philosophically interesting, illustration of the existence of non-referentially true analytic representations about the spatiotemporal world is the (*prima facie* synthetic) claim that water is H_2O. It is a relatively entrenched opinion today that this identity claim expresses a synthetic and necessary truth.[7] In contrast to this opinion, I believe that all knowable necessary truths are analytic in character, and the previous claim is not synthetic and necessarily true at the same time either. The idea that it is can be explained by reference to a conflation. Those who believe that the claim expresses a synthetic necessary truth conflate two senses in which the term 'water' can be applied. The first is the ordinary (pre-scientific) sense, which has no analytic relation to our scientific notion of H_2O. If one speaks about water in this sense, then the claim that water is H_2O expresses a synthetic, contingent and approximate truth. The second sense is the more technical (scientific) one, which is introduced by definition, by invoking our notion of H_2O. If the symbol is used in this sense, then the claim expresses an analytic and necessary truth, and serves as a further example of the existence of non-referentially true analytic representations of the spatiotemporal world. The two senses might be conflated, because their substitution with each other does not alter substantially the subject matter of the symbol.

Similarly to the case of the concepts of bachelor and unmarried man, once we acquired the scientific concepts of water, hydrogen, oxygen and the relevant molecular structure, we no longer need to observe our natural environment in order to know that water is H_2O. This is because, at some stage of our cognitive development, we adopt a certain representation (viz. the scientific concept of water) and establish a certain (semantically significant) natural relation between it and our earlier established

[7] Kripke (1972/1980) has famously argued that since the symbols 'water' and 'H_2O' are both rigid designators, once their referents turn out to be identical in the actual world, they must be identical in every possible world (where this referent exists) too. Accordingly, the claim that water is H_2O must be necessarily true if it is true at all.

representations of hydrogen, oxygen and the relevant molecular structure. Due to the obtaining of this natural relation, when we successfully access these representations in our head, we recall that they can be correctly used in each other's declarative representational contexts independently of what obtains (beyond the detected links, of course) in the spatiotemporal world. Since the relation whose obtaining we actually detect when we establish that water is H_2O is not what this analytic claim purports to be about, we can ascertain that the declarative use (or truth) conditions of this truth-apt representation are not referential in character.

The main contrast between the two illustrative examples is that in the former case the association of the relevant concepts, which guarantees the analyticity and the truth of the claim in question, has no empirical motivation at all. The semantic content of our term 'bachelor' was, as a matter of fact, never independent of that of our composite term 'unmarried man'. The only reason for which we introduced the notion of bachelor to our conceptual scheme is to possess a syntactically simple mental symbol by means of which we can more simply think of unmarried men.[8] The scientific reduction of water to the chemical compound H_2O, in contrast, was a more intricate intellectual manoeuvre, where the introduction of the new scientific notion of water had more complex pragmatic motivations and some empirical preconditions as well. Since the idea that this concept is different from its ordinary counterpart and has an analytic link to the previously acquired notion of H_2O is not widely recognised, let me say a bit more about how I think the story of this reduction should be best summarised.

[8] Had we ever used our term 'bachelor', in synthetic representational contexts, without keeping track of the obtaining or absence of the synthetic declarative use conditions of our complex expression 'unmarried man', the claim that bachelors are unmarried men would have amounted, like the claims that water (in the pre-scientific sense) is H_2O and that water (in the scientific sense) is transparent, to an empirical generalisation, whose referential truth conditions may not obtain in the actual world.

With some simplification, I believe we should distinguish three stages in the development of our representation of the stuffs we have classified under the term 'water' (or its translations) over the last thousands of years. In the long first stage, the term 'water' expressed a concept whose declarative use conditions were identical with some or another group of those easily observable conditions (such as fluidity, transparency, drinkability etc.) whose joint obtaining in our natural environment was so striking and important for our daily life. At some stage of our cognitive development, however, we gradually came to appreciate the practical value of those representations whose declarative use conditions were well-determined and explanatorily more significant. Our pre-scientific notion of water did not score high enough on this scale: first, it became never entirely determined which observable "water characteristics" were actually to be taken as constituents of the declarative use conditions of our concept of water; second, none of the conceivable groupings of these characteristics constituted an explanatorily outstanding (or sufficiently significant) entity in the natural world.

In search for a more comprehensive and simpler explanation of phenomena, at the beginning of the second stage of our conceptual development, we introduced an alternative, scientific concept (a mental symbol syntactically understood), also expressed by the term 'water', whose declarative use conditions were left unspecified and to be determined by our empirical science after further inquiry into the nature of the intended referents of our pre-scientific concept of water. The mode of introduction of this new mental symbol already forecast the future emergence of an analytic link, a semantically significant natural relation, within our system of representation about the natural world. It became clear that by determining the envisaged conditions we would connect our scientific concept of water with an alternative representation of the relevant conditions.

The "discovery" of the molecular structure of water was nothing other than the accomplishment of this scientific task, which brought us to a third stage in our conceptual development.

By determining (i.e. finding an articulate representation of) the conditions whose obtaining or absence in various segments of the spatiotemporal world is meant to govern the declarative use of our scientific notion of water in synthetic representational contexts, we determined the semantic content of this mental symbol, so that our term 'water' from now on expressed two concepts with more or less determined semantic contents.

The contrast between the expressed concepts is manifest in the systematic ambiguity we can observe in the declarative application of our term 'water' in various representational contexts. When the term is meant to express our pre-scientific notion of water, the claim that water is H_2O amounts to an empirical hypothesis, which, understood as a universal generalisation, is strictly speaking false. First, stuffs called water in the pre-scientific sense of the term are not composed purely of H_2O molecules. Second, some stuffs that are composed of H_2O molecules do not qualify as water in that sense, since they do not possess all the observable characteristics we take to be essential for them to do so. Under the same interpretation, the claim that water is a transparent and drinkable liquid is not merely true, but it may be analytic, and thus necessarily true. Whether it is depends on whether or not the predicated qualities are constitutive of the declarative use conditions of the subjects' pre-scientific concept of water.[9] When the term is meant to express our scientific notion of water, on the other hand, the claim that water is H_2O proves to be analytically true, while the claim that water is a transparent and drinkable liquid amounts to an empirical hypothesis, which, understood as a universal generalisation, is strictly speaking false.

[9] Those who have an explicit idea of exactly which observable "water-characteristics" constitute the declarative use conditions of their own pre-scientific concept of water in synthetic representational contexts are supposed to have an analytic connection in their head between the latter concept and those ideas of the relevant conditions. As a general rule, a concept is always analytically related to an alternative representation of its (referential) declarative use conditions.

What we had discovered by the end of this stage was that the obtaining of the (vaguely determined) declarative use conditions of our pre-scientific notion of water, in synthetic representational contexts, strongly correlates with that of the (relatively sharply determined) declarative use conditions of our notion of the chemical molecule H_2O. It was this empirical finding which convinced us that the characteristic of being constituted of H_2O molecules could be a convenient declarative use condition for our new scientific concept of water in synthetic representational contexts. By adopting this stipulation, we ensured that the content of this new symbol became not merely well-determined and explanatorily outstanding, but also sufficiently overlapping with that of our ordinary pre-scientific notion of water. [10]

Summing up, what the previous examples illustrate is that non-referential truth conditions emerge already in our discourses about the (epistemically accessible) natural world. The claims whose truth value is apparently determined by the obtaining or absence of such conditions are our analytic claims about the spatiotemporal realm. In addition, the suggested accounts assume that the conditions in question obtain (if they do) in the domain of representations within our head. Although the emergence of these conditions is the result of our cognitive-epistemic activity, once they are established they obtain independently of our actual thoughts or beliefs about this circumstance. In fact, we can even detect these obtainings without ever thinking of the relevant

[10] Note that the role of experience in motivating this content-determining move does not imply that the analytic truths emerging with this content are empirical in character. Questions about ways of knowledge and justification can be raised only if semantic contents have already been determined. The process of content-determination is utterly conventional, subject merely to practical considerations. The establishment of truth values, in contrast, is an activity whose success or failure (providing that success in this case means holding true what is true and holding false what is false) is an entirely objective issue, determined exclusively by what obtains in the world independently of what our actual opinions are concerning this circumstance. The conflation of the two processes may result not merely in radical empiricism in epistemology, but, as I noted in chapter 5, also in anti-realist conclusions concerning the nature of truth.

representational states of affairs. Finally, the examples also illuminate that, at least in the case of discourses about spatiotemporal domains, the declarative use conditions of our symbols may be referential in some representational contexts, while non-referential in others. With these conclusions in mind, I shall turn now to the second part of the current explanation, and show how the previous representational conditions can serve as declarative use conditions for our representations about abstract domains as well.

Abstraction: The Emergence of Non-Referential Truth Conditions in Discourses about Abstract Domains

In the previous subsection, I explained how our analytic claims about the spatiotemporal world acquire their non-referential truth conditions that obtain (if they do) in the domain of representations within our head. As we saw, a crucial feature of these conditions, which enables them to play this sort of semantic role, is that we can detect their obtaining or absence without actually thinking of them. In this section, I shall argue that, in consequence of a particular concept-forming process within our head, the very same conditions can serve as non-referential truth conditions in the case of our paradigm *a priori* claims about various abstract domains as well.

As a point of departure, let me recall a specific difficulty, briefly touched upon at the beginning of this section, that may seem to balk a proper account of the emergence of non-referential truth conditions in the semantics of discourses about causally inert domains in particular. Suppose someone asks us what makes the allegedly non-referential truth conditions of our mathematical claim *that there are three prime numbers between 70 and 80* in our head the truth conditions of this particular representation. In the case of our synthetic claims about causally effective entities, the corresponding question can be answered by invoking our referential intentions and the selecting work of our conscious attention. The declarative use conditions of our

concept of white and concept of snow, for instance, are presumably singled out by some causal mechanisms including those underlying the selecting work of our conscious attention in the course of the relevant semantic content-determinations. Together with the applied rules of composition, these mechanisms can be invoked in an explanation of what makes the referential truth conditions of our claim *that snow is white* the truth conditions of this particular representation. In the case of our analytic claims about causally effective entities, where the relevant truth conditions are supposed to be non-referential in character, the same question can be answered, as we saw, by invoking our own mental activity that associates more than one of the symbols figuring in the claim in question with the same (previously determined) referential contents. One may say, for instance, that the non-referential truth conditions of our claim *that bachelors are unmarried men* are the truth conditions of this particular claim, because the mental stipulations responsible for their obtaining concerned the content of the key constituents of this particular representation (i.e. the declarative use conditions of our concept of bachelor and our concept of unmarried man).

The apparent difficulty with providing a satisfying answer to the same question in the case of our claims about causally inert domains is that here we cannot suppose that the declarative use conditions of the semantically basic constituents of these claims are singled out from the intended causally inert domains by some causal mechanism including those underlying the selecting work of our conscious attention. This is simply because there can be no causal contact between our minds and the intended causally inert domains.[11] So, if there is an appropriate account of the emergence of non-referential truth conditions in our discourses about the latter domains, then it must presumably rely on an articulate conception of how we develop representations of causally inert entities without any interaction with the relevant

[11] In chapter 6, I argued that alternative conceptions of the existence of a non-causal contact between human minds and some platonic domains are both *ad hoc* and uninformative.

intended domains, and it must also illuminate what makes a certain natural link among some representations of the natural world in our head the truth condition of a particular claim about some causally inert entities. In what follows, I shall provide an outline of such an account.

The key element that I shall rely on in this account is our ability to create new concepts with non-spatiotemporal subject matter from earlier developed ones about some spatiotemporal entities by a certain cognitive process that I shall call abstraction. Unfortunately, the terms 'abstraction' and 'abstract' are used rather ambiguously in the philosophical literature, so before advancing the promised account, I must begin with a brief specification of the sense in which I shall use these terms in that account.[12]

In one received sense, abstractness is contrasted with concreteness and it can be characterised as the property of having no autonomous ontological status. Autonomous existence is often meant to be the privilege of concrete entities. The fact that we can think of entities that have no such existence is then thought to be the consequence of a mental operation, a separating act of human mind. It is this sense in which the particular colour of a concrete object (say, a red rose) can be regarded as an abstract characteristic. The classification reflects the natural, though hardly trivial, metaphysical assumption that the colour of a particular rose exists only as a feature of a concrete individual. If the flower ceased to exist, so would the particular colour as well. We can think of this colour merely because we can separate this feature from its concrete bearer in thought by abstraction. If we were not able to carry out this operation, we could not develop the concept in question, since we could never be acquainted with the intended subject matter independently of other characteristics of the concrete particular.

Abstraction in this sense can be held to be an important operation by which we are able to create new concepts from

[12] A brief overview of the alternative notions that I shall distinguish here has been given in the fifth section of chapter 1.

some others that we have acquired earlier. We carry out this operation by singling out a proper part of the declarative use conditions of an acquired concept and regard them as the declarative use conditions of the newly introduced one. In the case of our example, we may acquire a concept of a particular rose by developing a rich, though syntactically simple, representation of a flower in our direct natural environment. Later on, even in the absence of the intended referent, we can develop a new concept of its particular colour by separating the relevant feature of the concrete object in thought and regarding it as the declarative use condition of a new mental symbol in various synthetic representational contexts. Clearly, abstraction in this sense does not lead us out of our discourses about the natural world. The new concepts developed in this manner will not cease to represent aspects of a spatiotemporal universe.

In a second sense, abstractness is contrasted with particularity and it can be characterised as the property of having no unique location in space and time or, in other words, being a universal that may be instantiated in various spatiotemporal locations. It is this sense in which redness, a property appearing in space and time, qualifies as an abstract universal. Abstractness in the second sense is clearly different from abstractness in the first. A particular colour of a concrete individual is certainly not abstract in the second sense of the term, and a universal feature characterising a number of different individuals can hardly owe its existence to any one of these concrete particulars.

Abstraction in the second sense seems also an important operation by which we can create new concepts from some others that we have acquired before. In particular, we can create concepts by which we can think of spatiotemporal features that may appear in various spatiotemporal locations. In the case of our example, after acquiring representations of particular red and green colours along the lines specified in the previous paragraph, we can simplify our representational system and create some new concepts by which we can think of and speak about these colours in general, disregarding of their particular spatiotemporal location.

Notice that abstraction in this second sense does not lead us out of our discourses about the natural world either. Although our concept of redness resulting from this operation no longer stands for a particular red colour, it nevertheless represents a universal property that can be instantiated in space and time. When we apply these concepts for identifying a subject (what we do with the concept of water in *water is H_2O*) or predicating something about a certain subject (what we do with it in *this liquid is water*), we are still supposed to be thinking of features that can appear in the spatiotemporal world.[13]

It is important to see that, in synthetic representational contexts, the declarative use conditions of our representations of abstract entities in these first two senses of the term can be construed along the standard referentialist lines. This means that an advocate of standard referentialist realism about synthetic truths need not give up her theory in the semantics of our discourses about such abstract domains.[14]

The correctness of the previous tenet is by no means conspicuous. One may wonder, for instance, how an advocate of this referentialist view would explain the emergence of the posited referential relations in the case of our synthetic claims that involve concepts of universals actually not instantiated in

[13] As it has been observed in chapter 6, the concepts whose acquisition Hale and Wright explained by reference to our ability to carry out Fregean abstraction may have subject matters that are abstract only in this second sense of the term. Despite the authors' explicit stipulation, the way they invoke Frege's Abstraction Principle in their reasoning against their anti-platonist opponents seems to suggest that the operation they have in mind does not create concepts of strictly non-spatiotemporal entities from concepts of entities that can be instantiated, and thus known by acquaintance, in space and time. See esp. fn. 28 in chapter 6.

[14] Of course, in the case of analytic claims about these domains, such as *bachelors are unmarried men*, the considerations against a referentialist account of truth presented earlier seem still adequate. If the truth conditions of these claims (again, our analytic claims involving concepts of *in re* universals instantiated in the natural world) were referential in character, then we would have no suitable explanation of how we can learn about the necessary character of these truths, and how we are able to discover them, in possession of the relevant concepts, without any check upon what obtains in their intended referential domain.

space and time (e.g. the concept of unicorn or that of a man taller than 20 feet). The question to be answered here is how we can develop an idea of the intended uninstantiated declarative use conditions without ever getting acquainted with them in the course of our cognitive development.

The most intuitive answer to this question is to suppose that our ideas of uninstantiated universals are the combinatorial results of our previously acquired ideas of instantiated universals.[15] Our idea of a unicorn, for instance, may be thought to be constructed in our head from our ideas of a body, a head, a hind leg, a tail, a horn, a forehead, a horse, a stag, a lion, being a single and being in the middle.[16]

Some may want to challenge this answer by reference to the conceivability of simple (i.e. not constructed) uninstantiated universals. According to this objection, if in synthetic representational contexts the referentialist understanding of the declarative use conditions of our concepts of spatiotemporal universals were true, then we would have no suitable explanation of how the posited referential relations could emerge in the case of our concepts of simple universals that could be but are not instantiated in space and time.

To this challenge, a referentialist may reply that our ideas of the declarative use conditions of these concepts, in synthetic

[15] A recent defence of this response can be found in Armstrong (1989). Note that Armstrong's combinatorialism is meant to be not merely an account of the *semantic* content of our ideas of uninstantiated universals, but also a *metaphysical* account of unrealised possibilities. The extension of the combinatorialist perspective from semantics to metaphysics, however, seems necessary only if we think that the factual basis of modality is to be found in the intended domains of our modal thoughts. On a non-referentialist construal of modal truths, which I propose to adopt in this work, no such extension is required or appropriate. The truth value of our claims about unrealised possibilities will be determined by some contingent facts within our head, rather than by some modal features that obtain in the intended natural domains.

[16] According to the Merriam-Webster's Online Dictionary, a unicorn is a mythical animal generally depicted with the body and head of a horse, the hind legs of a stag, the tail of a lion, and a single horn in the middle of the forehead (http://www.m-w.com/dictionary/unicorn).

representational contexts, are merely negative. What we know of them is that they are not constructible from any conditions that are instantiated in the spatiotemporal world. In the light of this minimal knowledge, these concepts can be correctly applied in any synthetic representational contexts in which this application creates a claim whose truth requires the absence of the intended alien conditions at the intended spatiotemporal locations (as in the context 'there is no … in this room').[17] Since the intended subject matter of these concepts are, *ex hypothesi*, simple uninstantiated universals (i.e. actually non-existing thinkables), a referentialist semantics in their case can account for the above minimal knowledge even in the absence of any acquaintance on our part with these alien conditions.

Having shown that the semantic content of our synthetic representations of abstract entities in the first two senses of the term can be construed along the standard referentialist lines, we can turn now to the case of those representations whose subject matter is supposed to be abstract in a third sense of the term, a sense that excludes the spatiotemporal construal of the relevant intended entities.

It is abstractness in this third sense that characterises the subject matter of our paradigm *a priori* claims and underlies the received definition of platonism as realism about abstract entities. According to this construal, abstractness is contrasted with spatiotemporality, and it can be understood as the property of

[17] Within the referentialist framework under discussion, the declarative use conditions of our concepts of simple alien universals in synthetic representational contexts cannot differ from each other. This is simply because, contrary to our concepts of instantiated and constructed uninstantiated universals, these concepts have no positive semantic link to any particular aspects of reality that they could be about. This is why we know under what circumstances we should declaratively apply our concept of unicorn in the representational context *this creature in front of me is a* …, while we have no similar knowledge in the case of concepts of simple uninstantiated universals. Nevertheless, we can develop more of the latter concepts, since their declarative use conditions can differ in analytic representational contexts. The difference can be established by convention at the time of the introduction of these new concepts into our actual conceptual framework.

having no location whatsoever, whether unique or multiple, in space and time. Beyond the intended subject matters of our paradigm *a priori* discourses, *ante rem* universals or categorical norms and values may also be classified as abstract in this sense of the term. Clearly, this third understanding is different from the previous two. Particular and universal colours appearing in space and time are not abstract in the third sense of the term, while numbers and other non-spatiotemporal individuals are not abstract in the first and the second.

