THE CONTEXT PRINCIPLE AND IMPLICIT DEFINITIONS : TOWARDS AN ACCOUNT OF OUR A PRIORI KNOWLEDGE OF ARITHMETIC

Philip A. Ebert

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews

2005

Full metadata for this item is available in
St Andrews Research Repository
at:
http://research-repository.st-andrews.ac.uk/

Please use this identifier to cite or link to this item:
http://hdl.handle.net/10023/14916

This item is protected by original copyright

The Context Principle and Implicit Definitions

Towards an account of our a priori knowledge of arithmetic.

Philip A. Ebert

Submission Date: 9th of September 2005

Submission for a PhD in Philosophy at the University of St Andrews

BIBLIOTHECA UNIVERSITATIS ANDREANÆ

ProQuest Number: 10170996

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

ProQuest 10170996

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.
This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Th F109

Abstract

This thesis is concerned with explaining how a subject can acquire a priori knowledge of arithmetic. Every account for arithmetical, and in general mathematical knowledge faces Benacerraf's well-known challenge, i.e. how to reconcile the truths of mathematics with what can be known by ordinary human thinkers. I suggest four requirements that jointly make up this challenge and discuss and reject four distinct solutions to it. This will motivate a broadly *Fregean* approach to our knowledge of arithmetic and mathematics in general. Pursuing this strategy appeals to the context principle which, it is proposed, underwrites a form of platonism and explains how reference to and object-directed thought about abstract entities is, in principle, possible. I discuss this principle and defend it against different criticisms as put forth in recent literature. Moreover, I will offer a general framework for implicit definitions by means of which – without an appeal to a faculty of intuition or purely pragmatic considerations – a priori and non-inferential knowledge of basic mathematical principles can be acquired. In the course of this discussion, I will argue against various types of opposition to this general approach. Also, I will highlight crucial shortcomings in the explanation of how implicit definitions may underwrite a priori knowledge of basic principles in broadly similar conceptions. In the final part, I will offer a general account of how non-inferential mathematical knowledge resulting from implicit definitions is best conceived which avoids these shortcomings.

Declarations

I, Philip A. Ebert, hereby certify that this thesis, which is approximately 79.300 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

date: 3/12/05
signature of candidate:

I was admitted as a research student in September 1999 (for an M.Litt) and as a candidate for the degree of PhD in September 2000; the higher study for which this is a record was carried out in the University of St Andrews between October 2000 and August 2005

date: 3/12/05
signature of candidate:

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit the thesis in application for that degree.

date: 3/12/05
signature of supervisor:

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any bone fide library or research worker.

date: 3/12/05
signature of candidate:

Acknowledgments

My greatest thanks is to Crispin Wright. His comments and criticisms of earlier drafts and his support throughout the process of writing this thesis are greatly appreciated. Yet, not only am I very grateful for his philosophical comments but also for bringing about *Arché*, the AHRC – Centre for the Philosophy of Logic, Language, Mathematics and Mind. This centre provided a research environment next to none. I don't think that this thesis would have the form it has without the constant interaction and philosophical discussion with the many other PhD students, postdoctoral fellows, visiting professors and scholars at *Arché*.

More specifically, I would like to thank Roy Cook for reading and commenting on various parts of this thesis, and for being such a *cool* shadow-supervisor in my first year. I look forward to extending the Cook/Ebert collection! Carrie Jenkins for reading and commenting on the first chapter of the thesis. Patrick Greenough for making me trust my instincts and offering valuable comments on the fifth and early drafts of the sixth chapter of the thesis. Fraser MacBride for the many conversations during my first two years that helped shaping up chapter 1 and 2. Cyrus Panjvani who in the first two years of my PhD was a great companion and with whom I had some great philosophy 'chats'. Nikolaj Pedersen for his comments during the last 4 years and being a great office mate. Marcus "Kollege" Rossberg who read and commented on parts of the thesis. He was always ready for a chat and provided many comments that were always greatly appreciated. Stewart Shapiro who offered great comments and very valuable supervision during his time in St Andrews. Also, I would like to thank him for inviting me to Ohio State University where I stayed during Spring of 2003. Great thanks to Simon Robertson with whom I passed the M.Litt. in St Andrews. Despite his weakness for moral philosophy, I would like to thank him for the many philosophy conversations and his companionship during our time in St Andrews. Other members that I would like to thank for providing valuable comments during my time in Arché are: Elizabeth Barnes, Eline Busck, Ross Cameron, Aviv Hoffmann Darren McDonald, Josh Parsons, Agustín Rayo, Sonia Roca, Chiara Tabet, Robert Williams, Elia Zardini.

In addition to these specific thanks, I would like to thank all St Andrews Philosophy PhD students who created a community of young researchers that made the last five years a great experience. They all, not only contributed to making this philosophy department a highly fruitful research environment, but also, equally important, made it a very fun place to live! I shall also like to thank the audiences at Berlin, Lisbon, Nottingham, Ohio, Oviedo, St Andrews, Stirling, and Syracuse where papers from this thesis where presented.

I want thank, in addition, all my non-philosophy friends who, in the last five years, kept me sane: My old friends from Heidelberg, Felix Kaup, Johannes Schmitt-Widdersheim, Tilman "Wargi" Schubert and many others; and my climbing friends in Scotland, especially Francois Bocquet, the late Al Mathewson and Al, Guy and Simon Robertson.

Lastly, I wish to thank Lindsey Nicholson and my family, Klaus, Rumiana, Natalie Ebert and now wee little Karl. Their support has been much more valuable than they will think. This thesis is dedicated to them.

Contents

Introduction

This thesis is concerned with explaining how a subject can acquire a priori knowledge of arithmetic.[1] The first chapter is intended to outline and introduce the main topics of this thesis and to exhibit interconnections between the philosophy of mathematics and language, epistemology and metaphysics. To this end, I will first outline *Benacerraf's dilemma* as put forth in Paul Benacerraf's famous paper "Mathematical Truth". I will discuss and further clarify the semantic and epistemic constraints Benacerraf imposes, and show how they give rise to a dilemma that affects any approach in the philosophy of mathematics. I will then review four strategies of reply to this dilemma as they occur in the philosophical literature. This overview of various replies will help to elicit what I label the *fundamental assumption* which those four strategies all presuppose and whose rejection will help to give rise to a different type of platonistic response. At this point I will aim to show that Frege himself struggled with a kind of Benacerraf problem in his *Grundlagen der Arithmetik* and that his adoption of the context principle involves a rejection of the fundamental assumption. Hence, I believe that Frege himself would be sympathetic to this type of platonism and so I will label it *Fregean platonism*.

After having briefly outlined Frege's context principle and various interpretations of the

[1]Although the topic of this thesis is concerned with knowledge of arithmetic, I hope to show that this approach can be generalised to other mathematical theories.

context principle, I will focus on how contemporary *Neo-Fregeanism* adopts the context principle with the aim of overcoming Benaceraff's dilemma. The aim in the sequel is to develop this general approach and to assess various epistemic and metaphysical problems.

So, in the second chapter, I will be focussing on various types of criticisms of the context principle as it is used by the Neo-Fregeans. Although I won't be concerned with convincing the unconvinced of the cogency of this interpretation of the context principle, I hope to provide at least an internal justification for the context principle, by rebutting various arguments that claim to show that the principle, so understood, is either incoherent or else leads to unwelcome consequences.

In the third chapter, I will then focus on the idea of implicit definitions and the notion of stipulation. Here, I will first distinguish three aims, or *dimensions of achievement*, for implicit definitions which have been confused in current literature. Following this explanation, I will state and assess various necessary conditions one should impose on an implicit definition. The conditions that have to be fulfilled depend upon what the achievement of the definition is intended to be. Having provided this framework for implicit definitions, I will finally outline the so-called *traditional connection*: the view that implicit definitions can found a priori knowledge of first principles of mathematics and logic. To this end, I will explain a further epistemic dimension for implicit definitions which, according to the proponent of the traditional connection and the Neo-Fregean, will underwrite our knowledge of logic and mathematics and thereby meet Benacerraf's epistemic challenge.

The fourth chapter is concerned with a specific type of opposition to the traditional connection called *rejectionism*. I will here distinguish two types of rejectionism. The first I label *semantic* rejectionism, a position that is adopted by Paul Horwich. This conception rejects *tout court* the idea that implicit definitions can have ontological commitments and so any mathematical principle that entails the existence of infinitely many objects can't be regarded as an acceptable implicit definition. The other conception which is defended by Hartry Field and George Boolos, I label *epistemic* rejectionism. This position accepts that stipulated sentences – stipulated for the purpose of implicit definitions – can have ontological commitment. However, epistemic rejectionism repudiates the thought that notwithstanding such commitments this type of implicit definition can still be considered analytic, or a priori. I will discuss the two positions in turn and evaluate various responses to them. Finally, having evaluated these replies, I will offer a new argument that aims to destabilise the rejectionist position.

The fifth chapter is then concerned with explaining how a proponent of the traditional connection may aim to account for a thinker's a priori knowledge of logic and arithmetic by appeal to implicit definitions. The main issue is the demand for an explanation of how exactly a subject arrives purely in virtue of grasping an implicit definition at a *warranted belief* in the principle in question. This chapter will focus on one such explanation as originally proposed by Paul Boghossian. I will argue that his model ultimately fails, since his explanation commits a *transmission of warrant failure.* In addition, I argue that this worry generalises and affects other models of explanation, such as Bob Hale and Crispin Wright's.