Returning to our main line of thought, what I wish to suggest here is that by abstraction in this third sense we can create new concepts with non-spatiotemporal subject matters from earlier acquired ones about our spatiotemporal environment. By reference to this cognitive process we can explain how we learn determinately to refer to and make truth-apt claims about strictly non-spatiotemporal entities in the absence of a suitable causal interaction with the relevant intended domains.[18] Furthermore, the emerging relations between the respective input and output notions of this process will also explain what makes the earlier invoked natural links among our representations of the natural world in our head the truth conditions of our claims about the relevant abstract (i.e. non-spatiotemporal) domains as well.

Generally speaking, by abstraction in the third sense I mean a cognitive process by which we can create concepts of strictly non-spatiotemporal individuals from concepts of properties that can be instantiated in space and time.[19] The subject matters of two output concepts of this process are meant to differ exactly when the subject matters of the corresponding input concepts

[18] The account is not meant to explain the emergence of our concepts of causally inert subject matters, like values and normative properties, that are supposed to appear in the spatiotemporal world.

[19] As noted in chapter 6, if the stipulation that the instances of Frege's Abstraction Principle provide us with implicit definitions of concepts that are about strictly non-spatiotemporal objects is taken seriously with all its consequences, then the principle can be regarded as a formally adequate characterisation of abstraction in the currently intended third sense of the term.

differ too. Further, the process is meant not to affect the analytic relations of the input concepts. Consequently, any two concepts created by this process will be analytically related if and only if their antecedents were analytically related too. The correspondence is guaranteed by the fact that the operation keeps the declarative use conditions of the output concepts in each others' analytic representational contexts identical with those of the respective input concepts in each others' analytic representational contexts. Due to this identity, the truth conditions of our analytic claims composed of the output concepts will be also identical with those of our analytic claims composed of the respective input concepts of this operation. In principle, we can conceive the abstract counterpart of any property that can be instantiated in space and time.[20]

In synthetic representational contexts, such as '… exists in the non-spatiotemporal part of the world', the declarative use conditions of the resulting concepts preserve their referential character. Accordingly, the above explanation of the emergence of our concepts of abstract entities explains the emergence of our ideas of these synthetic declarative use conditions as well. In chapter 6, I argued that we have no reason to suppose that we could ever discover whether or not these referential conditions actually obtain. Consequently, I believe that we cannot hope to acquire synthetic knowledge of non-spatiotemporal domains. Of course, the impossibility of such knowledge does not imply that we cannot acquire any sort of knowledge of these domains. This is because our ideas of abstract entities in the third sense can be declaratively applied in analytic representational contexts as well, in which the conditions whose obtaining or absence determines whether or not we can correctly apply the relevant concepts in those contexts are non-referential in character.

The only prerequisite of the existence of such representational contexts is that at least some of the actual input notions of abstraction in the third sense must be analytically

[20] Plato's claim that every spatiotemporal characteristic is merely the reflection of an atemporal Form provides the clearest illustration of this capacity.

related to each other. Since the operation does not affect these analytic relations, any concepts which have been developed by abstraction in the third sense from analytically related notions of spatiotemporal entities will be analytically related too. If they appear in each other's truth-apt representational contexts, then our reliable epistemic access to the relevant representational facts in our head will (mostly) enable us to establish the truth value of these claims. We will not need to check what obtains in the represented abstract part of the world, because the declarative use conditions of the relevant concepts in each other's representational contexts will be non-referential in character. As we observed before, although the emergence of these conditions is the result of our cognitive-epistemic activity, once they are established they obtain independently of our actual thoughts or beliefs about this circumstance, and we can detect these obtainings without ever thinking of the relevant representational states of affairs.

Consider, for instance, the acquisition of numerical concepts applied in pure arithmetic, such as the concept of number three. The subject matter of this concept is an entity that cannot appear in space and time. It cannot appear there, because we do not think of it as something possibly spatiotemporal.[21] If my

[21] To be more exact, our *referential intentions* accompanying the use of this concept exclude the viability of the naturalist construal of this subject matter. If we maintained that number three is non-spatiotemporal merely because we never saw this object in the natural world, then our thought would not guarantee the abstractness of this entity in the relevant sense of the term. One may object that if the previous claim were true, then the sentence 'The number of chairs in front of me is three, just like the number of primes between 70 and 80, and the number of King Lear's daughters in Shakespeare's drama' could not be true. It could not be true, since a strictly non-spatiotemporal entity cannot be identical with a spatiotemporal one appearing in the real world (as the number of chairs in front of me) or in a fictive universe (as the number of King Lear's daughters in Shakespeare's drama). Since the above sentence can be true, our referential intentions accompanying the use of our concept of number three must allow for a domain-independent construal of the concept's subject matter. Note, however, that if we are asked whether the arithmetic object number three could appear in space and time, our answer is unanimously negative. This supports the original claim that the way we think of the objects of pure arithmetic excludes the

considerations against platonist contact theories of knowledge of causally inert entities advanced in chapter 6 are sound, then we must abandon the idea that the semantic content of this concept is determined by the interaction of our conscious attention with the domain of pure arithmetic. On the other hand, we clearly possess the concept and we use it in various representational contexts in a highly disciplined manner. Apparently, this use is informed, under normal epistemic circumstances, by the obtaining (or absence) of some non-epistemic conditions in the actual world that we can reliably detect by a purportedly *a priori* epistemic mechanism. So, how did we acquire this concept? How did this notion acquire its strictly non-spatiotemporal subject matter and its arguably spatiotemporal declarative use conditions in those representational contexts in which we are supposed to apply it in an evidence-governed way?

According to the account being proposed here, our concept of number three is developed by abstraction in the third sense from our concept of the number of objects in triples or, what is

naturalist construal of these subject matters. The apparent conflict between the quoted examples, however, disappears if we recall that in natural language a physical symbol can be applied to express various mental representations. The fact that, under some actual circumstances, our numerical expression 'three' can be correctly applied in the representational context '*The number of chairs in front of me is* …' would be in conflict with the fact that the same term can be correctly applied in the representational context '… *cannot appear in space and time*' as well only if it were supposed that the term expresses the same concepts in these applications. (For a currently irrelevant complication, see also fn. 27 in chapter 6.) Dropping this assumption, we can eliminate the conflict by maintaining that the term can be correctly applied in the former context, because it expresses there a concept of a numerical property that may characterise groups of individuals independently of whether these are fictive or real, or abstract or spatiotemporal, and it can be correctly applied in the latter context, because it expresses there a concept of an individual that cannot appear in space and time. What I wish to explain in the main text is the emergence of the semantic content of the latter notion and *its* linguistic expression. The conflation of the two senses in which mathematical symbols can be applied (viz. the one observed in pure mathematics and the other operative in the empirical sciences) is at the heart of the influential empiricist reasoning from the indispensability of mathematics in the empirical sciences to the existence of abstract mathematical objects and properties.

the same, from our concept of being a triple (i.e. a numerical property that characterises spatiotemporal groups with three members).[22] The input concept of this operation is supposed to be acquired, together with the concepts of some other basic numerical properties and operations, via acquaintance with the spatiotemporal instances of the intended entities.[23]

The process fully determines the subject matter of our notion of number three. It makes us apply this notion with the intention to think of an individual that cannot appear in space and time and is the strictly non-spatiotemporal correlate of the numerical property of being a triple. This much specificity in our referential intention accompanying the declarative use of our concept of number three is sufficient to distinguish this subject matter from the intended referents of any other notions developed by this type of operation, and, consequently, from any other entity that we can think of.

How about the determination of the respective declarative use conditions? Well, in synthetic representational contexts, the declarative use conditions of our concept of number three are supposed to preserve their referential character. Accordingly, by determining the previous subject matter, abstraction in the third sense can also fix the synthetic declarative use conditions of our

[22] Here I suppose that the semantic content of the alternative input expressions invoked is the same, so they express the same mental symbol or the same concept. By applying them declaratively in various representational contexts, we intend to speak about a numerical property that can be multiply instantiated in space and time. The expressions in question are still not interchangeable, because they have different syntactic roles: while the term 'the number of objects in triples' can identify a certain numerical property as the subject of a proposition, the term 'is a triple' predicates this property of some subjects identified otherwise. (Again, for a minor complication concerning the intended referent of the former term, see also fn. 27 in chapter 6.)

[23] Our concepts of non-instantiated or actually unobserved numerical properties are presumably developed by recursive analytic stipulations from the most basic ones, rather than by direct acquaintance with the spatiotemporal instances of these universals. The determination of exactly which numerical concepts are acquired by acquaintance, and thus independently of each other, and which are developed from these by analytic stipulations, is a task of our empirical sciences.

concept of number three. It is the obtaining or absence of these referential conditions that we are supposed to detect in order to know, for instance, whether number three, a platonic object, exists in the actual world. Again, if my arguments against the conceivable platonist epistemologies in chapter 6 are correct, then we have no reason to suppose that we could ever discover whether or not these referential conditions actually obtain, and thus whether the synthetic claim just mentioned about number three is true or false.

Of course, in pure arithmetic our concept of number three is used in a highly disciplined and evidence-governed way. When we are asked about the truth value of an arithmetic thought involving the notion of number three, our response is supposed to be non-arbitrary, based on the recognition of some real fact, the obtaining or absence of some non-epistemic conditions in the actual world.[24] In other terms, our cognitive and linguistic practice in pure arithmetic suggests that we can still acquire some knowledge of the strictly non-spatiotemporal domain of this discipline, including the subject matter of our concept of number three.

If this knowledge cannot be synthetic, then it must be analytic in character. As we saw, the declarative use conditions of our mental and physical symbols in analytic representational contexts cannot be adequately construed in referential terms, because on such a construal we could have no suitable explanation of how we can learn about the necessary character of analytic truths, and how we can discover them, in possession of the relevant concepts, without any check of (i.e. any cognitive interaction with) what obtains in the relevant intended referential domains. By adopting a non-referentialist, naturalistic construal of the truth conditions of our standard arithmetical claims, and thus our claims about number three, we can remove the main

[24] In chapter 4, I argued that without this substantive realist construal of the declarative use conditions of our mental and physical representations, we cannot suitably explain the objectivity of their correct declarative applicability or truth.

obstacle from the path of a proper contact theory of this type of knowledge as well.

Note, however, that this move implies also that the former account of how abstraction in the third sense determines the strictly non-spatiotemporal subject matter of our concept of number three does not explain how the notion acquires its non-referential declarative use conditions in analytic representational contexts, as those occurring in pure arithmetic. Apparently, a proper account of the latter phenomenon must invoke some fact about the suggested *explanans* whose explanatory role is not limited to the determination of the relevant strictly non-spatiotemporal subject matters.

The fact that I propose to invoke at this point is, again, that abstraction in the third sense fully preserves the analytic relations of the concepts acted upon. More specifically, I claim that the operation keeps the declarative use conditions of any two output concepts in each others' analytic representational contexts identical with those of the respective input concepts in each others' analytic representational contexts. In other terms, I suggest that the non-referential truth conditions of our analytic claims involving the former concepts are the same representational conditions that constitute the non-referential truth conditions of our analytic claims involving the latter.

In pure arithmetic, for instance, we think we know that one plus one plus one equals three. In other terms, we think we have good reason to assert that our concept of number three can be correctly applied in the representational context 'One plus one plus one equals …'. The suggestion we made earlier is that we can explain the possibility of this knowledge only if we suppose that the arithmetical thought in question is analytic, and the declarative use conditions of our notion of number three in the above context are non-referential in character, obtaining in our head, in the domain of our representations of the spatiotemporal world. What needs to be explained now is what determines the alleged semantic relations between the arithmetical concept under scrutiny and these non-referential conditions. Our answer to this question is briefly that the relations are established by those

cognitive mechanisms that underlie the acquisition of our arithmetical concepts, including our idea of number three.

In the case of the above example, the explanation runs as follows. In the course of our cognitive development, first we develop our concepts of numerical properties that are instantiated in space and time. These include our concepts of being a single and being a triple. During the same time, we acquire our concept of addition and our concept of being equal, whose subject matter (an operation and a relational property, respectively) can also occur in the natural world. The emergence of the existing semantic links between these basic concepts, on the one hand, and their referential use conditions in synthetic representational contexts, on the other, can be explained by invoking our acquaintance with the spatiotemporal instances of these subject matters.

At some point in this development, maybe after the recognition of the correlation between the obtaining of the referential declarative use conditions of our concept of being a triple, on the one hand, and that of our concept of being a single and another single and yet another single, on the other, we introduce a natural link between these representations in our head, which constitutes the obtaining non-referential truth condition of various analytic thoughts involving these concepts, and becomes constitutive of the semantic content of the related notions as well.[25] By detecting the obtaining of this condition through a reliable epistemic mechanism in our head, for instance, we can know that George's single apple, Peter's single apple and John's single apple on the table in front of us constitute a triple, without checking separately, after establishing the existence of the singletons on the table, whether the referential declarative use conditions of our concept of being a triple (in the above

[25] Before the introduction of this natural link, if there is such a stage, our belief that a group of three distinct entities is a group of a single entity, another single entity, and yet another single entity amounts to an empirical generalisation, whose truth conditions are referential in character.

representational context) actually obtain in the relevant part of the spatiotemporal world.[26]

In possession of the previous concepts we can form various synthetically or analytically true ideas of the two numerical properties, the relation and the operation in question, but we cannot develop thoughts of the strictly non-spatiotemporal domain of pure arithmetic. In particular, we cannot entertain the idea that one plus one plus one equals three in the sense we are supposed to do that in pure arithmetic. In order to formulate this thought, we must acquire the concepts of number one and number three as well.[27] According to the account under consideration, these concepts are developed by abstraction in the third sense from our earlier acquired concepts of being a single and being a triple, respectively. The operation determines the subject matter, and therewith the referential declarative use conditions (for synthetic representational contexts), of the two output concepts in the manner specified above. On the other hand, it also preserves the analytic relations of the concepts it acts upon by associating its output concepts with the non-referential declarative use conditions (for analytic representational contexts) of their respective antecedents.

Due to the latter feature, our concept of number three is analytically related to our concept of one plus one plus one just like our concept of being a triple is to our concept of being a single and another singe and yet another single. Moreover, the feature also ensures that the non-referential declarative use conditions of our concept of number three in the analytic

[26] The fact that the truth of this thought requires the obtaining of some conditions in the relevant intended domain (viz. the existence of the apples denoted by the definite descriptions on the table in front of us) indicates that, despite the analytic content mentioned in the main text, the whole idea cannot be analytically true. For a thoughtful discussion of the relation of our "arithmetic" and "perceptual" criteria for the applicability of our arithmetic terms, see Craig (1975).

[27] I suppose that the concepts of addition and equality that we apply in pure arithmetic do not differ from our ideas of addition and equality in the spatiotemporal world.

representational context 'One plus one plus one equals …' are identical with the representational conditions constituting the declarative use conditions of our concept of being a triple in the representational context 'One apple plus another apple plus yet another apple make … apples on the table in front of me'. As before, the analytic links in question are constitutive of the semantic contents of the concepts involved. Accordingly, the fact that we possess the concepts of number one, number three and being equal guarantees the truth of the thought that one plus one plus one equals three, independently of whether or not the intended platonic (i.e. real and strictly non-spatiotemporal) referents of this thought actually obtain.

The account clearly removes the main obstacle from the path of a naturalist contact theory of our (purportedly *a priori*) arithmetical knowledge of number three as well. If the truth conditions of our standard arithmetical beliefs about this abstract individual obtain in the domain of representations within our heads, then we can suppose that the formation of these beliefs is constrained by those reliable epistemic mechanisms that provide us with evidence of the obtaining or absence of these conditions.

The emergence of non-referential truth conditions can be explained in similar terms in the case of our purportedly *a priori* beliefs about other strictly non-spatiotemporal domains, such as the domain of pure geometry, set theory and logic, as well. Our concepts of geometrical, set theoretic and logical objects are supposed to be developed also by abstraction in the third sense from earlier acquired concepts of properties that can be instantiated in space and time.[28] In each case, the operation

[28] Our concepts of geometrical objects, for instance, are supposed to be rooted in our concepts of spatial properties, our notions of set-theoretic objects in our notions of group-member relations, while our ideas of logical objects (e.g. platonic ideas and propositions) in our ideas of semantic properties. Note also that the account advocated here does not imply that to each concept of a strictly non-spatiotemporal object we can find in our head an earlier acquired concept of a certain property from which the former notion is developed by abstraction. Once we acquired some basic ideas of an abstract domain, we can develop further

ensures that the declarative use conditions of the relevant output concepts in analytic representational contexts coincide with those representational conditions that constitute the non-referential declarative use conditions of the corresponding input concepts in such representational contexts.

Summing up, my primary aim in this section was to show how the advocates of a non-referentialist construal of truth in the semantics of our purportedly *a priori* discourses about abstract domains can account for the emergence of determinate semantic relations between our truth-apt representations within these discourses, on the one hand, and their arguably non-referential truth conditions, on the other, without assuming that we ever consciously attend to or think of these conditions in the course of the relevant content-determination. As a first step, I provided a brief review of those theoretical assumptions that we could make of the nature of the envisaged non-referential conditions in view of some explanatory *desiderata* put forward in chapter 2. The upshot of this survey was that the conditions in question are best understood as substantive real states of affairs that obtain (if they do) in our head, in the domain of our representations that we develop in the course of our cognitive engagement with the spatiotemporal world. Second, I explained how these representational conditions may become associated, as non-referential truth conditions, with our analytic claims about the natural world. The crucial assumption that I made in this part was that the suggested representational conditions can fulfil this particular type of semantic role, because we can reliably detect their obtaining or absence without actually thinking of them. Finally, I turned to the case of our purportedly *a priori* discourses about abstract domains, and argued that, due to a particular concept-forming mechanism in our head, which generates concepts of strictly non-spatiotemporal objects from earlier acquired notions of some properties that may be instantiated in

notions of it by definition (i.e. by composing new concepts from those basic ones) as well.

space and time, the non-referential truth conditions of our analytic claims about the spatiotemporal world can serve as the non-referential truth conditions of our standard (analytic) claims about the relevant abstract domains as well.

If the account presented in this section is correct, then we can conclude that the non-referentialist conception advocated in this work can satisfy the third adequacy condition set for a theory of truth in the semantics of our paradigm *a priori* discourses about abstract domains in chapter 2. Of course, this explanatory virtue in itself does not guarantee the superiority of this conception over its referentialist alternatives. The semantical positions criticised in chapters 4, 5 and 6 are clearly not devoid of explanatory potential either. Their inadequacy became manifest only against the background of a larger pool of *explananda*. What we have learned in the previous chapters is that none of these alternatives can account, either by itself or as part of a larger theory, for all those phenomena whose joint explanation we agreed to regard as a minimal condition of adequacy for a theory of the relevant truths. Now, if we want to make a convincing case in support of the suggested non-referentialist construal of truth in the semantics of our paradigm *a priori* discourses (i.e. our discourses about abstract domains), then we must show that the conception under scrutiny can meet all explanatory requirements that we took as a minimal condition of adequacy for a theory of *a priori* truth in chapter 2. It is this task that I intend to accomplish in the section that follows.