In the final chapter I will outline a general account of how non-inferential a priori knowledge resulting from implicit definitions is best conceived while avoiding the previously highlighted shortcoming of the other approaches. To this end, I offer a model for non-inferential a priori knowledge and discuss whether it is best placed within either externalist or internalist conceptions of knowledge, or whether Crispin Wright's recent proposal of *entitlement* should be invoked in order to accommodate this model. All three conceptions face various difficulties that will be highlighted in the course of this last chapter. These considerations give rise to a proposal – a broadly externalist conception that appeals to entitlements at the level of *claims* to knowledge – which, I argue, is best suited for an account of non-inferential a priori knowledge of basic mathematical principles.

Chapter 1

Benacerraf's dilemma and Frege's context principle

Introduction

The following chapter is intended to outline and introduce the main topics of this thesis and to exhibit interconnections between the philosophy of mathematics and language, epistemology and metaphysics. To this end, I will first outline *Benacerraf's dilemma* as put forth in Benacerraf's famous paper "Mathematical Truth".[1] I will discuss and further clarify the semantic and epistemic constraints Benacerraf imposes, and show how they give rise to a dilemma. I will then review four strategies of reply to this dilemma as they occur in the philosophical literature. The first two platonistic strategies comply with the semantic constraint but provide insufficient answers to the epistemic constraint, while the other two, nominalistic strategies either reject the idea of mathematical knowledge altogether or fail the semantic constraint. This overview of various replies will help to elicit what I label

[1](Benacerraf, 1973)

the *fundamental assumption* which those four strategies all presuppose and whose rejection will help to give rise to a different type of platonistic response. At this point I will aim to show that Frege himself struggled with a kind of Benacerraf problem in his *Grundlagen der Arithmetik*[2] and that his adoption of the context principle involves a rejection of the fundamental assumption. Hence, I believe that Frege himself would be sympathetic to this type of platonism and so I will label it *Fregean platonism.*

In section 1.3, I will briefly outline Frege's context principle and various interpretations of the context principle in the light of what Frege took to be the role of the context principle in the *Grundlagen.* In section 1.4, I will focus on how contemporary *Neo-Fregeanism* adopted the context principle with the aim of overcoming Benacerraf's dilemma as outlined before and thereby giving rise to a new version of Fregean platonism. My overall aim in this thesis is to develop this type of platonism and to assess various epistemic and metaphysical problems as they occur in the literature. So, this first chapter is intended, on the one hand, to introduce the favoured position and, on the other hand, to provide a framework of philosophical problems this position encounters.

1.1 Benacerraf's dilemma

In his seminal paper "Mathematical Truth"[3] Benacerraf outlines a dilemma that every account within the philosophy of mathematics faces. The dilemma arises from the need for any such conception to satisfy two constraints whose mutual resolution, however, seems impossible. The first constraint concerns the semantic theory adopted for our mathematical discourse. Here the demand is to have a "homogenous semantical theory in which the se-

[2](Frege, 1884). I will throughout refer with '*Grundlagen*' to this book.

[3]See (Benacerraf, 1973) all unspecified references in this chapter will be to the paper as published in (Benacerraf and Putnam, 1983).

mantics for the statements parallel the semantics for the rest of the language." (p.403) The second constraint is epistemological in nature and demands that "the account of mathematical truth mesh with a reasonable epistemology." (p.403) Benacerraf's dilemma, in its simplest form, arises when one attempts to conform to those two constraints by appeal to the prevalent views in semantics and epistemology: The standard view to comply with the first constraint – a Tarskian theory of truth – imposes an ontology which is incompatible with the standard view in epistemology, which, at the time of Benacerraf's publication[4], comprised causal constraints on knowledge of any type of object. In order to explain the dilemma and the resulting challenge more clearly, I will have a closer look at the two main constraints which each involve two components.

Although not explicitly in his writing, Benacerraf seems to argue for two requirements that jointly make up the semantic constraint.[5] The first requirement is that a theory of truth should be adopted that applies to any discourse, be it an empirical, a mathematical, or even an ethical discourse.[6] This requirement can be found in the following passage:

> "Another way of putting this first requirement is to demand that any theory of mathematical truth be in conformity with a general theory of truth – a theory of truth theories, if you like, – which certifies that the property of sentences that the account calls "truth" is indeed truth. This, it seems to me, can be done *only* on the basis of some *general theory for at least the language as a whole*." (p. 408, my italics)

The only candidate to satisfy this demand is, according to Benacerraf, a Tarskian account of truth, where truth is spelled out by appeal to reference and satisfaction. Here, the basic

[4]See for example (Goldman, 1967) or (Harman, 1973) as mentioned by Benacerraf.

[5]See especially (Hale and Wright, 2002) who distinguish these two issues. I draw on parts of their discussion of Benacerraf's dilemma.