3. The Explanatory Adequacy of a Representationist Construal of the Paradigms of A Priori *Truth*

In the second section of chapter 2, I claimed that a proper construal of *a priori* truth must explain, either in itself or as part of a larger theory, two kinds of characteristics of its subject matter: those which are possessed by any kind of truth, and those which are specific features of the *a priori* instances under scrutiny. Having said this, I compiled a list of the most important of these

explananda. The list included seven of the former and four of the latter type of phenomena. Since it was by no means meant to be complete, I argued that the joint explanation of these characteristics should be regarded as a minimal condition of adequacy for a construal of *a priori* truth.

In the previous section, I showed that the non-referentialist account advocated in this work can meet the third explanatory requirement specified in chapter 2. In particular, I argued that we can develop an (empirically confirmable) proposal of how our purportedly *a priori* claims about abstract domains may acquire both their representational truth conditions and their non-spatiotemporal intended referents. In this section, I shall show that the account can meet the other ten explanatory requirements as well. If my proposals prove to be correct, then we can conclude that the account under scrutiny provides an adequate characterisation of truth in the semantics of our paradigm *a priori* discourses, at least in view of the conditions of adequacy specified in chapter 2.

Fit with a General Construal of Truth

The first *desideratum* on our list toward a construal of the purported paradigms of *a priori* truth, which, arguably, must be maintained *vis-à-vis* a theory of any specific kind of truth, was that the account should harmonise with our general conception of truth. In particular, the semantic property characterised by the account must fall into the extension of our general concept of truth.

In chapter 3, I showed how the observation of this adequacy condition may lead someone with a broadly Tarskian, referentialist concept of truth to the denial of any non-referentialist response to Benacerraf's dilemma in philosophy of mathematics and the philosophy of our discourses about causally inert domains in general. The standard referentialist objection to the non-referentialist perspective is that an account that does not understand truth in terms of intended referents cannot be taken as a conception of truth. According to this perspective, what we

mean by this term is something essentially related to the subject matter of the relevant truthbearers. If truth is the property of having a subject matter (i.e. being about something) that obtains in the actual world, then our claims about abstract states of affairs cannot possess this property in virtue of the obtaining of some conditions in our head.

An advocate of the suggested representationist construal of truth in the semantics of our purportedly *a priori* discourses, of course, may argue that what our cognitive and linguistic practice in pure logic and mathematics (together with some assumptions of human knowledge acquisition) suggests is exactly that the received referentialist notion of truth is inadequate. According to this line of thought, what we mean by applying the term 'truth' in our discourses about abstract domains cannot be suitably specified in terms of the intended abstract subject matters.

Note that a referentialist may agree that the way we use the term 'truth' in our purportedly *a priori* discourses about abstract domains would disconfirm the referentialist reconstruction of our concept of truth if we could show that the term is applied in the same genuine sense in the latter discourses as in our empirical discourses about the natural world. Still, she may insist that we have no reason to suppose that the latter condition actually obtains. In fact, the very same evidence that makes her opponent abandon the standard referentialist construal of truth in the semantics of the relevant discourses can be invoked to query the idea that the term 'truth' is applied in the same genuine sense in these discourses as in the others.

The fact, however, that we assert something about truth in the semantics of certain discourses that we deny about it in the semantics of some others does not exclude that these claims are about the same semantic property. After all, most of our beliefs are not definitive of the nature of their subject matters. What is, on the other hand, clearly required from someone who insists that our cognitive and linguistic practice in discourses about abstract domains undermines the general referentialist construal of truth is an account of what makes the intended referent of our

notion of truth in the latter discourses the same as that in our discourses about the natural world.

Now, as has been briefly adverted to in chapter 3, from the perspective of a substantive realist, broadly use-theoretic semantics (a framework that emerged from our discussion of Horwich's minimalism in chapter 4), the opponents of standard referentialism may suggest that truth is best understood as *the property of possessing declarative use conditions that actually obtain in the world, in so far as the bearer of this property is a truth-apt mental or physical representation*. This understanding does not assume that truth has anything to do with the actual obtaining or absence of what its bearers purport to be about, and thus it is compatible with the suggested non-referentialist idea that our paradigm *a priori* claims about abstract domains are true or false in virtue of the actual obtaining of some representational conditions in our head. In so far as the construal that a non-referentialist provides is a construal of the correct declarative use conditions of our paradigm *a priori* beliefs, her account qualifies as an account of the paradigms of *a priori* truth, and thus satisfies the first adequacy condition set for such theories in chapter 2.

Objectivity of Truth

The second *explanandum* on our list for a construal of the purported paradigms of *a priori* truth (as well as, again, for a construal of any specific kind of truth) was the apparent objectivity of this semantic property, or in other terms, the fact that no one is ever conceptually prevented from committing epistemic mistakes (i.e. judge something to be true (or false) that is in fact false (or true)).

In chapter 4, I argued that the most natural way to explain this characteristic is to maintain that the truth value of our truth-apt representations is determined by the obtaining or absence of some conditions in the actual world, independently of anyone's actual knowledge or opinion of this circumstance. Putting it briefly, my suggestion was that the truth conditions of these representations are to be construed in a substantive realist way.

Apparently, the representationist construal advocated here of the truth conditions of our paradigm *a priori* claims about abstract states of affairs satisfies this explanatory requirement. One may object that it does not, since it implies that the obtaining or absence of the relevant truth conditions is the result of our epistemic activity, some stipulations that we make while developing our representations of the natural world. In response to this charge, however, an advocate of the suggested representationist account may observe that the reality of a certain condition requires merely that its obtaining or absence be independent of our actual knowledge, opinion or thought of this particular circumstance, rather than of our epistemic states and activities in general.[29] As has been emphasised earlier in this work, although the obtaining or absence of the crucial natural links among our representations is, indeed, the result of our epistemic activity in the spatiotemporal world, our knowledge, opinions or thoughts of these particular facts in our head are definitely not among the relevant epistemic determinants.

Summing up, since the suggested representationist construal offers a substantive realist understanding of the truth conditions of our paradigm *a priori* beliefs about abstract domains, an advocate of this account can explain the objectivity of *a priori* truth by reference to the fact that this property characterises its bearers in virtue of the obtaining of some substantive, real conditions.

Knowledge / Reliability of Evidence

The next phenomenon that we said an account of the paradigms of *a priori* truth, and in fact an account of any knowable kind of truth, must (arguably, in conjunction with our actual theory of human cognition) somehow explain is the possibility of

[29] Note that our intuition about truth in pure logic and mathematics is not that it obtains independently of any epistemic activity in the world, but instead that it obtains independently of what anyone ever actually thinks of this particular circumstance in the world.

knowledge acquisition, or reliable belief formation, of the relevant sort of truths.

As we have seen in chapter 3, the essential point behind Benacerraf's epistemological challenge to the standard platonist construal of mathematical truth (or its modified generalisation against the corresponding realist and referentialist construals of truth in discourses about causally inert domains) was that the account cannot satisfy the current adequacy condition, since it undermines the possibility of a reasonable explanation of how we can acquire knowledge or reliable evidence of what can be truly held about the relevant intended domains. In chapter 6, I argued that this charge against platonist theories of knowable truths is legitimate, and that a suitable construal of the paradigms of *a priori* truth must be naturalist and non-referentialist in character.

Obviously, the representationist account advocated in this chapter satisfies this requirement as well. It claims that the truth conditions of our paradigm *a priori* beliefs about abstract domains obtain (if they do) in our head, which means that our knowledge of or reliable evidence for what can be truly believed about these domains can be explained by reference to some causal mechanisms, whose nature is (in principle) just as open to systematic empirical study and characterisation as the nature of our knowledge or reliable belief formation of the natural world.

Intersubjectivity of Semantic Content

The next *explanandum* on our list in chapter 2 for an account of the paradigms of *a priori* truth, as well as for an account of any communicable kind of truth, was the intersubjectivity of the semantic content of the bearers of this semantic property. Although there are well-known sceptical considerations querying the existence of shared meanings, our successful daily communicative practice in pure logic and mathematics as well as in our empirical discourses about the natural world suggests that different people are still capable of entertaining (largely) the same ideas (i.e. thoughts that are about the same things and are applicable, in any given representational context, under the same

circumstances). So, a theory that cannot support an account of this phenomenon can hardly qualify as an adequate characterisation of the relevant type of truth.

In the first section of chapter 6, I conceded that the standard platonist construal of the semantic content (i.e. the truth conditions and the subject matter) of our claims about abstract domains can meet this explanatory requirement. If John's and Peter's respective thoughts *that one plus one plus one equals three* are supposed to be about, and also made true by the obtaining of, the same platonic conditions in the world, then any explanation of how John's and Peter's relevant mental symbols acquire their semantic contents (i.e. any account of the third *explanandum* specified in chapter 2) will *eo ipso* explain how these two subjects become able to share the above mathematical thought.

Note, however, that assuming the actual obtaining of the intended states of affairs cannot be a prerequisite for a successful account of our capacity to share and communicate ideas in different discourses. If it were, then we could have no such account of this capacity in the case of our beliefs about fictive or real-but-actually-uninstantiated entities.

The representationist account defended in this chapter preserves the idea that two people can entertain thoughts about the same abstract conditions (e.g. that John's and Peter's respective thoughts *that one plus one plus one equals three* are about the same abstract conditions), but it denies that the truth conditions of these thoughts are to be understood in terms of those abstract intended referents. Instead of adopting this referentialist construal, the account rather assumes that the truth conditions of our purportedly *a priori* thoughts about abstract domains obtain in our heads. What the advocates of this non-referentialist alternative suggest is that the fact whose reliable detection gives rise and justifies John's knowledge *that one plus one plus one equals three* is different from the one whose reliable detection is the grounding source of Peter's knowledge *that one plus one plus one equals three*. The former is supposed to be found in John's head, while the latter in Peter's. Now, the obvious question that an opponent of this account may raise is how this

difference between the respective truth conditions could ever be reconciled with the claim that the two representations under consideration have the same semantic content.

The non-referentialist answer to this question is that the truth conditions of a particular subject's (purportedly *a priori*) ideas of abstract domains obtain (if they do) not merely in this subject's head, but also in every single person's head who has acquired the ability to entertain those ideas in her own mind. Due to this construal, the semantic content of John's and Peter's respective thoughts *that one plus one plus one equals three* proves to possess the same content, despite the fact that the particular representational conditions whose obtaining gives rise to John's belief are clearly different from those whose obtaining inform Peter's belief. On the representationist account under scrutiny, both ideas are about the same abstract entities, and both are associated with the same (functionally identified) representational declarative use conditions in (the relevant) human heads.

One may wonder what makes it the case that John's and Peter's thoughts in the previous example are about the same abstract entities if there are no abstract entities in the world. As we said, the same question can be raised concerning our apparently co-referential thoughts about fictive or actually uninstantiated entities. The proper answer to this question can be derived from what has been said, in the previous section as well as in the first section of chapter 5, about our various capacities of developing new concepts from earlier acquired ones by subjecting the latter to some concept-forming operations, which alter the subject matter of their input concepts in a systematic way.

In the case of our particular example, John's and Peter's respective thoughts *that one plus one plus one equals three* are about the same abstract entities, because the corresponding conceptual elements of these thoughts also share their subject matter, while the latter condition obtains, because John and Peter developed these notions by the same kind of operations from corresponding concepts that represented the same obtaining conditions in the spatiotemporal world.

Similar response can be given to someone who wants to see why the relevant representational conditions in John's and Peter's heads can be regarded as fulfilling the same functional role. In the previous section, it has been argued that the conditions in question are established by the two subjects while they are developing their representations of certain properties and operations instantiated in the spatiotemporal world. Since in the case of the corresponding conceptual elements of the above thoughts in John's and Peter's heads these original natural properties and operations are supposed to be the same, it seems legitimate to maintain that the relevant representational features in the two heads have the same function in the natural world: they constitute the factual basis of analytic relations among representations, in different heads, of the same aspects of reality.

Summing up, similarly to her referentialist opponent, an advocate of the representationist construal articulated in this chapter can explain the intersubjectivity of the semantic content of our paradigm *a priori* beliefs about abstract domains by invoking the shared causal origin of the relevant semantically equivalent representations. An important novelty of the previous explanation is, however, that it does not presuppose that the truth conditions of these beliefs can be specified in terms of the intended abstract referents.

Observable Convergence of Beliefs

The next feature that occurred on our list in chapter 2 as an *explanandum* for an account of any type of truth was the observable convergence (or divergence) of our opinions concerning the distribution of the characterised semantic property. In our paradigm *a priori* discourses about abstract domains, the measure of convergence among various subjects' semantically equivalent beliefs is prominently high. Accordingly, an account of truth in the semantics of these discourses must support a suitable explanation of this high measure of convergence.

In the first section of chapter 6, it has been argued that if we had a belief-forming mechanism by which we could reliably detect the obtaining or absence of platonic truth conditions, then by reference to this capacity the advocates of the standard platonist construal of the paradigms of *a priori* truth could explain the observable convergence of our opinions concerning the distribution of this semantic property.[30] They could simply argue that different subjects' reliably generated beliefs about the obtaining or absence of the very same conditions must largely coincide.

As in the case of the previous *explanandum*, however, it must be noted that assuming the existence of a reliable information-conveying mechanism between minds and intended referential domains cannot be a prerequisite for a successful account of the observable convergence of our opinions in different discourses. If it were, then we could hardly develop such an account in the case of our converging beliefs about fictive domains.

The proponents of a representationist construal of the paradigms of *a priori* truth do not invoke the existence of such a mechanism in their account of this *explanandum* either. What they assume is rather the existence of a corresponding reliable information-conveying mechanism between minds and the domains of relevant truth conditions. In the case of our paradigm *a priori* beliefs about abstract states of affairs, the latter domain is, in their view, to be found in human heads. Accordingly, what they invoke in their account of the observable convergence of these beliefs is the existence of a reliable information-conveying mechanism between human minds and the alleged (obtaining or absent) representational truth conditions in human heads.[31]

[30] The primary purpose of chapter 6 was, of course, to show that we have no suitable ground to suppose the existence of such reliable belief forming mechanism.

[31] More precisely, the mechanism invoked must be "first-personal" between the subject's mind and the relevant representational conditions in her *own* head. The representationist construal under scrutiny does not exclude the possibility of empirical knowledge of the obtaining of these conditions in human heads.

Prima facie it may occur that a representationist cannot make use of the previous assumption, since, in her view, the particular conditions whose obtaining or absence is supposed to give rise, respectively, to John's and Peter's purportedly *a priori* beliefs about a certain abstract state of affairs are not the same: the conditions detected by John obtain in John's head, while those detected by Peter obtain in Peter's. Clearly, if these conditions were the same, then John's and Peter's reliable epistemic access to their obtaining or absence could explain the convergence of these subject's relevant co-referential beliefs. In absence of this identity, on the other hand, there seems to be no guarantee that the invoked reliable mechanisms give rise to the observed convergence.

Note, however, that what the explanation of this phenomenon seems to require, beyond the existence of the above reliable information-conveying mechanisms in our head, is not the identity, but merely the co-obtaining of the detected representational truth conditions. Returning to our previous example, if the representational conditions whose obtaining is meant to give rise to John's belief *that one plus one plus one equals three* obtain in John's head if and only if the corresponding representational conditions whose obtaining is supposed to inform Peter's belief *that one plus one plus one equals three* obtain in Peter's head, then the fact that both John and Peter develop their respective beliefs by reliable information-conveying mechanisms can explain the observable convergence of these beliefs.

So, what the proponents of the representationist account must show is merely that their theory implies the co-obtaining of the respective representational truth conditions of various subjects' purportedly *a priori* beliefs about a certain abstract state of affairs within these subjects' heads. Fortunately, the obtaining of this implication can be easily demonstrated. As we saw, the account assumes that the obtaining (or absence) of the relevant representational truth conditions of various subjects' purportedly *a priori* beliefs about a certain abstract state of affairs is constitutive of the semantic content of these beliefs. Accordingly, the fact that John and Peter are equally capable of entertaining

the idea *that one plus one plus one equals three* entails that the respective representational truth conditions of these beliefs in John's and Peter's heads co-obtain.

Summing up, the adoption of a representationist construal of the paradigms of *a priori* truth does not undermine the suitable explanation of the observable convergence of our paradigm *a priori* beliefs. Instead of assuming the existence of a reliable information-conveying mechanism between human minds and the intended abstract referential domains of these beliefs, an advocate of this construal can account for the current *explanandum* by invoking the existence of a similar cognitive mechanism which provides the subjects with reliable information of the obtaining or absence of the suggested representational conditions within their own heads. Since the account assumes that the truth conditions of these beliefs co-obtain in the heads of those subjects who can entertain the relevant paradigm *a priori* thoughts, the high reliability of the suggested belief-forming mechanism can explain the observable strong convergence of these beliefs.

Infinity of Semantically Non-Equivalent Truth-Apt Representations

The last feature on our list in chapter 2 that must be explained by an account of virtually any type of truth was that the property under scrutiny can characterise an infinite number of semantically non-equivalent bearers. In arithmetic, for instance, we can in principle entertain any particular member of the infinite thought-series *1+1=2, 2+1=3, 3+1=4*, etc. (i.e. any instance of the thought-scheme $n+1=k$, where k and n are natural numbers and k is the successor of n). Consequently, a proper construal of arithmetical truth must support an account of what makes it the case that there are infinitely many semantically non-equivalent representations that may be the bearer of arithmetical truth.

In chapter 6, it was noted that the standard platonist construal of the examined paradigms of *a priori* truth can meet this explanatory requirement by endorsing a simple referentialist construal of the semantic content of our paradigm *a priori* claims

and observing that the intended referential domain of these truth-apt representations is infinite in character. *Prima facie*, a similar explanation should be available for the advocates of the suggested representationist construal of the paradigms of *a priori* truth as well. After all, a representationist queries neither the referential aspects of the semantic content nor the infinity of the intended referential domain of our paradigm *a priori* beliefs.

Opponents, however, may object that the substantive realist (use-theoretic) semantical framework put forward earlier in this work undermines the adequacy of this standard explanatory strategy in the representationist's case. According to that framework, semantic contents are to be specified in terms of correct declarative use conditions. In the case of our paradigm *a priori* claims about abstract domains, representationists hold that these conditions obtain, if they do, within our heads. Supposing that there is only a finite number of these conditions, the representationist construal seems to entail that there can be only a finite number of semantically distinguishable purportedly *a priori* beliefs about abstract domains.

In response to this objection, a representationist may recall that claims with identical truth conditions may differ in semantic contents, because declarative use conditions are in general more fine-grained than truth conditions. As it was noted in chapter 4, the truth conditions of a truth-apt representation can be identified with the declarative use conditions of this complex symbol in an unembedded state. Clearly, the fact that two representations can be applied under the same circumstances in an unembedded state does not entail that they preserve their equivalence in larger representational contexts as well.[32] In view of this relation between truth conditions and semantic contents, a representationist may argue that her finitist conception of the domain of the relevant truth conditions is fully compatible with the idea that there are infinitely many semantically different, purportedly *a priori* claims about abstract domains.

---

[32] See fn. 15 in chapter 4.

A second objection to the representationist's view that the obtaining of finitely many conditions in our heads determines the truth value of all (determinately true or determinately false) paradigm *a priori* claims focuses on the representationist assumption that the conditions in question are natural relations among some representational constituents of the relevant claims. This time, opponents may argue that there seem to be infinitely many semantically basic constituents that we can combine into paradigm *a priori* claims about abstract domains. Moreover, it seems that we can conceive various infinite series of these truth-apt representations, in which no two members can possess the same analytically related constituents. If the above representationist assumption is true, then the existence of such series implies that our paradigm *a priori* claims must have infinitely many truth conditions.

Consider, for instance, the series of truth-apt arithmetic representations mentioned above. According to the representationist construal, the truth conditions of the idea *that one plus one equals two* are some links among the semantically basic constituents of this representation in our heads, and the same is supposed to hold of the truth conditions of every other member of the series. Now, it appears that no two members have exactly the same analytically related constituents. This implies that none of them can, according to the suggested representationist assumption, possess exactly the same truth conditions. Since the series consists of infinitely many arithmetic claims, it seems that a representationist cannot consistently hold that the truth value of these members is determined by the obtaining of a finite number of conditions.

In response, a representationist may query two premises in the previous line of thought. First, she may observe that in the actual world we can possess only a finite number of representations in our heads. Consequently, in the metaphysically thick sense, there are not infinitely many (mental or physical) representations in the world. The conceivability of the above infinite series of arithmetic thoughts requires, in a representationist framework, merely the conceivability, not the

actual existence, of an infinite number of distinct representational truth conditions, something that clearly obtains in the actual world.

The second representationist reaction concerns the opponent's assumption that no two members of the arithmetical series under scrutiny can have exactly the same analytically related constituents. *Prima facie* the assumption seems true. If we focus on the mental or physical symbols applied in these representations, then we find that each member of the series differs in at least one crucial constituent from any others. The first doubt concerning this claim emerges when we recognise that each member of the series is an instance of the same general arithmetical thought, the idea *that the sum of the number one and an arbitrary natural number equals the successor of that number*. Since we can establish the truth value of this general thought with our finite cognitive capacities, a representationist may reasonably assume that the truth conditions to be detected in this case obtain within the actual finite domain of representations in our heads. Further, since the truth of this general thought guarantees the truth of its infinitely many instances, she can also assume that the former conditions constitute the truth conditions of these instances as well.[33]

[33] One may object that the identity of the truth conditions of a general analytic thought with those of its instances would entail the logical equivalence of these representations, which is incompatible with the logical laws that govern universal quantification. My view is that in the case of analytic generalisations (i.e. generalisations that are analytically true), the equivalence holds whenever the analytic link guaranteeing the truth of the general idea obtains among its constituents other than the universal quantifier. For instance, I hold that the generalisation *that bachelors are men* is logically equivalent with, rather than merely entails, its own instances (e.g. the idea *that the tallest bachelor is a man*, the idea *that the second tallest bachelor is a man* etc.). Due to the analytic character of these truths, indeed, we cannot conceive any of them holding without the holding of the others. Of course, the representations are semantically different, because they do not share all their constituents. Nevertheless, the presence of their specific components does not affect their inferential relations, since these relations are fixed by the analytic link between the relevant common components of these representations. (Thanks to Daniel Isaacson for calling my attention to this objection.)

The opponent may insist that this point illuminates merely the inadequacy of the representationist assumption that the truth conditions of our paradigm *a priori* claims are natural relations among some constituents of these claims. In other terms, the opponent may grant that each member of the above series is made true by the obtaining of the same representational conditions, insist that no two members have exactly the same analytically related constituents, and therefore conclude that the suggested representationist characterisation of the particular truth conditions of our paradigm *a priori* claims cannot be adequate. In order to save this part of her conception, a representationist must show that our paradigm *a priori* thoughts may have analytically related common constituents even if no such constituents can be identified on the level of those mental or physical symbols that appear in the actual formulation of these thoughts.

The crucial observation that a representationist may invoke at this point is that the way we conceive an infinite set of truth-apt representations which has no two members with the same analytically related symbolic constituents is always by conceiving an infinite number of instances of a finite number of more general thoughts. The instances of a general thought, however, usually possess as semantic constituents most constituents of the general thought. Of course, at least one of these constituents must occur in the instances in infinitely many specified forms. This is what distinguishes the instances from the general thought. Nevertheless, these specifications do not annul the presence of the specified constituent in the particular instances. It merely terminates the explicit symbolic representation of this common semantic component in those thoughts. It is due to this fact that we cannot completely identify the common constituents of a certain multiplicity of thoughts merely by comparing those mental symbols that occur in the actual formulation of these representations.

In the case of our infinite arithmetic series, for instance, a representationist may argue that, despite appearances, most semantic constituents of the general thought *that the sum of the number one and an arbitrary natural number equals the successor of that*

number (viz. our concepts of addition, number one, natural number, successor, sameness and equality) appear in each of the conceivable instances of this thought, and that, corresponding to the suggested representationist assumption, it is the actual obtaining of a finite number of suitable natural relations among some of these common constituents that guarantees the non-referential truth of the relevant infinitely many arithmetic beliefs.

The illusion that there are no such common constituents emerges merely because some of these elements have no explicit symbolic representation in the chosen conceptualisations of the purported arithmetic contents. Our notion of natural number, for instance, does not appear explicitly in the thought *that one plus three equals four*. Note, however, that our concept of number three is a symbolic shortcut for the symbolically more complex representation of the successor of the successor of the only natural number that is not a successor. The two representations are associated with the same arithmetic content. Accordingly, the idea *that one plus three equals four* must involve, at least as an implicit semantic constituent, our notion of natural number as well. The only constituent of the above general thought that is clearly absent in its particular instances is the idea of universality (the notion of everything-of-a-certain-kind or everything *simpliciter*). This component is replaced by various individuating concepts, which turn our general notion of natural number into ideas of particular natural numbers. As we saw, our notion of number three can be regarded as a specification of our general concept of natural number.[34] The most important observation from the representationist's perspective is that the semantic relations of the applied individuating concepts need not be constitutive of the truth conditions of the resulting infinitely many arithmetic claims.

Summing up, an advocate of the suggested representationist construal can explain the conceivability of an infinite number of

[34] The claim that certain concepts can be regarded as specifications of some others is meant to have no implication concerning the actual genesis of the related symbolic elements. New concepts can be developed from earlier acquired ones by specification as well as by abstraction.

semantically non-equivalent paradigm *a priori* claims even if her account implies that the truth value of these claims is determined by the obtaining of a finite number of conditions in our heads. Following her standard referentialist opponent, a representationist may argue that the previous phenomenon is a consequence of two semantic facts. The first is that claims with different intended referents have different semantic contents, while the second that the intended referential domain of our paradigm *a priori* claims is infinite in character. The finitist implications of the construal concerning the domain of relevant truth conditions turned out to be compatible with this account on three different considerations: first, the representationist's general use-theoretic notion of truth conditions implies that semantic contents are more fine-grained than truth conditions; second, the finitist implications emerge only in so far as the construal is meant to characterise the truth conditions of our paradigm *a priori* representations in the actual world, where (in the metaphysically thick sense) there is only a finite number of such representations; and third, even those infinitely many paradigm *a priori* representations that we can at least conceive (as opposed to separately entertain) in this world are instances (and logical consequences) of a finite number of more general representations, whose truth value can arguably be determined by the obtaining of a finite number of conditions in our heads.

Apriority of Evidence

Beyond the above general features, whose proper explanation is arguably a minimal adequacy condition for a construal of virtually any kind of truth, in chapter 2 I collected four further characteristics that constitute an *explanandum* for a suitable account of the paradigms of *a priori* truth in particular. The first of these characteristics was that our knowledge of these paradigms is based on *a priori* evidence, or, in other terms, that it can be justified without reliance on experience.

Of course, in a sense this explanatory task can be trivially accomplished by any construal of the subject that supports an

acceptable account of our knowledge within the relevant discourses. The fact that we regard certain sorts of thoughts as the paradigms of *a priori* representations implies that our knowledge of their truth value must be also *a priori* in character. The situation is, however, far from being so simple, since our notion of apriority is not entirely void of empirical content. In fact, as we saw, it is understood in terms of independence of experience. Now, even if we have no definite concept of experience, nevertheless we understand that, for instance, any knowledge of the external part of the spatiotemporal world is based on experience. If this is so, however, then a construal which locates the truth conditions of the examined representations in the external part of the natural world cannot be consistently regarded as a construal of the paradigms of *a priori* truth.

In chapter 6, I acknowledged that in possession of an acceptable account of knowledge acquisition or reliable belief formation of abstract domains, the advocates of the standard platonist construal of those paradigms could meet this explanatory requirement. If they stood in the contact theorist camp, they could simply observe that our access to the alleged platonic truth conditions cannot be empirical, since it requires a specific epistemic capacity (different from the use of our external or internal senses) which connects our minds with entities outside the natural world. In contrast, if they opted for a no-contact epistemology, they could argue that our knowledge of abstract domains cannot be empirical, since it requires no contact whatsoever between our minds and the relevant platonic truth conditions.

Now, can a proponent of the suggested representationist construal of the truth conditions of our purportedly *a priori* claims about abstract domains also consistently assume that her account of knowledge acquisition or reliable belief formation within the relevant discourses is apriorist in character? Well, at the beginning of the second section of this chapter, it was noted that the main motive behind the non-referentialist's adoption of the suggested representationist version of naturalism concerning the above

conditions is that this construal seems to observe our (slightly indefinite) idea of experience, which implies that all knowledge of the external part of the natural world and almost all knowledge of our own bodily states is empirical.

Of course, some construals of the *a priori* / *a posteriori* distinction would be incompatible with the representationist position. For instance, if one supposed that any knowledge of the spatiotemporal world is by definition empirical, then one could not consistently maintain that the truth conditions of our paradigm *a priori* claims obtain in the domain of representations within our heads. As it was noted in the first section of chapter 1, however, the conflict between the representationist conception and such an understanding is merely terminological. The substantive representationist point is that our claims about abstract domains (and our analytic claims in general) are true or false in virtue of the obtaining of some representational conditions in our heads, and our knowledge of these representational facts constitute a natural kind, which can be contrasted with our knowledge of other parts of the natural world. The representationist classifies this kind as *a priori*, because it is traditionally regarded as the paradigm of this type knowledge, but if her opponent wants to use this term with theoretical connotations that make radical empiricism a plausible position in epistemology, she can accept this alternative convention as well, and merely insist on the previous substantive part of her doctrine.

Reconciled to the representationist terminology, one may still wonder what a representationist could tell us about the essential feature of what she calls *a priori* knowledge. The simplest representationist response to this question is that the desired feature is exactly that the facts detected by this type of knowledge are representational, and they obtain (among others) within the thinker's own head. Our knowledge of them is *a priori*, because it can be acquired before getting any (empirical) information of what obtains in the rest of the actual world. Beyond this response, a representationist may add that the exact specification of the nature of *a priori* knowledge and justification is the task of our empirical sciences, and it must emerge from a painstaking

empirical inquiry into the work of human brains in the course of the relevant types of knowledge acquisition.

In possession of these responses, we can conclude that an advocate of the suggested representationist construal of truth in the semantics of our purportedly *a priori* discourses about abstract domains can adequately account for the fact that our knowledge within these discourses is based on *a priori* evidence by maintaining that a piece of evidence, or the justification that it provides for a belief that is based on its recognition, is *a priori* if and only if it is generated by a reliable causal mechanism within human heads that conveys information of the obtaining or absence of truth conditions within a subject's system of representation to her knowing mind.

Necessity of (the Paradigms of) A Priori *Truth*

The second specific *explanandum* for an account of the paradigms of *a priori* truth that appeared on our list in chapter 2 was that the relation of this semantic value (or its opposite) to its bearers is necessary in character. My claim that there are three chairs in front of me could be false even if it is actually true. In contrast, my idea that there are three prime numbers between 70 and 80 is not merely true, but it is necessarily true.

In chapter 6, I argued that an advocate of the standard platonist construal of the paradigms of *a priori* truth can provide a relatively simple explanation of this phenomenon. Namely, she can maintain that the necessity of the relevant truths (or falsities) is due to the necessity of the obtaining (or absence) of those platonic truth conditions that our paradigm *a priori* claims purport to be about.[35] The proponents of the suggested representationist

[35] In absence of a reasonable account of how we could actually detect the modal character of the obtaining of various truth conditions, however, this referentialist conception of the factual ground of modality creates a serious explanatory problem in the epistemology of our cognitive discourses in general. Note that Lewis's alternative strategy, which explains the modal status of necessary truths and falsities by reference to the simple (non-modal) obtaining of referential conditions in realistically construed non-actual worlds results in similar

construal cannot follow this route, since on their view the relevant truth conditions obtain (or not) contingently in some segments of the actual spatiotemporal world. But is there an alternative account of this *explanandum*? Can the simple (i.e. non-necessary) obtaining of some natural conditions in the actual world guarantee the necessary truth (or necessary falsity) of our paradigm *a priori* beliefs?

My claim is that the proper response to this question is also positive. To see why, we should merely recognise that the necessity of the truth (or falsity) of a certain thought can be also a consequence of the semantic fact that the obtaining (or absence) of the relevant truth conditions is constitutive of the semantic content of this thought. In section 2, I argued that the suggested representationist construal of the paradigms of *a priori* truth is compatible with such an account of this *explanandum*. After all, the analytic relations obtaining among our representations, the suggested representational truth conditions of our paradigm *a priori* claims, can be arguably regarded as constitutive of the semantic content of the related symbols as well as the claims composed of them. According to this understanding, the claim that there are three prime numbers between 70 and 80 cannot be false, because in absence of the (actually obtaining) non-referential truth conditions of this representation the declarative use conditions of at least some of its constituents would also differ from the actual ones.

Putting it briefly, the representationist construal advocated here can explain the observable contrast between contingent and necessary truths and falsities by invoking the difference between two kinds of truth conditions in the actual world: those whose simple obtaining (or absence) merely determines the truth value of some claims, and those whose simple obtaining (or absence),

epistemological difficulties in so far as it leaves no room for a viable account of how we could actually detect what obtains in the suggested realistically understood but causally closed alternative possible worlds. Lewis (1986).

beyond determining this value, also contributes to the semantic content of those claims.[36]

Applicability of A Priori *Knowledge in the Empirical Sciences*

The third specific feature to be explained by an appropriate account of the paradigms of *a priori* truth on our list in chapter 2 was the applicability of our knowledge of these truths in our empirical sciences (as well as in our ordinary theorising) of the natural world. To take the simplest example, if we learned from experience that Peter put two apples on an empty table, and then John added three others to these two, then we can know without counting the group again, merely relying on our *a priori* knowledge that two plus three equals five, that there must be five apples lying on the table now.[37]

[36] Note that this account can explain the existing ambiguity in the modal character of those truths and falsities as well, whose subject matter is the actual obtaining (or absence) of the invoked representational truth conditions in our heads. Consider, for instance, the claim that the representational truth conditions of the idea that there are three prime numbers between 70 and 80 obtain in my head. According to the suggested representationist construal, this claim is presumably true. But is this truth necessary or contingent? Well, we may easily conceive a world in which the representational conditions in question do not obtain. Note, however, that in such a world those missing conditions would not constitute the truth conditions of this mathematical claim. In fact, in such a world the claim could not be made at all. (In a purely syntactical sense, the symbols applied could exist there as well, but the semantic contents associated with these entities would be different from the actual ones.) What this means is that there is a sense in which the above truth is necessary, and an other in which it is contingent in character. We can easily conceive a world in which the truth conditions of an actually thinkable analytic truth would not obtain, but we cannot conceive a world in which this content could be entertained in thought without being true. (Thanks to Hartry Field for turning my attention to this consequence of the representationist construal advocated in this work.)

[37] The history of science provides more complex illustrations of this phenomenon. The greatest insights behind the early modern emergence and later development of our current scientific conception of the world are famously associated with the invention of illuminative thought experiments, *a priori* reasonings that seem to provide new synthetic knowledge of some laws in nature. The best known examples of such applications of *a priori* (logical and

The crucial question to be answered in order to meet this explanatory requirement is what sort of relation between the truth conditions of the relevant claims makes it the case that our purportedly *a priori* knowledge of abstract domains can help us discover what is necessarily true about the spatiotemporal world. In chapter 6, I observed that, in view of the stipulated metaphysical gap between the intended platonic and natural realms, an advocate of the standard platonist construal of the truth conditions of our paradigm *a priori* beliefs seems to have no satisfactory response to this question, and, consequently, her theory would remain inadequate even if it proved to be compatible with an acceptable account of our knowledge of the relevant abstract domains.[38]

In contrast, an advocate of the suggested representationist framework can provide a relatively simple answer to the previous question. As we saw, in this framework the truth conditions of our paradigm *a priori* claims about abstract domains are relational states of affairs obtaining among our representations that we

mathematical) knowledge in natural science can be found in Stevin (1605), Galilei (1632), Galilei (1638), Newton (1687), Maxwell (1871), Einstein (1905), Einstein, Podolski and Rosen (1935), and Schrödinger (1935). The most influential classical theories of the nature of thought experiments include Mach (1897), Duhem (1906), Koyré (1939), Popper (1959), and Kuhn (1977b). For more recent literature on this topic, see Brown (1991), Horowitz and Massey (1991), Miščević (1992), Sorensen (1992), Hull, Forbes and Okruhlik (1992), Haggqvist (1996), McAllister (1996), Norton (1996), Gendler (2000), and Hitchcock (2004).

[38] One may try to defend the platonist construal in this regard by first maintaining that the alleged metaphysical gap does not exclude the existence of certain isomorphisms between what obtains in the contrasted platonic and natural realms, and then declaring that the existence of these isomorphisms would perfectly well explain the observable applicability of our purportedly *a priori* knowledge of platonic domains in our empirical sciences of the natural world. The problem with this response is that a platonist has no independent evidence of the obtaining of this enormous coincidence between the facts of the two realms. Her evidence for her pure logical and mathematical beliefs cannot be taken as such, because, as I attempted to show in chapter 6, we have no reason for assuming that our epistemic grounds in these paradigm *a priori* disciplines provide reliable information of what actually obtains in a platonic realm. To this measure, the platonist's failure in the two explanatory respects mentioned here is still related to each other.

develop in our head in the course of our cognitive engagement with our natural environment. In section 2, we saw also what a representationist can in principle say about the emergence of these conditions and their envisaged semantic links to the above claims. Most importantly, the account in question assumed that the conditions are analytic relations among our representations of abstract entities that we developed by abstraction in the third specified sense of the term in our head from some earlier acquired ones of various aspects of the natural world. So, the question to be answered by a representationist boils down to this: what makes it the case that our knowledge of the obtaining or absence of these analytic links can help us discover what is (necessarily) the case in the spatiotemporal world, or rather, in non-epistemic terms, why do these analytic relations, the obtaining truth conditions of our true representations of abstract domains, reflect so well what actually obtains in the spatiotemporal world?

The representationist answer to this question is the following. If the account advanced in section 2 is correct, then the truth conditions of our paradigm *a priori* claims about abstract domains are identical with the truth conditions of their various "applications" to the natural world. The truth conditions of our arithmetic claim that two plus three equals five, for instance, are identical (among others) with those of our "applied arithmetic" claim that Peter's two and John's three apples must constitute a group of five apples on the table. The identity in question is ensured by the way we develop our ideas of the relevant abstract entities from our ideas of some properties that may obtain, and thus be observed, in the spatiotemporal world.

An important consequence of this semantic fact is that our *a priori* acquired logical and mathematical knowledge cannot fail to help us discover or clarify, in principle, what (necessarily) obtains in the natural world, unless the relevant "applications" fail to be true about this world too. The latter circumstance, however, could obtain only if the analytic relations determining the truth value of these "applications" were introduced by us among symbols with unrelated (fine-grained) referential contents (i.e.

unrelated synthetic declarative use conditions).[39] A careless act like this would turn (the relevant part of) our system of representation inconsistent. For instance, if we stipulated that water is whatever has the chemical structure CO_2, then the claim that water is partly constituted of carbon would become analytically (and thus necessarily) true. On the other hand, unless we radically changed the actual referential content of our notion of water (or that of our notion of carbon dioxide), the same claim would be clearly false about the actual world. Note, however, that an analytically true representation cannot be consistently false about the actual world (if the referents of its constituents exist in that world). Apparently, maintaining consistency requires that we create analytic links between symbols only if the referential contents of these entities stand in a suitable (part-whole or identity) relation with each other.