[6]So a similar "Benacerraf's dilemma" can also be posed for ethical and modal discourse.

idea is to assign semantic values to the different semantic components of a sentence and to then correlate the truth or falsity of a sentence with those semantic values.[7]

The second requirement of the semantic constraint adopted by Benacerraf is to take the surface grammar of sentences of the discourse at face value. Consider his two examples:

1. There are at least three large cities older than New York.

2. There are at least three perfect numbers greater than 17.

If we take the surface grammar at face value we should regard both sentences as having the same structure, namely:

3. There are at least three F/G's that bear R to a.

Benacerraf reinforces this second requirement when he writes that "we should not be satisfied with an account that fails to treat (1) and (2) in parallel fashion, on the model of (3)." (p. 408) Hence, in addition to adopting a Tarskian – referential – theory of truth, the semantic constraint also involves the demand to respect the surface grammar and hence to take it without further qualification that what *seem* to be singular terms, such as "New York" and "17", *are* singular terms.

To sum up, the semantic constraint demands a general and systematic theory of truth to be applied to the mathematical language; and it also demands that we respect the surface-grammar of the mathematical discourse in so applying our theory of truth. Consequently, the *standard view* which satisfies these demands does interpret mathematics by analogy with empirical sentences. Both are analysed using predicates, quantifiers and singular terms and satisfaction-conditions for truth. Crucially, in the case of mathematics, number-terms are regarded as singular terms and as such are required to denote objects. In addition, the

[7]Much more can be said about a Tarskian theory of truth, but the details are not important here.

objects that number-terms refer to are intuitively regarded as non-spatio-temporally located and are thus abstract. The resulting position – the standard view – is one which genuinely deserves the label 'platonistic', as Benacerraf notes.

Importantly, the thought that singular termhood and its occurrence in presumably true sentences (it is implicit in Benacerraf's treatment that our mathematical language is truth-apt and that statements such as '$2 + 2 = 4$' are true) suffices for claiming that there is an object that is the referent of the term (and presumably a sui-generis one in the case of number-terms) is a substantial claim. Also, it mirrors considerations that were appreciated by Frege in his *Grundlagen* and is part of what we shall be regarding as the minimal content of the context principle. We will come back to these issues in section 1.3 and more prominently in 1.4.

The merit of the standard view for the semantic constraint is, according to Benacerraf, that the semantics of mathematics nicely meshes with that of other discourses. Consequently, mixed discourses – where mathematical and empirical terms are used in the same context – pose no additional problem. Further to these virtues of parsimony and simplicity, Benacerraf notes that a Tarskian theory of truth "is the only viable systematic general account we have of truth" in order to account for the first requirement. Finally, note that taking the surface-grammar at face value keeps the truth-conditions of mathematical discourse in line with the thoughts intuitively expressed by such sentences. Hence, respecting the surface-grammar guarantees that what *seems* to be the subject of our mathematical discourse *is* its subject and thus what *seems* to be known and thought about by the subject *is* what is known and thought about by the subject.

Let us now turn to the second *epistemological* constraint proposed by Benacerraf. At first sight, it merely comprises a minimal assumption that hardly seems controversial: we have mathematical knowledge, which "is no less knowledge for being mathematical" (p.409).

Subsequently, the demand of the second constraint is just that the account of mathematical truth be compatible with an account of knowledge that renders such truths knowable.

According to Benacerraf, the *standard way* of complying with this second constraint, and the only viable general account of knowledge, is a causal account. The idea, roughly, is that for a subject X to know that p, there has to obtain some type of causal relation between the subject X and the objects, or other items, involved in the subject matter of p. This account can be easily motivated by appeal to our knowledge of medium-sized, everyday objects, which does seem to involve causal relations.

However, adopting the standard view – a Tarskian theory of truth – to comply with the semantic constraint and adopting the standard way – causal theory of knowledge – to comply with the epistemological constraint seems to lead to an impasse, and brings *Benacerraf's dilemma* to the fore: If the truth-conditions of mathematical statements are given by the standard semantical account, then an arithmetical statement involving number terms should be regarded as making reference to numbers as objects that exist in an abstract realm. However, these objects seem unaccountable within the standard view in epistemology, as clearly no causal connection between an object of this sort and the presumably knowing subject can be made out. Thus, no explanation of how we can come to know that the truth-conditions of a mathematical statement obtain can be offered. Crucially, note that what Benacerraf outlines is a genuine dilemma in that *either* of the two constraints (or even both) can be relaxed – either of the two standard views can be dismissed and a new theory of truth or new epistemology can be put into its place. It is not a direct attack (as often thought to be) on the platonistic view of semantics.

Before I outline various general strategies for resolving the dilemma, I will briefly consider a misled interpretation of the epistemological constraint. This will lead to a strengthening of the simple version of the dilemma.