In section 2, we saw that there are two major types of situation in which we actually introduce an analytic link between two elements of our system of representation. Sometimes, the link emerges when we define a new concept to represent a certain aspect of the world in a simpler way than its *definiens* did before. In the actual world, this is how we develop our concept of bachelor. In such cases, the referential contents of the related symbols cannot fail to stand in the required relation with each other. The synthetic declarative use conditions of our notion of bachelor, for instance, have presumably included from the outset as a proper part those of our concept of man.

At other times, we seem to introduce the link among symbols with independently established referential contents. This is what we seemed to do when we reduced our idea of water to our chemical concept of H_2O, and maybe also when we first defined our notion of being a double in terms of our notion of being a single and our idea of addition. In these cases, the required relation of the relevant referential contents does not

[39] I insert the term 'fine-grained' here to recall that the notion of referential content that I rely on in this part is not the coarse-grained Fregean one, which was shown to lead us to the collapsing conclusion of the slingshot in chapter 4.

seem to be automatically guaranteed. As we observed in section 2, however, the link that we actually introduce when we "reduce" a non-analytically acquired notion to some others is (1) strictly speaking not among these semantically independent representations, and (2) even loosely speaking among these entities only if the intended referential contents in question sufficiently overlap (as in the case of "ordinary" water and H_2O) or coincide (as in the case of being a single together with another single, on the one hand, and being a double, on the other). In fact, what happens in the case of these theoretical reductions is that we replace the relevant semantically independent notions with some others, which are analytically connected and possess suitably related referential contents that are also either identical or just slightly different from those of the respective antecedents.[40]

Due to this aspect of our concept-developing practice, the referential contents of our analytically linked notions never fail to meet the semantical requirement specified above. Accordingly, our analytic claims never fail to be necessarily true about their intended referential domains.[41] This implies that the analytic "applications" of our paradigm *a priori* claims of abstract domains to the natural world are also always true about this world, which in turn explains why our purportedly *a priori* knowledge of the former domains can help us discover what is necessarily true about the latter realm.

[40] It is this identity or sufficient overlap between the initial and modified referential contents that "legitimises" or motivates the strictly speaking rather confusing linguistic practice that we express the newly introduced notions with the same physical symbols as their analytically unrelated antecedent (e.g. we keep the term 'water' to speak about the newly conceptualised stuff that is essentially constituted of H_2O).

[41] They may, of course, fail to apply to the world if applicability is understood in the stronger sense requiring the actual obtaining of the synthetic declarative use conditions (or the existence of the intended referents) of all basic constituents of the relevant analytic claims. Note, however, that our current *explanandum* is not the applicability of our *a priori* knowledge of abstract domains in this strong sense of the term, but its applicability when the referents of the symbolic elements of its "applications" do exist in the natural world.

Summing up, an advocate of the suggested representationist construal of the paradigms of *a priori* truth can explain the applicability of our knowledge of these truths in our empirical sciences of the natural world by invoking two facts about the truth conditions of the relevant *a priori* claims: first, that they are identical with those of the conceivable "applications" of these representations to the natural world; second, that their obtaining is a consequence of our stipulative work, whose nature guarantees that the "applications" in question are never false about the natural world.

Abstractness and Infinity of Intended Domains

The final specific characteristic that occurred on our list in chapter 2 as an *explanandum* for an account of the paradigms of *a priori* truth was that the intended referential domains of the bearers of these paradigms are abstract and often infinite in character.[42] The question to be answered in this case is how the bearers of these paradigms can determinately refer to entities within such domains.

Earlier I argued that an advocate of the standard referentialist construal of the subject can provide a suitable answer to this question and successfully explain our capacity to think of or speak about abstract and infinite domains. Nonetheless, as we saw, the referentialist conception cannot be regarded as adequate, since it can explain the objectivity of its subject matter only if it involves a platonist understanding of this entity, a construal that supports no reasonable account of how we can acquire this type of knowledge, and why we can successfully apply it in our empirical sciences.

Fortunately, the adoption of an alternative non-referentialist conception of the same subject affects merely our idea of the

[42] Note that the possible infinity of referential domains, in itself, is not a specific *explanandum* for an account of the paradigms of *a priori* truth. After all, our capacity to develop ideas of infinite universes is manifest in our broadly physicalistic discourses as well.

relation of the truth conditions of the relevant bearers to their intended referents. What an advocate of the standard construal took to be plain identity, a proponent of the suggested representationist construal regards as a less intimate relation. In particular, she maintains that the truth conditions of our paradigm *a priori* representations are identical with some analytic relations that may obtain among the basic constituents of these contentful entities.

In accounting for the current *explanandum*, the phenomenon that our paradigm *a priori* beliefs can determinately refer to entities in abstract and often infinite domains, a representationist can simply follow her referentialist opponent, and argue that the referential content of the basic constituents of these beliefs is fixed by those cognitive mechanisms that underlie the actual development of these representations in human heads. In so far as at the bottom they are all developed from earlier acquired notions (with determinate spatiotemporal referents) by abstraction in the third specified sense of the term, their fixed referential content will be, quite understandably, non-spatiotemporal in character.[43] On the other hand, since in possession of a limited number of basic concepts (again, with determinate referents) of a certain domain we can recursively develop, at least in principle, an unlimited number of new (determinately referring) representations of previously not represented entities of that domain, the possible infinity of what we can think of or speak about in the relevant *a priori* discourses should not strike us as a surprise either.

With this result, I have finished the examination of the adequacy of the suggested referentialist construal of the paradigms of *a priori* truth. If the accounts advanced in this section are correct, then we can conclude that in contrast to its referentialist alternatives discussed in chapters 4, 5 and 6, the non-referentialist conception proposed in this work can satisfy all major

[43] In chapter 5, I argued that determinate reference to spatiotemporal entities can be explained within a "metaphysical realist" perspective as well.

explanatory requirements set for an account of this subject in section 2, and thus qualifies as a minimally adequate conception of what we took to be the paradigms of *a priori* truth. In the last section of this chapter, I shall return to the broader perspective of our current investigation and argue that the representationist conception advocated here can be taken as a suitable defining characterisation of the nature of *a priori* truth in general, and in possession of this definition, we can derive a corresponding characterisation of the nature of *a priori* knowledge, justification and evidence as well.

4. A Representationist Construal of A Priori *Knowledge and Truth*

I started this work with the announcement that I would argue for a particular naturalistic characterisation of *a priori* truth and knowledge, a conception that preserves the categorical distinction between *a priori* and empirical beliefs, largely observes the received application of the contrasted terms, and, together with our best empirical theories of the world and human cognition, explains the most important characteristics of our purportedly *a priori* claims and beliefs. In order to ensure the greatest initial agreement concerning the premises of our investigation, I proposed to start with the examination of truth within what we usually take to be the paradigms of *a priori* discourses (i.e. within pure logic and mathematics), develop an adequate account of this subject, and then consider whether the resulting conception would suggest us a suitable real definition (i.e. a characterisation of the nature) of *a priori* truth and knowledge in general.[44]

With the previous section of this chapter, the first part of the above plan has been completed. After presenting a list of the most important conditions of adequacy that the envisaged

[44] The project thus conceived corresponds to standard reductive analyses in the empirical sciences, such as the earlier discussed scientific reduction of (our concept of) water to (that of) H_2O.

account of the paradigms of *a priori* truth must (arguably without exception) satisfy, I argued for two major tenets concerning the subject under scrutiny. First, I attempted to establish that the truth conditions of our paradigm *a priori* beliefs cannot be understood along the received referentialist lines. Second, I argued that a particular representationist construal of these conditions can pass the above test, and thus qualify as a minimally adequate characterisation of the paradigms of *a priori* truth. In possession of these results, we can turn now to the second part of our project, and consider whether any of these findings could be invoked in a minimally adequate characterisation of the nature of *a priori* truth, knowledge, evidence and justification in general.[45]

First, it may be worth noting that the two features mentioned in the previous paragraph are independent of each other. In other terms, the fact that the truth conditions of certain beliefs are representational does not entail that they are non-referential, and *vice versa*. As it was mentioned earlier, the truth conditions of our beliefs *about* the obtaining of the alleged representational truth conditions of our paradigm *a priori* claims are arguably also representational but referential. On the other hand, one may argue that the most plausible construal of the truth conditions of our truth-apt normative claims (e.g. in ethics or epistemology) is non-referentialist but not representationist in character. So, what would follow concerning the extension of our notion of *a priori* truth, if we invoked either or both of these features in the envisaged characterisation of the nature of this property?

[45] As it was mentioned in chapter 1, in this work apriority is meant to be a property that is primarily attributed to justifications. Derivatively, however, it can be also attributed to pieces of knowledge, beliefs, propositions, judgements, sentences, utterances and truths due to the relevant epistemological features of these entities. Here I suppose that an *a priori* truth is a truth that can be known *a priori*, and a truth can be known *a priori* if and only if our belief in it can be justified *a priori*, without reliance on experience. For more on this topic, see section 1 in chapter 1.

Well, consider first what would happen if we stipulated that both non-referentiality and representationality are essential for certain truths to qualify as knowable *a priori*. Such a construal would entail that neither our claims about the obtaining of the alleged representational truth conditions of our paradigm *a priori* beliefs, nor our claims about the occurrence of causally inert normative values in the spatiotemporal world, could be taken as *a priori* in character. On the other hand, as was argued in section 2, our standard analytic claims about the spatiotemporal world do possess the above two characteristics, and therefore, according to this understanding, they would also fall into the category of (truth-apt) *a priori* representations.

As a second option, we could stipulate also that the essential feature of an *a priori* knowable truth is that it is non-referential. This would render our standard analytic claims as well as, arguably, our claims about the occurrence of causally inert normative values in the spatiotemporal world *a priori*, while our claims about the obtaining of the alleged representational truth conditions of our paradigm *a priori* beliefs empirical in character.

Finally, we could say that a certain truth is *a priori* if and only if it is representational. This would imply that beyond our standard analytic beliefs our claims about the obtaining of the representational truth conditions of these beliefs are *a priori* too.[46] On the other hand, this understanding would be incompatible with the idea that our knowledge of the occurrence of causally inert normative values in the spatiotemporal world, if possible at all, is *a priori* as well.[47]

[46] Whether or not we acknowledge that our claims about the obtaining of these representational conditions can be analytic in character, it is relatively easy to see that they can be taken as synthetic claims about a certain segment of the actual spatiotemporal world. To say that we can acquire *a priori* knowledge of the obtaining of these conditions is nothing else than to recognise the existence of synthetic *a priori* claims about the actual world.

[47] A further notable consequence of all these construals would be that our synthetic thoughts about abstract domains could not be regarded as *a priori* either. Of course, this does not mean that we must classify them as empirical. This would follow only if every thought would be either *a priori* or empirical. But why should we assume this? After all, was not our earlier conclusion of these truths that they cannot be known by us anyhow?

Note that neither of these definitions would preserve the purely epistemological character of our distinctions between *a priori* and empirical truth, knowledge, evidence and justification: while the contrast between referential and non-referential conditions is clearly a semantic one, that between representational and non-representational conditions is one (with semantic connotations) within our actual ontology. So, whichever option we choose, our notion of apriority will lose its traditional epistemological significance.[48]

Nonetheless, I believe that our notion of non-referentiality and notion of representationality are not equally appropriate candidates for becoming an *analysans* of our notion of *a priori* knowledge and truth. In so far as we regard apriority as a property that primarily characterises ways of justifications, the semantical question whether or not certain truth-apt representations are *about* their actual truth conditions seems largely irrelevant. The fact, for instance, that our beliefs about the obtaining of the (allegedly representational) truth conditions of our (pure) logical and mathematical claims are about their own truth conditions does not seem to affect at all the way in which we actually justify these beliefs. In contrast, the envisaged location of these conditions within the spatiotemporal world, and especially the relation of this location to that of knowing minds seems much more significant from the perspective of our current investigation.

For these reasons, I believe that from among the three analytic proposals specified above the intuitively best strategy for us is to opt for the third (i.e. to stipulate that a certain truth is *a priori* if and only if it consists in the obtaining of some representational conditions in a subject's head). As we observed, this understanding implies that our knowledge of causally inert

[48] As it was mentioned in chapter 1, a contrast is supposed to be purely epistemological if it can be characterised by reference to epistemologically significant properties, such as fallibility, transparency or fundamentality, alone.

normative properties or values must be either analytic or empirical. A proper assessment of this implication is clearly beyond the scope of the current work. Here and now it suffices to say that the observable characteristics of our linguistic and cognitive practice in normative discourses, such as ethics and epistemology, do not seem to stand in obvious conflict with the above assumption.

In possession of this representationist account of *a priori* truth, we can formulate now the corresponding representationist construals of *a priori* knowledge, evidence and justification. According to these construals, a piece of knowledge is *a priori* if and only if it is justified *a priori*, by invoking *a priori* evidence, and a piece of evidence, or the justification that it provides for a belief that is based on its recognition, is *a priori* if and only if it is generated by a reliable causal mechanism within human heads that conveys information of the obtaining or absence of truth conditions within a subject's system of representation to her knowing mind.[49]

Despite their largely "ontological" (or maybe semantical) character, these construals nevertheless preserve an important connotation of the traditional epistemological notion of apriority. Namely, they preserve the idea that, in a sense, *a priori* knowledge is prior to any further knowledge of the actual world. The priority in question was already briefly indicated in section 2. There we observed that our synthetic knowledge acquisition of the world is realised by two consecutive epistemic mechanisms. First, we develop an access to our representations, and recall those circumstances under which they can be correctly applied in the particular declarative representational context under scrutiny. Second, we discover whether or not the recalled circumstances actually obtain in the represented world. By adopting the suggested representationist construal of apriority, we can say now that the first sort of mechanism is what underlies our *a priori* knowledge acquisition of various domains. Accordingly, we may

[49] Note that the resulting notions enable us to distinguish in a relatively sharp way between *a priori* and introspective knowledge, justification and evidence as well.

conclude that on a representationist understanding of apriority, some *a priori* knowledge is constitutive of any propositional knowledge of the actual world.

With this conclusion, we can close now the second part of our investigation as well. After arguing that truth within our paradigm *a priori* discourses is best understood in a non-referentialist and representationist manner, in this section I suggested that by reference to the latter characteristic we can provide a minimally adequate specification of the nature of *a priori* truth, knowledge, justification and evidence in general.[50]

Summary

In this chapter, I attempted to show that, contrary to its referentialist alternatives, a specific naturalist version of non-referentialist realism about truth in the semantics of our paradigm *a priori* discourses satisfies all major adequacy conditions set for such a theory in chapter 2. I called this version a representationist account of the relevant truths, since it assumes that the truth conditions of our purportedly *a priori* claims about abstract domains obtain (if they do) in the realm of representations within our heads, rather than in the domain of the represented abstract states of affairs.

In section 1, I argued that, in view of our actual cognitive and linguistic practice, we have no reason to suppose that a non-referentialist construal of certain truths is conceptually objectionable.

In section 2, I started the demonstration of the adequacy of the representationist construal by providing an (in principle empirically confirmable) ontologically naturalist account of how our purportedly *a priori* claims about abstract domains may

[50] As it was emphasised in chapter 2, one can challenge the adequacy of the representationist construal advocated here by identifying some further characteristic of *a priori* truth and knowledge that the account under scrutiny can explain neither in itself, nor as part of a larger theory of the world.

acquire their determinate semantic relations to their (non-referential) representational truth conditions, on the one hand, and their non-spatiotemporal intended referents, on the other. In possession of this account, I concluded that the construal under scrutiny can satisfy the third explanatory requirement set for such a theory in chapter 2.

In section 3, I confronted the proposed construal with the ten other explanatory requirements collected in chapter 2, and argued that this version of non-referentialism is compatible with an appropriate account of all those phenomena whose joint explanation we agreed upon to regard as a minimal condition of adequacy for a conception of the relevant truths, and thus it constitutes also a suitable response to Benacerraf's original or modified and generalised dilemma in the philosophy of our purportedly *a priori* discourses about abstract domains.

Having shown that truth within our paradigm *a priori* discourses is best understood in a non-referentialist and representationalist way, finally, in section 4, I returned to the initial broader perspective of our investigation, and argued that by reference to the latter characteristic we can provide a minimally adequate specification of the nature of *a priori* truth, knowledge, justification and evidence in general.

CONCLUSIONS

In this work, I have argued for a naturalistic construal of *a priori* truth and knowledge. My reasoning has consisted of three consecutive steps. Having elucidated the central notions and the major methodological assumptions of this inquiry, first, I argued that in the semantics of our paradigm *a priori* discourses (i.e. in the semantics of pure logic and mathematics) truth cannot be adequately interpreted along the standard referentialist line, in other words, assuming that the truth conditions of the relevant beliefs are to be specified in terms of those entities that these beliefs purport to be about ("non-referentialism about the paradigms of *a priori* truth"). Second, I argued that the truth conditions (as opposed to the intended referents) of our paradigm *a priori* beliefs are natural conditions in our heads, or, more exactly, that they are analytic relations among representations that we develop in our heads in the course of our cognitive interaction with our direct natural environment ("representationism about the paradigms of *a priori* truth"). Finally, I argued that we have good reasons to take representationality as an essential, defining trait of apriority, and thus as a necessary characteristic of *a priori* truths in general ("representationism about *a priori* truth, knowledge, justification and evidence").

My case against the standard referentialist construal of the paradigms of *a priori* truth was based on the observation that no conceivable version of the referentialist conception can support an acceptable account of all striking characteristics of these paradigms, and thus none of them can offer a viable solution to Benacerraf's updated and generalised dilemma in the philosophy of the relevant *a priori* disciplines. From among the *prima facie* plausible versions of the referentialist doctrine, the deflationist and anti-realist forms were shown to be incompatible with a suitable explanation of the objectivity of the examined truths, while the realist (i.e. platonist) form proved to support no

reasonable account of how we can acquire knowledge or reliable beliefs of the intended abstract domains.

My case for the suggested representationist construal of the same paradigms was based also on explanatory considerations. In particular, I argued that an advocate of this construal can explain all striking characteristics of the subject, and thus demonstrate that the conception amounts to a suitable response to Benacerraf's dilemma and at least a minimally adequate characterisation of the paradigms of *a priori* truth.

Finally, my case for the corresponding representationist construal of apriority in general was based on pragmatic considerations. My primary motive at this point was that by adopting this construal we obtain a relatively clear, informative, and scientifically confirmable idea of the nature and extension of apriority, which largely observes the received application of the term, which together with our best empirical theories of the world and human cognition explains the most important characteristics of our purportedly *a priori* claims, and which, last but not least, preserves an important connotation of the traditional epistemological notion, namely that *a priori* knowledge is, in a sense, prior to any further knowledge of the actual world.

The establishment of the explanatory adequacy of the naturalistic construal just summarised required a number of conceptual and theoretical assumptions concerning apriority, knowledge, truth, realism, objectivity, reference, abstractness, nature, representation, analyticity and semantic content, which have been made with an eye on our actual cognitive and linguistic practice and the most obvious, observable characteristics of the intended phenomena. For the sake of clarity, many of these assumptions were stated explicitly in chapter 1. Here and now, I do not wish to repeat all these statements. Instead, in what follows, I shall briefly review the most important generalisations that occurred among or can be drawn from the conceptual and theoretical claims made in this work and spell out what the collected findings imply concerning some neighbouring issues in the current literature and the relation of armchair philosophy to the deliverances of empirical inquiry.