It would be an inadequate platonist response to the dilemma to merely reject the causal conception of knowledge in the light of recent criticisms. Certainly, the causal conception of knowledge does not always seem adequate for medium sized objects which provided its initial motivation, nor is it *prima facie* compatible with knowledge of the future or even facts that can be construed as involving arguably less suspect abstract objects such as the University of St Andrews, Dundee United F.C. or Apple Computers. Thus, so the platonist could argue, the causal theory of knowledge should not be adopted, since it clearly fails to be a simple and general account of knowledge that would qualify as the "only viable systematic general account we have of" knowledge. Consequently, Benacerraf's dilemma would vanish in its initial form since there are no fully general causal constraints on knowledge.

Although I am in general sympathetic to the idea that the causal theory of knowledge is insufficient as an account of our knowledge in general, this response, however, misses the crucial point of Benacerraf's dilemma. The dilemma, in its strongest form, not only points to an incompatibility between two standard views within two areas in the philosophy of mathematics: it also aims to highlight an *integration problem* in the philosophy of mathematics.[8] Accordingly, the epistemological constraint does not necessarily depend on adopting a causal account of knowledge. Rather, what is needed is an epistemological account that is able to *integrate* mathematical truths and thereby provide an *explanation* of how it is possible that mathematical truths, whose truth-conditions are spelled out using platonistic ontology (according to the the standard view of the semantic constraint), can be known to obtain by the subject.[9] Rejecting the causal account does little to *explain* how such truth-conditions can be *reliably known* to obtain.

[8]This terminology was first introduced by (Peacocke, 1999) who himself refers to Benacerraf's dilemma as a prototype of the integration problem.

[9]Note that a solution to the integration problem need not start with a semantic theory as I do here. It is open to start with an epistemology and then aim to integrate a semantic theory within it.

It is this stronger version of the Benacerraf dilemma – in the form of an integration problem – which will be the focus of the following sections. But, for clarity, I believe we should distinguish two requirements towards providing a fully integrated account. An epistemology for mathematics – which involves abstract objects – not only needs to explain how we know that the truth-conditions of a mathematical statement obtain, but, in addition, it also needs to explain how we can have (directed) beliefs about the abstract objects that make the mathematical statements true. To put the point differently, it needs to explain how in thought we can have access or refer to objects that are not spatio-temporal.[10]

This idea can be further explicated by considering the following passage of Hartry Field's interpretation of the dilemma,[11] which also dispenses with the purely causal constraint. He writes:

> "Benacerraf's challenge – or at least, the challenge which his paper suggests to me – is to provide an account of the mechanisms that explain how our beliefs about these remote entities can so well reflect the facts about them." ((Field, 1980), p. 26)[12]

Here, it seems to me, two issues are at hand in the challenge. The first issue is to explain how, in principle, we can successfully talk or have beliefs about abstract objects in the first place. The second issue is to explain, more specifically, how we are justified in thinking that those mathematical beliefs are true, or, as Field would put it, how those beliefs reliably track

[10]The question of what our mathematical beliefs are about is underlying Benacerraf's earlier discussion in (Benacerraf, 1965).

[11]See (Field, 1980) especially the introduction to this book, p. 20-30

[12]Dialectically there is a difference between Field's presentation and mine. He continues the above quote in the following way: "The idea is that *if it appears in principle impossible to explain this*, then that tends to *undermine* the belief in mathematical entities, *despite* whatever reason we might have for believing in them." (op.cit.) I don't think that this is how Bencacerraf's dilemma should be understood. I believe it is a genuine dilemma while Field regards it more or less as a challenge to the platonist.

the mathematical facts.[13]

In conclusion to these considerations, I think we can now see that there are four separate requirements (the first two involved in the semantic constraint and the other two in the epistemological constraint) that collectively make up Benacerraf's dilemma and that collectively need to be addressed:

1. *Homogeneous semantic theory*

   The demand that we adopt a general and systematic theory of truth, which – for Benacerraf – should be a Tarskian Theory of truth.

2. *Surface-grammar*

   The demand to respect the surface grammar of mathematical discourse.

3. *Reference and object-directed thought*

   The demand to explain how the objects posited by the semantic theory can, in principle, be in the range of directed thought and talk of the subjects.

4. *Knowledge*

   The 'integration challenge': The demand to reconcile the truths of the subject matter with what can be known by ordinary human thinkers. Crucial here is to provide an explanation of how a subject can have mathematical knowledge and on what basis the subject can claim such knowledge.[14]

[13] At this stage I don't read too much into the notion of *reliability* which Field uses, but again more on this in later chapters.

[14] These issues – knowledge and claiming knowledge – will become prominent in chapter 6.

1.2 Strategies to resolve Benacerraf's dilemma

In this section I will review four different strategies to resolve Benacerraf's dilemma. I will begin with a version of platonism which regards mathematical knowledge as a special kind of knowledge that has its own special source and so is distinct in kind from knowledge of other subject matters.