The first general claim that I wish to emphasise was defended in chapter 3, and concerns the scope of the dilemma that initiated my case against the standard referentialist understanding of the paradigms of *a priori* truth. As has been observed, in its most generic form, the dilemma provides a challenge to the standard referentialist and realist construal of truth in any discourse in which we are supposed to acquire knowledge of conditions whose obtaining or absence can have no causal impact upon our knowing minds. Beyond the intended abstract referents of our beliefs in pure logic and mathematics, genuine (irreducible) normative properties and values, the entities of realistically construed non-actual worlds, and the future constituents of the actual spatiotemporal world are arguably also incapable of exerting causal influence on our present and actual cognitive faculties. If my arguments in chapter 6 are sound, then nothing in our currently available evidential pool can be reasonably taken as a reliable sign of the obtaining or absence of truth conditions involving such (realistically construed) normative, non-actual or future entities. So, if one believes that we can have knowledge of normative, non-actual or future states of affairs, then one must abandon the idea that the truth conditions of these knowledge claims are to be understood in terms of the respective intended referents.

My own view is that in the case of synthetic claims about realistically interpreted future or non-actual states of affairs referentialism about truth can be maintained, because we can acquire no synthetic knowledge of what will obtain in the future or what obtains in a non-actual world. Our synthetic claims about the future may be more or less well supported by our current evidence *if* future tokens of various entity types will occur along the same patterns as they did in the past, but the truth value of these claims will be fixed, and thus become knowable, only by the actual future obtaining or absence of the relevant (referential) truth conditions. Our synthetic claims about non-actual realms are similar to our synthetic claims about platonic entities: we can never reasonably take anything as evidence for the truth or the falsity of these representations. In chapter 7, I argued that the

truth conditions of analytic claims are non-referential and representational independently of what these representations are about. In line with this view, I do not query the possibility of analytic knowledge of the future or of alternative non-actual realms, but I suppose that this knowledge is knowledge of the (current and actual) obtaining or absence of some representational conditions in the subject's head, rather than of those future or non-actual conditions that these analytic claims purport to be about.

Our normative claims or value judgements are typically synthetic, so a representationist response to the dilemma in their case does not seem to be adequate. Note, however, that representationism is not the only conceivable form of non-referentialism. Simon Blackburn's quasi-realist in meta-ethics, for instance, as I observed in chapter 1, is best understood as an anti-realist about intended referents and a realist about truth conditions. This implies that she must be also a non-referentialist about ethical truths. Nonetheless, her construal is presumably not representationist in character. As Blackburn's or Gibbard's projectivist meta-ethics may exemplify, certain truth conditions can be regarded as non-referential, with realistically construed constituents (e.g. affective and conative attitudes) in the evaluating subject that are outside the domain of the subject's acquired representations. Be it as it may, despite the availability of this alternative non-referentialist construal, I believe that the semantic content of our normative or evaluative utterances is much more idiosyncratic (i.e. intersubjectively much less harmonised) and much less stable (i.e. changing considerably through contexts and time) than that of our utterances about abstract, not-actual or future states of affairs. If this belief is correct, then an adequate construal of truth in the case of these normative or evaluative beliefs may prove to be also much more heterogeneous than one in the semantics of other truth-apt representations.

In chapter 5, I argued that the explanatory failure of standard referentialism laid bare by Benacerraf's dilemma is merely epistemological, not semantical as sometimes believed.

What a referentialist cannot suitably explain is, in my view, our capacity to acquire knowledge or reliable beliefs about some problematic domains. Our ability to develop ideas or make claims about particular entities within those domains provides no explanatory puzzle for a referentialist, since a sufficiently sophisticated causal account of reference determination does not require that each mental representation be causally related to its intended referent. As we were able to learn from Putnam's model-theoretic argument, the semantic relation of model-theoretic satisfaction (specified by reference to the usual theoretical and operational constraints) is necessarily indeterminate. This, however, could undermine the possibility of a suitable explanation of determinate reference only if the two relations (i.e. reference and model-theoretic satisfaction) were identical. If we suppose that, beyond the existing (though merely Humean) causal links between our minds and the perceived aspects of the external world, our referential intentions and various concept-generating mental operations (as causally effective natural factors) also contribute to the determination of what our mental and physical symbols refer to, then the best-known arguments against the standard realist idea that we can determinately think of various (epistemically accessible or inaccessible) extra-mental entities prove to be equally inadequate.

A further important result about reference, anticipated in chapter 1 and then defended in chapter 4, is that the symbol-world relation that is (in my view rightly) invoked in the standard referentialist construal of some truths must be more fine-grained than the semantic link that Frege called reference. If the analysis that I provided of the slingshot arguments is correct, then we can lay down that the relation that stands at the heart of a suitable (referentialist) construal of most synthetic truths must be at least as fine-grained as the semantic aspect that Frege denoted by the term 'sense'.

It may be worth recalling also that the denial of standard referentialism in the semantics of various discourses is a doctrine about truth conditions, not a doctrine about referents. In other terms, by abandoning the idea that the truth conditions of our

truth-apt representations are to be specified always in terms of those conditions that the representations in question purport to be about one does not endorse any thesis about the nature or metaphysical status of the relevant intended referents. In particular, a non-referentialist need not deny the reality or the actual existence of these entities. In fact, she may embrace realism or anti-realism about them. In line with the intentionalist account of reference determination hinted at above, I believe that the adequacy of the latter doctrines (at least partly) hinges upon the subjects' referential intentions that accompany the application of the relevant symbols. Although these intentions are predominantly implicit in the subjects' cognitive and linguistic practice, they can be explicitly formulated in thought and expressed in public language as well. In pure mathematics, for instance, we do not entertain ideas or make claims about the nature or metaphysical status of mathematical objects. This, however, does not mean that our referential intentions in this discipline impose no constraint upon an appropriate "philosophical" characterisation of mathematical entities. Presumably, if questioned, most people would reject the idea that mathematical objects (i.e. the subject matters of pure mathematics) can appear in space and time, that they can be causally effective, or (I presume) that they are merely positions in a structure characterised categorically by a mathematical theory. These judgements are not mathematical, but they inform us about the speakers' referential intentions while doing mathematics, and they provide evidence against the adequacy of various revisionist theories (such as naturalism or structuralism) of the nature of mathematical entities. As regards the adequacy of realism about these referents, the decisive question is whether people conceive them as entities that can exist in the real world or as entities that can "exist" merely in a fictive domain invented by human minds. My guess is that most people would not exclude the possibility of the real existence of mathematical entities, as they would, in contrast, that of Little Red Riding Hood. If this is true, then realism may be an adequate doctrine about the referents of pure mathematics.

The observation that an adequate general construal of truth implies nothing about the metaphysical status of the subject matter of truth-apt representations may sound as an indirect embrace of deflationism in the theory of truth. Note, however, that the non-committal character of a general (not referent-invoking) construal of truth concerning the metaphysical status of subject matters need not prevent this construal from stating something substantive about the metaphysical status of truth itself. Beyond this conceptual point, in chapter 4, I argued that a proper explanation of objectivity calls for a realist construal of truth (or correct declarative applicability in general). Consequently, we can say that the general use-theoretic conception advocated in this work is not merely not entailing, but also diametrically opposed to deflationism in the theory of truth. As was emphasised in chapter 4, it is in fact a substantive realist account, according to which truth is a sort of correspondence, the property of having suitable semantic relation to some conditions that obtain in the actual world. The only difference between this and the existing referentialist versions of the correspondence theory is that this account does not assume that the worldly terms of the relevant correspondence relations are necessarily those facts that the true representations in question purport to be about.

Beyond the establishment of a viable naturalist construal of *a priori* truth and knowledge, the possibly most significant result of this work is the challenge it articulates to the standard, broadly Tarskian, referentialist conception of truth. Again, if the argumentation presented in the previous chapters is correct, then the idea that truth in general can be defined in terms of reference or satisfaction must be false. Now, one may think that the alternative use-theoretic construal advocated here is anti-Tarskian not merely in that it is not referentialist in character, but even in the stronger (and intuitively more problematic) sense that it fails to observe Tarski's T-schema or material adequacy condition for a theory of truth. As is well-known, the condition holds that a viable theory of truth must entail, for each sentence p of the language under scrutiny, the corresponding instance of the

schema "p is true if and only if *p*", where *p* stands for the metalinguistic translation of p. The conflict between the proposed use-theoretic construal and this requirement occurs if the truth or falsity of some sentences of the object language is not referential, while it is supposed that the metalanguage used for formulating the instances of Tarski's T-schema is still so (i.e. that the conditions "specified" by the metalinguistic expressions on the right-hand sides of the instances are always those that these expressions purport to be about). For instance, on a referentialist reading, Tarski's material adequacy condition holds that a proper theory of truth must entail, among other things, that the arithmetical sentence 'two plus three equals five' is true if and only if the abstract arithmetical conditions denoted by the metalinguistic translation of this sentence actually obtain. Under such a reading, the adequacy condition is certainly not observed by a general use-theoretic account that allows for a non-referentialist construal of this arithmetical truth. After all, if the truth conditions of arithmetical sentences are indeed non-referential, then this instance of Tarski's T-schema is not merely underivable from the use-theoretic conception of truth advocated in this work, but also false in the actual world.

Note, however, that Tarski's material adequacy condition need not be read along the standard referentialist line. What the condition intuitively requires is merely that the metalinguistic translation of each sentence of the object language be equivalent with the claim that the sentence in question is true, or (more exactly, as was observed in chapter 4) that the translations on the right-hand sides of the T-schema's instances in some way "specify" the truth conditions of the respective sentences mentioned on the left-hand sides. Now, if the truth of a sentence mentioned on the left-hand side of a schema instance is referential, then it is quite natural to suppose that the truth of its metalinguistic translation will be referential too, or (again, what is the same) that the translation in question does the required specification by referring to (i.e. being about) the relevant truth conditions. However, if the truth of the object sentence under scrutiny is not referential, then it is hard to see why it should be

supposed that the truth of its metalinguistic translation is still so. The intuitive assumption in this case is rather that the latter property is not referential either, which means that the translation "specifies" the relevant truth conditions merely by possessing them as such (i.e. possessing them as truth conditions, rather than as intended referents). Someone who understands the metalanguage, and therewith the translations provided on the right-hand side of the equivalence instances, will presumably understand what the truth or falsity of the corresponding sentences of the object language consists in even if the conditions "specified" are not referential. If Tarski's material adequacy condition is read along this line, then the suggested use-theoretic conception of truth turns out to observe this requirement just as much as its standard referentialist alternative. So, from the perspective of the current work, the problem with Tarski's condition is not that it imposes too strong a constraint upon a construal of truth. It is rather that its satisfaction does not guarantee the adequacy of a characterisation of this subject.

Having collected the most important general implications of this work concerning truth, reference and the relation of these semantic entities, in the remaining part of this conclusion, let me briefly review those general claims that an advocate of the proposed representationist construal must embrace concerning those characteristics that she has attributed to the paradigms of *a priori* truth in particular. In one sentence, the construal in question assumed that our logical and mathematical claims are analytic, their truth conditions non-referential, natural and representational, and their truth value necessary and *a priori* knowable. In view of the conceptual stipulations and theoretical assumption adopted in the previous chapters, it may be worth realising that some of these characteristics are more intimately related than the others.

In particular, I wish to draw attention to seven universal generalisations concerning the actual occurrences of these characteristics that seem to follow from the conceptual and doctrinal commitments made in this work. The first pair is a direct consequence of the suggested representationist construal of

a priori truth, and it observes that all *a priori* knowable truths are representational and all representational truths that are knowable from a first-personal perspective are *a priori* knowable. Since representationality is, at least in principle, an empirically discernible feature of truth conditions, whose occurrence can be established by a careful examination of the subject's cognitive or linguistic behaviour, the identification of apriority with this feature may promote the emergence of a systematic empirical theory of the actual nature and scope of *a priori* knowledge, justification and evidence as well as the sources of *a priori* epistemic mistakes.

The second pair of universal generalisations that I wish to bring into relief states that all knowable (explicitly or implicitly) modal truths are analytic and all analytic truths are (explicitly or implicitly) modal in character. This result can be derived from the semantical explanation of analytic modal truths presented in chapter 7, the idea that knowledge requires causal contact with obtaining truth conditions (so we cannot acquire knowledge of what obtains in realistically construed non-actual worlds), and Hume's (here merely granted) observation that we have no experience of modal (e.g. necessary, possible, probable and chancy) facts in the actual world. Note that this generalisation does not deny the existence of synthetic modal truths (or else claims made true by such modal facts). What it denies is merely that such truths can be discovered by a cognitive subject (unless it is one who, *pace* Hume, is capable of experiencing modal aspects in the actual world). The implication is nonetheless significant, since it runs counter to the popular view that we can acquire synthetic knowledge of the occurrences of various modal aspects (e.g. the necessity of identities, causal relations and essential property possessions, and the probability or chancy character of the obtaining of particular conditions) in the actual world.

Beyond subscribing to the premises mentioned in the previous paragraph, I advanced a few other considerations as well to weaken the appeal of this popular view. Discussing Putnam's case against metaphysical realism in chapter 5, for instance, I delineated a broadly Humean construal of causation, which does

not assume the necessary character of this natural relation. If the construal is correct, then our synthetic knowledge of causal relations will not qualify as knowledge of the occurrence of modal aspects in the actual world. Further, in chapter 1, I argued that the apparent *de re* necessity of essential property possession can be explained by reference to our (content constitutive) referential intentions or explicit stipulations, without invoking modal occurrences in the actual world or non-modal occurrences in realistically construed non-actual worlds. If this explanatory strategy is correct, then our knowledge of necessarily possessed essential properties proves to be analytic in character. In chapter 7, I applied this strategy to the case of the scientific reduction of water to H_2O, and argued that if water possesses this chemical structure necessarily, then it is merely because we take it to be something that must do so (i.e. must possess this structure), rather than because of the obtaining of some modal property in the actual world. A complete case against the idea that we can acquire knowledge of synthetic modal truths would require a careful examination of our claims about identities and probabilities as well, something that could not be properly done in this work. Nevertheless, I believe that the result of such an investigation would be also in line with the above generalisation. While our identity claims are arguably also analytic in character, our synthetic claims about chances and probabilities are either unjustified or made true by non-modal truth conditions (e.g. relative frequencies) in the actual world, whose obtaining or absence can be established by human minds.

The fifth universal generalisation that I wish to mention here holds that all analytic truths are representational. The claim is a consequence of the representationist construal of analytic relations elaborated in chapter 7. The opposite generalisation (that all representational truths are analytic) does not follow from the commitments of this work. This is because the analytic relations of representations were not supposed to exhaust the representational conditions that may or may not obtain in the natural world. For instance, the truth conditions of someone's belief that a subject has representations in her head are clearly

representational. Still the belief is synthetic, since its truth value is not determined by the relations of the concepts that occur in its formulation.

The above characterisation of the relation of analyticity and representationality has at least two significant implications in philosophy. First, it provides a clear view of the factual ground of analytic truths, something that has been called for by many critics of post-Fregean moderate empiricist analytic philosophy. In particular, it suggests that analytic claims are true or false in virtue of the obtaining of some content-constitutive natural relations among representations in human heads (and in other representation-manipulating devices). Second, by acknowledging the existence of synthetic representational truths (which might be established from a first-personal perspective as well), the characterisation also concedes the possibility of synthetic *a priori* knowledge of a limited segment of the natural world.

The construal of analyticity adopted in this work gives rise to a sixth universal generalisation as well, according to which all analytic truths are non-referential. This time the fundamental observation to rely on is that the representational conditions whose obtaining or absence was supposed to determine the truth value of analytic beliefs are always different from those conditions that the beliefs in question purport to be about. Having recognised this fact, nevertheless, we must note that in the case of a metalinguistic claim about the obtaining of the truth conditions of an analytic claim (i.e. the obtaining of a content-constitutive natural relation between some constituents of the relevant analytic claim) the obtaining of the denoted conditions may be constitutive of the content of the metalinguistic claim as well. This situation occurs *if* the constituents of the analytic claim *mentioned* in the metalinguistic claim are supposed to be identified with respect to their semantic contents (rather than merely to their syntactic shape). What this means is that the truth value of such a metalinguistic claim may depend at least as much on the analytic relations of its own constituents as on the obtaining or absence of the conditions that it purports to be about.

Finally, the seventh universal generalisation that I wish to stress declares that all knowledge is knowledge of the actual (past or present) obtaining of some natural (i.e. spatiotemporal) conditions. The claim can be derived from the epistemological principle adopted here that knowledge is always knowledge of the obtaining of some truth conditions that must stand in a suitable causal relation with the knowing mind. Again, the opposite generalisation (that the obtaining or absence of any natural condition is knowable) is not meant to be implied in this work. Although there are reasons for assuming the existence of unknowable facts in the natural world, the argumentation developed in the previous chapters did not require taking a stance on this particular issue.

With this note, I have finished the promised survey of the general implications of the conceptual and doctrinal commitments of this work. The main purpose of this overview was to provide a sense of the broader significance of the proposals and claims that have been adopted and defended in the previous chapters. Although the central concern of the inquiry advanced in those chapters was the elaboration of a viable construal of *a priori* truth and knowledge, the realisation of this task required the development of a relatively clear notion of a number of other phenomena that are essentially related to these subjects. Thus we have witnessed, among other things, the emergence of a general realist use-theoretic construal of truth, a fine-grained, intentionalist, deflated construal of reference, a communicable metaphysical construal of realism that does not imply actual existence, a strictly naturalist causal contact theory of knowledge acquisition, a related representationist construal of analyticity that gives rise to a naturalist semantical explanation of knowledge of (analytic) modal truths, and a broadly causal account of how we learn to think of or speak about epistemically inaccessible entities or domains. Inflated with these conceptual and theoretical elements, the core tenets of this work acquire a wider significance than one might attribute to them after a brief reflection upon the primary subject of this investigation. In this conclusion, I wished to make sure that this collateral impact is

understood as clearly as the central claims and suggestions made concerning the nature of *a priori* truth, knowledge, justification and evidence.

Of course, wider significance means wider ground for potential criticism. In chapter 2, I presented a list of those phenomena whose proper explanation, with the assistance of other theories, I regarded as a condition of minimal adequacy for any construal of *a priori* truth. The stress on minimality was meant to keep awareness of the fact that a fully adequate notion of apriority must face the tribunal of a much larger evidential pool than the one (with the most striking characteristics of the subject) advanced in chapter 2. The implications collected here highlight a few latent segments of this larger evidential pool.

Having said that, I wish to conclude this work with a brief remark on the relation of the outcome of this philosophical inquiry to our current empirical (scientific and ordinary) theories of the spatiotemporal world. As often in philosophy, the primary purpose of this investigation was the development of a relatively sharp and theoretically useful notion of a phenomenon (viz. the apriority of truths and pieces of knowledge) that acquired a certain significance in the traditional concerns of the discipline. Although the work has been carried out from an "armchair", without the inclusion of new empirical data to test the adequacy of the adopted conceptual and theoretical elements, the commitments made here were nevertheless grounded on a considerable amount of empirical evidence of how things are in the natural world. In particular, conceptual proposals were made with an eye on the observable characteristics of our cognitive and linguistic practice (i.e. on how we used to apply the relevant mental and physical symbols in various symbolic and situative contexts) and the expected theoretical benefits of the suggested conventions, while theoretical claims were made on the basis of explanatory considerations and some epistemologically fundamental (though fallible) observations of the intended phenomena. Beyond attempting to satisfy these empirical adequacy conditions, one major methodological consideration behind the search for a viable naturalist (as opposed to, say,

platonist) construal of *a priori* truth and knowledge was that the schematic picture delineated by such an account is expected to be both confirmable and further specifiable by the results of future empirical inquiries into the relevant segments of the spatiotemporal world. In other terms, the suggested representationist construal of *a priori* truth and knowledge differs from its referentialist alternatives not merely in that it is supported by our current observations of the world, but also in that it can gain evidential support as well as further sophistication from the future findings of the sciences of human brain and mind. By calling for such support and sophistication, an advocate of this construal may inspire various inquiries in the above sciences. The continuity between her philosophy, on the one hand, and her ordinary or scientific theory formation, on the other, does not reduce to the regular adjustment of the former to the empirical adequacy conditions imposed by the latter. Beyond this adjustment, she can also inspire and shape new empirical research programmes in the purported disciplines. The fact that her philosophy is methodologically naturalised does not entail that it can have no impact on the development of our empirical theories of the world.