1.2.1 *Intuitive* platonism

Intuitive platonism adopts the standard view of the semantic constraint, conceiving of the mathematical language as referring to self-subsistent, abstract mathematical entities, and also respects the surface-grammar of the mathematical discourse. Our knowledge of such 'remote' entities is explained by the fact that in the case of mathematics we are concerned with a special type of knowledge, which in the relevant respect is *basic*. The idea here is to break with the demand that a *generally* applicable account of knowledge is needed for every discourse. Instead, the axioms of mathematics and the rules of inference – from which the theorems of mathematics are derived – are regarded as basic in the sense that they cannot be inferred from, and so be known in virtue of, even more fundamental principles. Rather, a subject's *non-inferential* knowledge of the axioms and the rules of inference has its source in the special faculty of *intution*, which, similar to the faculty of perception, provides direct knowledge of the truth of the basic axioms.

The main proponents of intuitive platonism within the philosophy of mathematics have been Kurt Gödel and more recently Charles Parsons[15]. To characterise this type of platonism more precisely, the following often-cited quotation from Gödel highlights the role intuition is supposed to play here:

[15]For example in (Parsons, 1979).

> "But despite their remoteness from sense-experience, we do have something like a perception of the objects of set theory, as is seen from the fact that the axioms force themselves upon us as being true. I don't see any reason why we should have less confidence in this kind of perception, i.e. mathematical intuition, than in sense-perception." (Gödel, 1947), p.483-4

This version of platonism, however, faces various difficulties in providing a satisfactory answer to the two epistemic requirements in Benacerraf's dilemma. For one thing, note the transition in this quotation from knowledge of objects ("the objects of set theory") with which mathematical intuition is concerned to knowledge of the truths of axioms ("the axioms force themselves on us") which Gödel aims to underwrite by the faculty of intuition. Leaving aside what underwrites this transition, it remains unclear what the mark of a successful intuitive grasp of such an abstract object is. Just claiming that we can perceive these objects and thereby regard the mathematical axioms as intuitively compelling, obvious, or as somehow "forcing themselves upon us", seems insufficient as a genuine justification of our cognitive beliefs in the truth of the mathematical axioms.

To highlight this further, consider in analogy the scenario in which we postulate a faculty of "perceiving other minds" which provides immediate knowledge of other minds and thereby accounts for the obviousness or intuitive compellingness of certain beliefs about them. No-one (in their right mind) would regard this as a sufficient explanation and justification of our beliefs about other minds.

Also, and connected to this weakness, intuitive platonism has to be able to account for the *fallibility* of this faculty. It would be insufficient to say that Frege's faculty of intuition did let him down when he postulated Basic Law V as an axiom, without explaining why it did and why it does not in other cases where the target statements are consistent.

Various additional concerns could be raised but what seems to be at the heart of most criticisms[16] is that a postulated faculty of intuition fails to provide a genuine *explanation* of our access to and knowledge of abstract entities, since it is just built into the faculty of intuition – as a brute fact – that it does enable such access and knowledge. Hence, crudely put, this type of platonism seems more like an acknowledgement of the inability to provide an explanation of our knowledge of mathematics than a genuine solution.[17]

In addition, the *intuitive* platonist also runs the risk of dislocating our mathematical knowledge from everyday and, more crucially, scientific knowledge. How on this view is mathematical knowledge embedded and interactive within the scientific corpus of knowledge? It is this version of the well-known *application problem*, which concerns how mathematical knowledge is applicable in empirical science, that proves especially challenging for the *intuitive* platonist. And exactly at this point is where a new type of platonism can be located, one which rejects both the view that mathematics has it's own epistemology, and that it gives rise to a special sort of *a priori* knowledge.

1.2.2 *Naturalised* platonism

Naturalised platonism, whose principal author is Quine[18], regards mathematical knowledge as being on a par with scientific knowledge. So, mathematical knowledge is part of our theoretical knowledge, just like knowledge of physics or chemistry, and the objects of mathematics are theoretical objects just like electrons, neutrinos or strings are theoretical objects posited by the physical theories. Therefore, there is neither the need for a special faculty of intuition to explain mathematical knowledge nor does mathematical knowledge enjoy a

[16]See for example (Hale and Wright, 2002) for extensive criticism of Parson's and Gödel's views.

[17]There is however a recent resurrection of the idea of a faculty of intuition in current epistemology in the works of BonJour in his (BonJour, 1998) and (Sosa, 2005), which I won't be able to cover here.

[18]See his (Quine, 1986) but there are various others who hold similar positions, such as (Resnik, 1997) and (Shapiro, 1997)

special status – as a type of a priori knowledge.

But, crucially, then how does the naturalised platonist explain mathematical knowledge, even if it is merely theoretical? After all, this version of platonism also adopts a Tarskian semantic theory to comply with the semantic constraint of Benacerraf's dilemma and, so, regards the objects of mathematics as abstract objects.