BIBLIOGRAPHY

Armstrong, D. (1989), *A Combinatorial Theory of Possibility*, Cambridge: Cambridge University Press.

Ayer, A. J. (1935), 'The Criterion of Truth', *Analysis*, 3 (1-2): 28-32.

Ayer, A. J. (1946), *Language, Thought and Logic*, 2nd rev. and reset ed., London: Gollancz.

Balaguer, M. (1998), *Platonism and Anti-Platonism in Mathematics*, Oxford: Oxford University Press.

Bealer, G. (1999), 'The A Priori', in Greco, J. and Sosa, E. (1999): 243-270.

Benacerraf, P. (1965), 'What Numbers Could not Be', *Philosophical Review*, 74 (1): 47-73.

Benacerraf, P. (1973), 'Mathematical Truth', *Journal of Philosophy*, 70 (19): 661-679.

Benacerraf, P. and Putnam, H. (1983), *Philosophy of Mathematics*, 2nd ed., Cambridge: Cambridge University Press.

Blackburn, S. ed. (1975), *Meaning, Reference and Necessity*, Cambridge: Cambridge University Press.

Blackburn, S. (1993), *Essays in Quasi-Realism*, New York: Oxford University Press.

Block, N. (1986), 'Advertisement for a Semantics for Psychology', in French, P. A., Uehling T. E., Wettstein, H. K. (1986): 615-678.

Boghossian, P. and Peacocke, C. eds. (2000), *New Essays on the A Priori*, Oxford: Clarendon Press.

BonJour, L. (1985), *The Structure of Empirical Knowledge*, Cambridge, MA: Harvard University Press.

BonJour, L. (1998), *In Defence of Pure Reason*, Cambridge: Cambridge University Press.

Brandom, R. B. (1988), 'Pragmatism, Phenomenalism and Truth Talk', in French, P. A., Uehling, T. E. and Wettstein, H. K. (1988): 75-93.

Brandom, R. B. (1994), Making it Explicit: Reasoning, Representing, and Discursive Commitment, Cambridge, MA: Harvard University Press.
Brouwer, L. E. J. (1949), 'Consciousness, Philosophy, and Mathematics', *Proceedings of the Tenth International Congress of Philosophy*, 3: 1235-1249.
Brown, J. R. (1991), The Laboratory of the Mind: Thought Experiments in the Natural Sciences, London: Routledge.
Burgess, J. P. and Rosen, G. (2005), 'Nominalism Reconsidered', in Shapiro, S. (2005): 515-536.
Carnap, R. (1934), *Logische Syntax der Sprache*, Wien: Springer.
Casullo, A. (2003), *A Priori Justification*, Oxford: Oxford University Press.
Chihara, C. (1990), *Constructibility and Mathematical Existence*, Oxford: Oxford University Press.
Chisholm, R. M. (1989), *Theory of Knowledge*, 3rd ed., Englewood Cliffs, NJ: Prentice Hall.
Church, A. (1943), 'Review of Carnap's *Introduction to Semantics*', *Philosophical Review*, 52 (3): 298-304.
Church, A. (1956), *Introduction to Mathematical Logic*, Princeton, NJ: Princeton University Press.
Churchland, P. M. (1981), 'Eliminative Materialism and the Propositional Attitudes', *Journal of Philosophy*, 78 (2): 67-90.
Churchland, P. M. (1988/1992), *Matter and Consciousness*, rev. ed., Cambridge, Mass.: MIT Press.
Colyvan, M. (2001), *The Indispensability of Mathematics*, New York: Oxford University Press.
Colyvan, M. (2007), 'Mathematical Recreation Versus Mathematical Knowledge', in Leng, M., Paseau, A. and Potter, M. (2007): 109-122.
Colyvan, M. and Zalta, E. N. (1999), 'Mathematics: Truth *and* Fiction? Review of Mark Balaguer's *Platonism and Anti-Platonism in Mathematics*', *Philosophia Mathematica*, 7 (3): 336-349.
Craig, E. J. (1975), 'The Problem of Necessary Truth', in Blackburn, S. (1975): 1-31.
Davidson, D. (1969), 'True to the Facts', *Journal of Philosophy*, 66 (21): 748-764.

Davidson, D. (1970), 'Mental Events', in Foster, L. and Swanson, J. (1970): 79-101.
Davidson, D. (1984), *Inquiries into Truth and Interpretation*, Oxford: Oxford University Press.
Devitt, M. (1981), *Designation*, New York: Columbia University Press.
Devitt, M. (1984), *Realism and Truth*, Oxford: Blackwell.
Devitt, M. (1991), 'Aberrations of the Realism Debate', *Philosophical Studies*, 61: 43-63.
Dewey, J. (1925), *Experience and Nature*, Chicago: Open Court.
Duhem, P. (1906), *La théorie physique: son objet et sa structure*, Paris: Chevalier & Rivière.
Dummett, M. (1963), 'Realism', in Dummett, M. (1978): 145-165.
Dummett, M. (1973), 'The Philosophical Basis of Intuitionistic Logic', in Dummett, M. (1978): 215-247.
Dummett, M. (1978), *Truth and Other Enigmas*, London: Duckworth.
Dummett, M. (1991), *The Logical Basis of Metaphysics*, Cambridge, MA: Harvard University Press.
Dummett, M. (1993), *The Seas of Language*, Oxford: Clarendon Press.
Dworkin, R. (1996), 'Objectivity and Truth: You'd Better Believe it', *Philosophy and Public Affairs*, 25 (2): 87-139.
Einstein, A. (1905), 'Zur Elektrodynamik bewegter Körper', *Annalen der Physik*, 17: 891-921.
Einstein, A., Podolski, B. and Rosen, N. (1935), 'Can Quantum Mechanical Description of Reality Be Considered Complete?' *Physical Review*, 47 (10): 777-780.
Evans, G. (1973), 'The Causal Theory of Names', *Proceedings of the Aristotelian Society*, suppl. vol. 47: 187-225.
Field, H. (1972), 'Tarski's Theory of Truth', *Journal of Philosophy*, 69 (13): 347-375.
Field, H. (1980), *Science without Numbers: A Defence of Nominalism*, Princeton, NJ: Princeton University Press.
Field, H. (1984), 'Platonism for Cheap? Crispin Wright on Frege's Context Principle', *Canadian Journal of Philosophy*, 14: 637-662.

Field, H. (1986), 'The Deflationary Conception of Truth', in MacDonald, G. and Wright, C. (1986): 55-117.
Field, H. (1989), *Realism, Mathematics and Modality*, Oxford: Blackwell.
Field, H. (1994), 'Deflationist Views of Meaning and Content', *Mind*, 103 (411): 294-285.
Field, H. (2001), *Truth and the Absence of Fact*, Oxford: Clarendon Press.
Fine, A. (1984), 'The Natural Ontological Attitude', in Leplin, J. (1984): 83-107.
Fine, K. (2001), 'The Question of Realism', *Philosophers' Imprint*, 1 (1): 1-30.
Fine, K. (2002), *The Limits of Abstraction*, Oxford: Clarendon Press.
Fitch, F. B. (1963), 'A Logical Analysis of Some Value Concepts', *Journal of Symbolic Logic*, 28 (2): 135-142.
Fodor, J. A. (1975), *The Language of Thought*, Cambridge, MA: MIT Press.
Fodor, J. A. (1981), *Representations*, Cambridge, MA: MIT Press.
Fodor, J. A. (1987), *Psychosemantics*, Cambridge, MA: MIT Press.
Fodor, J. A. (1994), *The Elm and the Expert*, Cambridge, MA: MIT Press.
Forrai, G. (2001), *Reference, Truth and Conceptual Schemes*, Dordrecht: Kluwer.
Foster, L. and Swanson, J. eds. (1970), *Experience and Theory*, London: Duckworth.
Frege, G. (1879), Begriffsschrift: eine der arithmetischen nachgebildete Formelsprache des reinen Denkens, Halle a. S.: Louis Nebert; transl. by Bauer Mengelbert, S. as Concept Notation: A Formula Language of Pure Thought, Modelled upon that of Arithmetic, in van Heijenoort, J. (1967)
Frege, G. (1884), Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl, Breslau: Wilhelm Koebner; transl. by Austin, J. L. as The Foundations of Arithmetic, Oxford: Basil Blackwell, 1950.
Frege, G. (1892), 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik*, 100: 25-50.

Frege, G. (1893), *Grundgesetze der Arithmetik*, Vol. 1, Hildesheim: Georg Olms.
French, P. A., Uehling, T. E. and Wettstein, H. K. eds. (1986), *Midwest Studies in Philosophy*, Vol. X, Minneapolis, MN: University of Minnesota Press.
French, P. A., Uehling, T. E. and Wettstein, H. K. eds. (1988), *Midwest Studies in Philosophy*, Vol. XII, Minneapolis, MN: University of Minnesota Press.
Galilei, G. (1632), Dialogo sopra I due massimi sistemi del mondo Tolemaico e Copernicano, Firenze: Battista Landini.
Galilei, G. (1638), Discorsi, e dimostrazioni matematiche, intorno à due nuove scienze attenenti alla mecanica, e i movimenti locali, Leiden: Elsevier.
Gendler, T. S. (2000), Thought Experiment. On the Powers and Limits of Imaginary Cases, New York: Garland.
George, A. ed. (1989), *Reflections on Chomsky*, Oxford: Basil Blackwell.
George, A. ed. (1994), *Mathematics and Mind*, Oxford: Oxford University Press.
Gettier, E. L. (1963), 'Is Justified True Belief Knowledge?' *Analysis*, 23 (6): 121-123.
Gibbard, A. (1990), *Wise Choices, Apt Feelings: A Theory of Normative Judgement*, Cambridge, MA: Harvard University Press.
Gibbard, A. (1994), 'Meaning and Normativity', in Villanueva, E. (1994): 95-115.
Gödel, K. (1944), 'Russell's Mathematical Logic', in Schilpp, P. A. (1944): 125-153.
Gödel, K. (1951), 'Some Basic Theorems on the Foundations of Mathematics and Their Implications', in Gödel, K. (1995): 304-323.
Gödel, K. (1964), 'What Is Cantor's Continuum Problem', in Benacerraf, P. and Putnam, H. (1983): 470-485.
Gödel, K. (1995), *Collected Works*, Vol. III, Oxford: Oxford University Press.
Goldman, A. I. (1967), 'A Causal Theory of Knowing', *Journal of Philosophy*, 64 (12): 357-372.

Goodman, N. (1954), *Fact, Fiction, and Forecast*, London: Athlone Press.
Goodman, N. (1978), *Ways of Worldmaking*, Indianapolis: Hackett.
Greco, J. and Sosa, E. eds. (1999), *The Blackwell Guide to Epistemology*, Malden, MA: Blackwell.
Grover, D. L., Camp, J. L. and Belnap, N. D. (1975), 'A Prosentential Theory of Truth', *Philosophical Studies*, 27 (2): 73-125.
Gunderson, K. ed. (1975), *Minnesota Studies in the Philosophy of Science*, Vol. VII, Minneapolis, MN: University of Minnesota Press.
Haggqvist, S. (1996), *Thought Experiments in Philosophy*, Stockholm: Almqvist & Wiksell International.
Hahn, L. and Schilpp, P. A. eds. (1986), *The Philosophy of W. V. Quine*, La Salle, IL: Open Court.
Hale, R. (1987), *Abstract Objects*, Oxford: Basil Blackwell.
Hale, R. (1994), 'Dummett's Critique of Wright's Attempt to Resuscitate Frege', *Philosophia Mathematica*, 2 (2): 122-147.
Hale, R. (1997), 'Realism and Its Oppositions', in Hale, R. and Wright, C. (1997a): 271-308.
Hale, R. (2001), 'A Response to Potter and Smiley: Abstraction by Recarving', *Proceedings of the Aristotelian Society*, 101: 339-358.
Hale, R. and Wright, C. eds. (1997a), *A Companion to the Philosophy of Language*, Oxford: Basil Blackwell.
Hale, R. and Wright, C. (1997b), 'Putnam's Model-Theoretic Argument Against Metaphysical Realism', in Hale, R. and Wright, C. (1997a): 427-457.
Hale, R. and Wright, C. (2001a), The Reason's Proper Study: Essays toward a Neo-Fregean Philosophy of Mathematics, Oxford: Clarendon Press.
Hale, R. and Wright, C. (2001b), 'To Bury Caesar...' in Hale, R. and Wright, C. (2001a): 335-396.
Hale, R. and Wright, C. (2002), 'Benacerraf's Dilemma Revisited', *European Journal of Philosophy*, 10 (1): 101-129.
Hanson, P. P. and Hunter, B. A. eds. (1992), *The Return of the A Priori*, Calgary: The University of Calgary Press.

Hare, R. M. (1952), *The Language of Morals*, Oxford: Clarendon Press.
Harman, G. (1973), *Thought*, Princeton, NJ: Princeton University Press.
Hatfield, G. (1990), The Natural and the Normative: Theories of Spatial Perception from Kant to Helmholtz, Cambridge, MA: MIT Press.
Hellman, G. (1989), *Mathematics without Numbers*, Oxford: Clarendon Press.
Hilbert, D. (1925), 'On the Infinite', in Benacerraf, P. and Putnam, H. (1983): 183-201.
Hitchcock, C. ed. (2004), *Contemporary Debates in the Philosophy of Science*, Malden, MA: Basil Blackwell.
Horowitz, T. (2006), *The Epistemology of A Priori Knowledge*, Oxford: Oxford University Press.
Horowitz, T. and Massey, G. eds. (1991), *Thought Experiments in Science and Philosophy*, Savage, MD: Rowman and Littlefield.
Horwich, P. (1998a), *Meaning*, Oxford: Clarendon Press.
Horwich, P. (1998b), *Truth*, 2nd ed., Oxford: Clarendon Press.
Horwich, P. (2005), *Reflections on Meaning*, Oxford: Clarendon Press.
Hull, D. M., Forbes, M. and Okruhlik, K. eds. (1992), *PSA 1992, Proceedings of the 1992 Biennial Meeting of the Philosophy of Science Association*, Vol. 2, East Lansing, MI: Philosophy of Science Association.
Isaacson, D. (forthcoming), 'The Reality of Mathematics and the Case of Set Theory', in Novák, Zs. and Simonyi, A. (forthcoming-b)
Jackson, F. and Smith, M. eds. (2005), *The Oxford Handbook of Contemporary Philosophy*, Oxford: Oxford University Press.
Kant, I. (1781/1787), *Kritik der reinen Vernunft*, transl. by Guyer, P. as *The Critique of Pure Reason*, Cambridge: Cambridge University Press, 1998.
Katz, J. (1981), *Language and Other Abstract Objects*, Totowa, NJ: Rowman and Littlefield.
Katz, J. (1995), 'What Mathematical Knowledge Could Be', *Mind*, 104 (415): 491-520.

Kitcher, P. (1980), 'Apriority and Necessity', *Australasian Journal of Philosophy*, 58: 89-101.
Kitcher, P. (1984), *The Nature of Mathematical Knowledge*, New York: Oxford University Press.
Knowles, J. (2003), *Norms, Naturalism and Epistemology*, Houndmills: Palgrave Macmillan.
Koyré, A. (1939), *Études galiléennes*, Paris: Hermann.
Kripke, S. (1972/1980), *Naming and Necessity*, rev. and enl. ed., Oxford: Basil Blackwell.
Kuhn, T. S. (1977a), The Essential Tension: Selected Studies in Scientific Tradition and Change, Chicago: The University of Chicago Press.
Kuhn, T. S. (1977b), 'A Function for Thought Experiments', in Kuhn, T. S. (1977a): 240-265.
Lance, M. and Hawthorne, J. (1997), *The Grammar of Meaning: Normativity and Semantic Discourse*, Cambridge: Cambridge University Press.
Leeds, S. (1978), 'Theories of Reference and Truth', *Erkenntnis*, 13 (1): 111-129.
Leng, M., Paseau, A. and Potter, M. eds. (2007), *Mathematical Knowledge*, Oxford: Oxford University Press.
Leplin, J. ed. (1984), *Scientific Realism*, Berkeley: University of California Press.
Lewis, D. (1984), 'Putnam's Paradox', *Australasian Journal of Philosophy*, 62 (3): 221-236.
Lewis, D. (1986), *On the Plurality of Worlds*, Oxford: Basil Blackwell.
Liggins, D. (2006), 'Is There a Good Epistemological Argument against Platonism?' *Analysis*, 66 (2): 135-141.
Loar, B. (1987), 'Truth Beyond All Verification', in Taylor, B. (1987): 81-116.
MacDonald, C. and MacDonald, G. eds. (1995), *Connectionism: Debates on Psychological Explanation*, Oxford: Blackwell.
MacDonald, G. and Wright, C. eds. (1986), Fact, Science and Morality. Essays on A. J. Ayer's Language, Truth and Logic, Oxford: Basil Blackwell.

Mach, E. (1897), 'Über Gedankenexperimente', Poskes Zeitschrift für den physikalischen und chemischen Unterricht, January: 1-5.
Mackie, J. L. (1977), *Ethics: Inventing Right and Wrong*, Harmondsworth: Penguin Books.
Maddy, P. (1980), 'Perception and Mathematical Intuition', *Philosophical Review*, 89: 163-196.
Maddy, P. (1990), *Realism in Mathematics*, Oxford: Oxford University Press.
Maddy, P. (1997), *Naturalism in Mathematics*, Oxford: Oxford University Press.
Marr, D. (1982), *Vision*, New York: W. H. Freeman and Company.
Maxwell, J. C. (1871), *Theory of Heat*, London: Longman.
McAllister, J. W. (1996), 'The Evidential Significance of Thought Experiment in Science', *Studies in History and Philosophy of Science*, 27 (2): 233-250.
McCulloch, W. S. and Pitts, W. (1943), 'A Logical Calculus of the Ideas Immanent in Nervous Activity', *Bulletin of Mathematical Biophysics*, 5: 115-133.
McDowell, J. (1981), 'Anti-Realism and the Epistemology of Understanding', in McDowell, J. (1998): 314-343.
McDowell, J. (1998), *Meaning, Knowledge and Reality*, Cambridge, MA: Harvard University Press.
Mill, J. S. (1843), *A System of Logic*, London: Longmans, Green and Company.
Miller, A. (2002), 'What Is the Manifestation Argument?' *Pacific Philosophical Quarterly*, 83 (4): 352-383.
Millican, P. and Clark, A. eds. (1996), *Machines and Thought: The Legacy of Alan Turing*, Vol. 1, Oxford: Clarendon Press.
Miščević, N. (1992), 'Mental Models and Thought Experiments', *International Studies in the Philosophy of Science*, 6 (3): 215-226.
Morton, A. and Stich, S. P. eds. (1996), *Benacerraf and His Critics*, Oxford: Basil Blackwell.
Nadel, L., Cooper, L. A*., et al.* eds. (1989), *Neural Connections, Mental Computation*, Cambridge, MA: MIT Press.