This integration challenge has received an answer by the naturalised platonist in the form of the well-known *Quine-Putnam* indispensability argument.[19] The argument can be presented as follows:

Premise 1 Mathematics is *indispensable* to our scientific theories, in that they can neither be formulated nor practised without mathematical vocabulary and inferences.

Premise 2 If mathematics is indispensable to our accepted scientific theories, then if those scientific theories are true then the mathematics involved in scientific theorising is true.

Intermediate Conclusion 1 If scientific theories are true then the mathematics involved in scientific theorising is true.

Premise 3 Scientific theories are true.

Intermediate Conclusion 2 The mathematics involved in scientific theorising is true.

Premise 4 If mathematics is true, then there are the abstract entities to which it purportedly refers, such as numbers, functions, sets.

Conclusion Abstract entities, such as numbers, functions and sets that are appealed to in mathematical theories which are involved in scientific theorising, exist.

[19]Locus classicus is (Putnam, 1971). For an extensive discussion and recent defence of the indispensability argument, see (Colyvan, 2001)

It is in virtue of these pragmatic considerations that the *naturalised* platonist aims to incorporate his platonist conception of mathematics within a naturalised epistemology, whereby all knowledge is merely empirical. Obviously crucial here is that mathematics actually *is* indispensable to science in the relevant respect – a claim that has been challenged by Field, and which will be discussed in the next section. But even granting that mathematics is indispensable, two issues remain: Firstly, is a conception for which every statement is empirical and as such "up for revision" stable and, secondly, if it is stable, how exactly do these pragmatic considerations resolve the two epistemic challenges?

The first question has received much discussion in recent years. (Wright, 1986) argues that a Quinean position – which is a form of global empiricism – is intrinsically incoherent. The argument itself is very intricate and I don't propose to discuss it here. The second question, however, is more pertinent to the current discussion and concerns the adequacy of the *naturalised* platonist answers to the epistemological issues about our knowledge of mathematical objects.

Just like the *intuitive* platonist, the *naturalised* platonist does not provide much in terms of an *explanation* of our access to, or knowledge of, abstract objects. The indispensability argument might at best provide an argument *that* we are justified in thinking that there are abstract objects conceived of as theoretical entities. However, note that on this perspective a plausible element of mathematical thinking is lost. As Frege noted, "in arithmetic we are concerned with objects that we come to know not as something alien, from without through the medium of the sense, rather they are directly posited to reason, which, as its nearest kin, it can completely grasp." ((Frege, 1884), §105 my translation[20]) Hence, the "charm" of

[20] Austin's official translation writes: "In arithmetic we are not concerned with objects which we come to know as something alien from without through the medium of the senses, but with objects given directly to our reason and, as its nearest kin, utterly transparent to it." The differences of detail here are irrelevant to the point I'm making.

mathematics as "the reason's proper study" is lost from a naturalised platonist perspective, as knowledge of mathematics and knowledge of the objects of mathematics is merely justified indirectly by its involvement in scientific theories that are true.

Moreover, it is worth noting that knowledge of mathematics is parasitic upon (global) scientific realism (premise three) which is needed to arrive at the conclusion that mathematics is true. And, even granting the soundness of the indispensability argument, only a small part of mathematics will be justified by its application in science and thus the line between applied and pure mathematics becomes of crucial epistemological significance, since only the former, and not the latter (provided the above indispensability argument is all we have), can be justifiably regarded as true.[21]

Thus, one misgiving about naturalised platonism is that, even granting for the moment that the indispensability explains how a subject can have knowledge of parts of mathematics, this position is failing to address Benacerraf's dilemma in full generality. The challenge is how mathematics in general can be integrated into a thinker's corpus of knowledge, and not how some parts can be so integrated. So, I think a position that aims to explain all of mathematics and so tackles Benacerraf's dilemma in full generality is what is needed and desirable. Also, I find it hard to regard the indispensability of mathematics as an adequate explanation of our knowledge of mathematical entities and mathematical statements. Appealing to the need for mathematics in a presumably truth-apt scientific discourse does not – so at least it seems to me – provide the *right type* explanation of a thinker's access to and knowledge of mathematical entities.[22] Lastly and rather worringly, the very idea that

[21]There are further issues, for example how and if at all classical logic could be justified in virtue of application in science. See for example (Shapiro, 2005) for such discussion.

[22]A naturalised platonist will respond that asking for further explanations over and above provided by the indispensability argument is to ask for too much! Hence, the naturalist platonist won't be moved (much at least) by the presumed explanatory shortcomings of his position. I will have to grant this possible reply but I still believe that from a neutral ground there is prima facie a lack of epistemic explanation of our knowledge of abstract objects.

mathematics is *indispensable* to science is challengeable (and we shall review this challenge to naturalised platonism in the next section). Therefore, I think that these misgivings raise serious doubts as to whether the naturalised platonist position provides an adequate response to Benacerraf's dilemma.