Neale, S. (1995), 'The Philosophical Significance of Gödel's Slingshot', *Mind*, 104 (416): 761-825.
Neale, S. (2001), *Facing Facts*, Oxford: Clarendon.
Newell, A. and Simon, H. A. (1976), 'Computer Science as Empirical Inquiry: Symbols and Search', *Communications of the Association for Computing Machinery*, 19: 113-126.
Newton, I. (1687), *Philosophiae naturalis principia mathematica*, London: Joseph Streater.
Norton, J. (1996), 'Are Thought Experiments Just What You Thought?' *Canadian Journal of Philosophy*, 26 (3): 333-366.
Novák, Zs. (2006), 'Review of Jonathan Knowles' Norms, Naturalism and Epistemology', International Studies in the Philosophy of Science, 20 (1): 113-124.
Novák, Zs. and Simonyi, A. eds. (2010a), *Truth, Reference and Realism*, Budapest: Central European University Press.
Novák, Zs. and Simonyi, A. (2010b), 'Introduction', in Novák, Zs. and Simonyi, A. (2010a)
Parsons, C. (1980), 'Mathematical Intuition', *Proceedings of the Aristotelian Society*, 80: 145-168.
Parsons, C. (1994), 'Intuition and Number', in George, A. (1994): 141-157.
Peacocke, C. (2005), 'The A Priori', in Jackson, F. and Smith, M. (2005): 739-763.
Plantinga, A. (1993), *Warrant and Proper Function*, New York: Oxford University Press.
Popper, K. R. (1959), *The Logic of Scientific Discovery*, London: Hutchinson.
Posner, M. I. (1989), *Foundations of Cognitive Science*, Cambridge, MA: MIT Press.
Potter, M. and Smiley, T. (2001), 'Abstraction by Recarving', *Proceedings of the Aristotelian Society*, 101: 327-338.
Putnam, H. (1967), 'Mathematics without Foundations', *Journal of Philosophy*, 64 (1): 5-22.
Putnam, H. (1971), *Philosophy of Logic*, New York: Harper and Row.
Putnam, H. (1975a), *Mathematics, Matter and Method. Philosophical Papers*, Vol. 1, Cambridge: Cambridge University Press.

Putnam, H. (1975b), 'The Meaning of "Meaning"', in Gunderson, K. (1975): 131-193.
Putnam, H. (1975c), *Mind, Language and Reality. Philosophical Papers*, Vol. 2, Cambridge: Cambridge University Press.
Putnam, H. (1975d), 'What Is Mathematical Truth?' in Putnam, H. (1975a): 60-78.
Putnam, H. (1977), 'Models and Reality', in Putnam, H. (1983): 1-25.
Putnam, H. (1981), *Reason, Truth and History*, Cambridge: Cambridge University Press.
Putnam, H. (1983), *Realism and Reason. Philosophical Papers*, Vol. 3, Cambridge: Cambridge University Press.
Putnam, H. (1987), *The Many Faces of Realism*, La Salle, IL: Open Court.
Putnam, H. (1988), *Representation and Reality*, Cambridge, MA: MIT Press.
Putnam, H. (1989), 'Model Theory and the 'Factuality' of Semantics', in George, A. (1989): 213-232.
Putnam, H. (1992), *Renewing Philosophy*, Cambridge, MA: Harvard University Press.
Putnam, H. (1999), *The Threefold Cord: Mind, Body, and World*, New York: Columbia University Press.
Quine, W. V. (1948), 'On What There Is', *Review of Metaphysics*, 2 (5): 21-38.
Quine, W. V. (1951), 'Two Dogmas of Empiricism', *Philosophical Review*, 60 (1): 20-43.
Quine, W. V. (1953a), *From a Logical Point of View*, Cambridge, MA: Harvard University Press.
Quine, W. V. (1953b), 'Reference and Modality', in Quine, W. V. (1953a): 139-159.
Quine, W. V. (1960a), 'Carnap and Logical Truth', *Synthese*, 12 (4): 350-374.
Quine, W. V. (1960b), *Word and Object*, Cambridge, MA: MIT Press.
Quine, W. V. (1970), *Philosophy of Logic*, Englewood Cliffs, NJ: Prentice-Hall.

Quine, W. V. (1975), 'On Empirically Equivalent Systems of the World', *Erkenntnis*, 9: 313-328.
Quine, W. V. (1986), 'Reply to Charles Parsons', in Hahn, L. and Schilpp, P. A. (1986): 396-403.
Quine, W. V. (1995), *From Stimulus to Science*, Cambridge, MA: Harvard University Press.
Quinton, A. (1972), 'The *A Priori* and the Analytic', *Proceedings of the Aristotelian Society*, 64: 31-54.
Ramsey, F. (1927), 'Facts and Propositions', *Proceedings of the Aristotelian Society*, suppl. vol. 7: 153-170.
Resnik, M. (1982), 'Mathematics as a Science of Patters: Epistemology', *Nous*, 16 (1): 95-105.
Resnik, M. (1992), 'A Structuralist's Involvement with Modality (review of Hellman 1989)', *Mind*, 101 (401): 107-122.
Resnik, M. (1997), *Mathematics as a Science of Patterns*, Oxford: Oxford University Press.
Rosen, G. (1998), 'Blackburn's *Essays in Quasi-Realism*', *Noûs*, 32 (3): 386-405.
Rumelhart, D. E. (1989), 'The Architecture of the Mind: A Connectionist Approach', in Posner, M. I. (1989): 133-159.
Rumelhart, D. E. and McClelland, J. L. (1986), *Parallel Distributed Processing*, Vol. 1, Cambridge, MA: MIT Press.
Russell, B. (1905), 'On Denoting', *Mind*, 14 (56): 479-493.
Russell, B. (1912), *The Problems of Philosophy*, London: Williams and Norgate.
Schilpp, P. A. (1944), *The Philosophy of Bertrand Russell*, Evanston, IL: Northwestern University Press.
Schirn, M. ed. (1998), *Philosophy of Mathematics Today*, Oxford: Clarendon Press.
Schrödinger, E. (1935), 'Die gegenwärtige Situation in der Quantenmechanik', *Naturwissenschaften*, 23: 807-812, 823-828, 844-849.
Shapiro, S. (1989), 'Structure and Ontology in Philosophy of Mathematics', *Philosophical Topics*, 17 (2): 145-171.
Shapiro, S. (1997), *Philosophy of Mathematics: Structure and Ontology*, New York: Oxford University Press.

Shapiro, S. (2000), *Thinking about Mathematics*, Oxford: Oxford University Press.
Shapiro, S. (2005), The Oxford Handbook of Philosophy of Mathematics and Logic, Oxford: Oxford University Press.
Smolensky, P. (1988), 'On the Proper Treatment of Connectionism', *Behavioral and Brain Sciences*, 11: 1-74.
Smolensky, P. (1989), 'Connectionist Modeling: Neural Computation / Mental Connections ', in Nadel, L., Cooper, L. A.*, et al.* (1989): 49-67.
Soames, S. (1984), 'What Is a Theory of Truth?' *Journal of Philosophy*, 81 (8): 411-429.
Soames, S. (1997), 'The Truth about Deflationism', *Philosophical Issues*, 8, Truth: 1-44.
Sorensen, R. A. (1992), *Thought Experiments*, Oxford: Oxford University Press.
Steiner, M. (1975), *Mathematical Knowledge*, Ithaca, NY: Cornell University Press.
Steiner, M. (1998), *The Applicability of Mathematics as a Philosophical Problem*, Cambridge, MA: Harvard University Press.
Stevenson, C. L. (1944), *Ethics and Language*, New Haven: Yale University Press.
Stevin, S. (1605), *Wisconstige gedachtenissen*, Leiden: Ian Bouvvensz.
Stich, S. P. (1983), *From Folk Psychology to Cognitive Science*, Cambridge, MA: MIT Press.
Strawson, P. (1950), 'Truth', *Proceedings of the Aristotelian Society*, suppl. vol. 24: 129-156.
Tarski, A. (1944), 'The Semantic Conception of Truth', *Philosophy and Phenomenological Research*, 4: 341-75.
Taylor, B. ed. (1987), *Michael Dummett: Contributions to Philosophy*, Dordrecht: Martinus Nijhoff.
Tennant, N. (1997), *Taming of the True*, Oxford: Clarendon Press.
Turing, A. (1950), 'Computing Machinery and Intelligence', *Mind*, 59: 433-460.
Van Heijenoort, J. ed. (1967), *From Frege to Gödel: A Sourcebook in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press.

Villanueva, E. ed. (1994), *Philosophical Issues 5: Truth and Rationality*, Atascadereo, CA: Ridgeview Publishing Co.

Williams, M. J. (1986), 'Do We (Epistemologists) Need a Theory of Truth?' *Philosophical Topics*, 14 (11): 223-242.

Williamson, T. (2000), *Knowledge and Its Limits*, Oxford: Oxford University Press.

Wittgenstein, L. (1953), *Philosophische Untersuchungen*, transl. by Anscombe, G. E. M. as *Philosophical Investigations*, Oxford: Basil Blackwell, 1953.

Wright, C. (1983), *Frege's Conception of Numbers as Objects*, Aberdeen: Aberdeen University Press.

Wright, C. (1992), *Truth and Objectivity*, Cambridge, MA: Harvard University Press.

Wright, C. (1993), *Realism, Meaning and Truth*, 2nd ed., Oxford: Basil Blackwell.

Wright, C. (1998), 'On the Harmless Impredicativity of N= (Hume's Principle)', in Schirn, M. (1998): 339-368.

INDEX

SAMENVATTING

De opvatting dat we kennis van objectieve waarheden kunnen verkrijgen door middel van bewijs anders dan zintuiglijke ervaring is in de wijsbegeerte gangbaar en plausibel. Maar als we eenmaal geloven in de mogelijkheid van *a priori* kennisverwerving, rijst vanzelf de volgende vraag: wat gebeurt hier? Hoe leren we wat objectief en noodzakelijkerwijs het geval is zonder te vertrouwen op wat de waarneming ons vertelt? Het is duidelijk dat een antwoord op deze vraag zwaar zal leunen op wat we redelijkerwijs kunnen denken over de aard van die condities, waarvan het (niet) vervuld zijn de waarheidswaarde van de relevante claims geacht wordt te bepalen. Een gedegen verklaring van *a priori* kennis vraagt om een adequate conceptie van de betekenis van *a priori* claims en de aard van *a priori* waarheden.

In de filosofie van de wiskunde is de wederzijdse afhankelijkheid van enerzijds theorieën over betekenis en waarheid en anderzijds theorieën over kennisverwerving reeds lang gemeengoed. Dit komt in grote mate door twee baanbrekende papers van Paul Benacerraf, gepubliceerd in 1965 en 1973, die praktisch elke denker die sindsdien over deze onderwerpen heeft geschreven hebben beïnvloed.

De tweede paper ontvouwt een dilemma dat gezien kan worden als de bron van een groot aantal discussies en ontdekkingen in vroeg 20e-eeuws werk op het gebied van de grondslagen van de wiskunde. Het dilemma is het volgende: Als we volhouden dat de waarheidswaarde van onze wiskundige overtuigingen bepaald wordt door het oordeelsonafhankelijk (niet) vervuld zijn van de abstracte condities, waarover deze overtuigingen beweren te gaan (i.e. of bepaalde bestaande wiskundige objecten bepaalde wiskundige eigenschappen hebben), dan blijken we niet in staat te begrijpen waarom we, middels onze natuurlijke cognitieve mechanismen, kunnen ontdekken of deze condities (niet) vervuld zijn. Als we, anderzijds, veronderstellen dat de verwerving van wiskundige kennis tot stand komt door een adequaat causaal contact tussen

de kennende geest en de vervulde waarheidscondities van ware overtuigingen, dan lijken we gedwongen te concluderen dat de waarheidscondities van onze gevestigde wiskundige theorieën niet begrepen kunnen worden langs de bovengenoemde referentialistische lijnen. Met andere woorden, in de filosofie van de wiskunde lijkt onze standaard referentialistische conceptie van waarheid niet te rijmen met onze standaard causale theorie van kennis.

De betekenis van Benacerrafs constateringen blijft niet beperkt tot de filosofie van de wiskunde. Vergelijkbare vragen kunnen gesteld worden in de semantiek van elk discours waarin we geacht zijn kennis te verwerven over causaal inerte onderwerpen.

Volgens Benacerraf is de voornaamste reden waarom filosofen zouden moeten vasthouden aan het (in grote lijnen Tarskiaanse) referentialistische waarheidsbegrip dat deze theorie de enig beschikbare, gearticuleerde en *prima facie* plausibele algemene conceptie van waarheid is. Bij het ontbreken van een dergelijke conceptie is de veronderstelling dat bepaalde condities de *waarheids*condities zijn (in tegenstelling tot, bijvoorbeeld, de condities van rationele aanvaardbaarheid) van bepaalde overtuigingen nauwelijks meer te onderbouwen.

Gezien dit conceptuele probleem en daarnaast bepaalde verklaringsproblemen, zijn de meeste filosofen heden ten dage ervan overtuigd dat het juiste antwoord op Benacerrafs dilemma aan de standaard referentialistische conceptie van waarheid moet vasthouden. Mijn voornaamste doel in deze studie is te laten zien dat die overtuiging onjuist is.

De studie is verdeeld in zeven hoofdstukken. De eerste twee hoofdstukken bevatten voorbereidend materiaal voor de argumentatie die in de overige vijf ontwikkeld wordt. De voorbereidende hoofdstukken zijn bedoeld om de voornaamste conceptuele en methodologische vooronderstellingen van de daarna volgende redeneringen te verhelderen. Het argumentatieve deel begint met hoofdstuk 3, dat Benacerrafs dilemma presenteert en de theoretische opties schetst die men zou kunnen kiezen in reactie daarop. De overige vier hoofdstukken van deze studie zijn

gewijd aan de evaluatie van de *prima facie* levensvatbare kandidaten voor een antwoord. Hoofdstukken 4, 5 en 6 bevatten mijn argumenten tegen de bestaande referentialistische antwoorden (d.w.z. deflatoir referentialisme, anti-realistisch referentialisme en realistisch referentialisme), terwijl hoofdstuk 7 zich concentreert op de uitwerking en verdediging van het voorgestelde non-referentialistische alternatief.

De centrale claim van de voorgestelde non-referentialistische conceptie is dat de waarheidsvoorwaarden (in tegenstelling tot de beoogde referenten) van onze paradigmatische *a priori* (i.e. logische en wiskundige) overtuigingen natuurlijke toestanden in mensenhoofden zijn. Meer specifiek zijn ze analytische relaties tussen representaties die subjecten ontwikkelen in hun hoofden gedurende de cognitieve omgang met hun directe natuurlijke omgeving. In het licht van deze veronderstelling, noem ik het voorstel een *representationistische* conceptie van de waarheden in kwestie. Deze conceptie behoudt het realistische idee dat de waarheidswaarde van logische en wiskundige overtuigingen onafhankelijk is van enige daadwerkelijke opvatting ten aanzien daarvan. Hij onderschrijft de gematigd empiristische claim ten aanzien van de analytische aard van logische en wiskundige overtuigingen, maar identificeert de feitelijke basis van analytische waarheden in de werkelijke, natuurlijke wereld. Een essentiële component van deze conceptie is de veronderstelling dat een subject het (niet) vervuld zijn van de relevante representationele condities in haar hoofd kan bespeuren zonder er daadwerkelijk aan te denken. Hierdoor kunnen deze condities fungeren als non-referentiële waarheidscondities voor de overtuigingen. Ten slotte kan het (niet) vervuld zijn van de condities, omdat die natuurlijk zijn, de cognitieve capaciteiten van het subject causaal beïnvloeden, wat betekent dat de constructie ook compatibel is met de algemeen aangenomen "contacttheorie" van kennisverwerving. Waarheid en kennis in zuivere logica en wiskunde blijken natuurlijke verschijnselen, die onderworpen kunnen worden aan systematisch empirisch onderzoek.

Na de verdediging van deze specifieke non-referentialistische conceptie van de paradigma's van *a priori* waarheid en kennis, betoog ik ten slotte dat het doorslaggevende kenmerk van representionaliteit meer dan een contingente eigenschap van *a priori* waarheden kan zijn. In feite hebben we goede redenen om dit kenmerk te zien als een essentiële, bepalende trek van aprioriciteit, en dus als een noodzakelijke eigenschap van *a priori* waarheden. Volgens de resulterende representationalistische conceptie van aprioriciteit, is een bepaalde waarheid *a priori* (i.e. *a priori* kenbaar) te noemen, precies dan wanneer die bestaat in het (niet) vervuld zijn van bepaalde condities in het domein van de representaties in het hoofd van een subject, terwijl een bewijs, de rechtvaardiging die dat verschaft voor een overtuiging die gebaseerd is op de herkenning ervan en de kennis verworven door die rechtvaardiging, *a priori* genoemd kunnen worden, precies dan wanneer die gegenereerd zijn door een (in principe) te observeren cognitief mechanisme in het hoofd van een subject, dat betrouwbare informatie overbrengt over het (niet) vervuld zijn van representationele waarheidscondities in de geest van het subject.

Na het afronden van de argumentatie geef ik in de conclusie een kort overzicht van de voornaamste beweringen in deze studie en een uitleg van wat de hier bijeengebrachte bevindingen ons kunnen vertellen over naburige kwesties in de huidige literatuur en over de rol van empirisch onderzoek in de ontwikkeling van concepten van groot filosofisch belang.

CURRICULUM VITAE

Zsolt Novák werd op 6 juli 1970 geboren in Boedapest. In 1988 behaalde hij zijn diploma op het Berzsenyi Dániel Gimnázium. Na de militaire dienst in het Hongaarse leger heeft hij eerst architectuur gestudeerd aan de Technische Universiteit van Boedapest. Vervolgens studeerde hij aan het Institut für Waldorf-Pädagogik, Witten-Annen, Duitsland, in 1990. Na zijn terugkeer in Boedapest studeerde hij geschiedenis, taalkunde en wijsbegeerte aan de Eötvös Loránd Universiteit. Hij behaalde daar een MA-diploma in de wijsbegeerte in 1999 en een MA-diploma in de geschiedenis in 2006.

Novák begon in 2000 aan een promotietraject in de wijsbegeerte aan de Central European University (CEU) te Boedapest. In 2000 en 2001 bracht hij acht maanden door aan het Collegium Hungaricum in Wenen, waar hij onderzoek deed naar het logisch empirisme aan het Institut Wiener Kreis. In 2002 ontving hij een Chevening Scholarship van de Britse overheid om filosofie van de wiskunde te studeren aan de University of Oxford. In 2004 verhuisde hij zijn project naar de Faculteit der Wijsbegeerte (tegenwoordig het Instituut voor Wijsbegeerte) van de Universiteit Leiden, waar hij het promotieonderzoek voortzette met een Huygens Scholarship in 2004-2005 en een aanstelling als Assistent in opleiding in 2005-2006. Hij was in 2005 medeorganisator van de eerste Oxford-Budapest Conference over waarheid, referentie en realisme, die plaatsvond aan de CEU. In 2006 verbleef hij voor onderzoek in New York met een beurs van het Leids Universiteits Fonds. Hij ontving een Kathleen Wilkes Scholarship ten behoeve van deelname aan de 2007 Annual Philosophy of Science Conference, gehouden aan het Inter-University Centre, Dubrovnik. In 2008 en 2009 werkte hij als onderzoeksassistent aan het Netherlands Institute for Advanced Study te Wassenaar. Hij is co-editor van *Truth, Reference and Realism* (Boedapest: CEU Press, in druk). Hij heeft Hongaarse vertalingen gepubliceerd van verscheidene Engelse en Duitse filosofische werken.