To summarise, both intuitive and naturalised platonist positions suffer from an inability to provide satisfactory answers to the epistemic issues of Benacerraf's dilemma. Both adopt the standard view to comply with the semantic constraint, yet they fail to account for the epistemological challenges. These failures can be regarded as motivating two alternative positions next to be discussed. The first, *error-theoretic* nominalism, aims to avoid the problematic ontology, without thereby rejecting the standard view in semantics, by claiming that mathematics, taken at face-value, is actually false. This conception is – in some respects – in the tradition of *naturalised* platonism in that it regards all knowledge as theoretical, if it exists at all, but rejects the *indispensability* of mathematics to science, thereby leaving no possible theoretical reason to accept mathematical entities in the first place. The second conception I shall call *reconstructive* nominalism. It also accepts a Tarskian account, but regards the surface-grammar of the mathematical discourse as misleading. Mathematical ontology, accordingly, is not what it seems to be.

1.2.3 *Error-theoretic* nominalism

This strategy is famously proposed and explored by Hartry Field in several of his writings.[23] It can be motivated as a reply to the *naturalised* platonist who, according to the nominalist, has not gone far enough. What motivated *naturalised* platonism was the idea that we only have theoretical, i.e. empirical knowledge and thus no additional faculty has to be appealed to in order to account for mathematical knowledge. The *error-theoretic* nominalist adopts

[23]See his (Field, 1989) and (Field, 1980). Chapter 2 of (Field, 1982) provides a non-technical introduction.

the main feature of *naturalised* platonism – that there is only empirical knowledge – while additionally avoiding any commitments to abstract entities whose knowledge is hard to explain. .

As the *naturalised* platonist conception was characterised above, the *only* reason it provides for thinking that there are numbers, sets, etc is the previously outlined *indispensability* argument. Field regards this a valid argument but denies its soundness by rejecting the first premise – the claim that mathematics is indispensable to science. Crucially, Field works with an understanding of the underlying notion of indispensability which has to do not merely with the expressive resources gained by using mathematics, but also with the fact that mathematics is *essential* in establishing (proving) theorems and making predictions. Hence, to undermine the indispensability of current mathematical theory, it needs to be shown that there is an alternative theory that does equally well in establishing theorems and making predictions, but which does not involve commitments to numbers, sets, etc.

Field attempts exactly this and provides a framework that, according to him, does equally well, but that makes (arguably) no reference to abstract objects.[24] Consequently, Field rejects the indispensability argument and with it he rejects what the regards as the only good motivation to believe in abstract objects. Hence, he adopts an error-theory for standard mathematics in that, taken at face-value, mathematics is false since it has ontological commitments to things that we have no reason to believe exist.[25] Nevertheless, he thinks that we are still entitled to use a false theory and that it is desirable to do so since it is simpler and helps to "speed up inferences" so long as it is *conservative*. The relevant notion of

[24]Field is able to provide a theory without reference to mathematical objects that is able to capture Newtonian physics. Field's theory, however, makes reference to space-time points that, some have argued, are best regarded as abstract entities. I won't pursue this line of worry here.

[25]It can be challenged that this consideration suffices to establish an error-theory about mathematics. It seems rather that *agnosticism* should follow from this consideration. I will leave further discussion of this issue aside and grant that an error-theory about mathematics can be motivated in this way.

conservativeness is the following[26]:

Field's notion of conservativeness

"A mathematical theory M is conservative iff for any assertion A about the physical world and any body N of such assertions, A doesn't follow from N + M, unless it follows from N alone."((Field, 1982), p. 58)

The idea is that if mathematics is conservative, it is acceptable to use mathematics since it won't lead to any conclusions that could not be arrived at without mathematics. So, mathematics can be used to "speed-up" inferences or, in general, to make life easier for scientists without the need to endorse its truth and thereby its ontological commitments. Consequently, the basic notion for the *error-theoretic* nominalist is conservativeness and his credo is that "a mathematical theory must be conservative but need not be true".[27]

As can be expected of such a radical view, there is an extensive literature on Field's approach[28] that I won't attempt to survey here.[29] Rather I will assess how such an account

[26] In chapter 3 we will come back to this issue – see especially 3.5.2.

[27] (Field, 1982), p. 58.

[28] A chronological collection of the most important literature is: (Malamet, 1982), (Shapiro, 1983), (Shapiro, 1984), (Hale, 1987), as well as most of the papers in the collection (Irvine, 1990). For a very detailed survey of Field and his critics consult (MacBride, 1999) and for a detailed account and criticism of the technical framework see (Urquhart, 1990), as well as (Burgess and Rosen, 1997) who also provide a very nice reconstruction of the nominalism-platonism debate.

[29] One often-noted difficulty with Field's approach – which was first mentioned by Field himself – see his (Field, 1989) – and further pursued by (Shapiro, 1983) and mentioned in (Shapiro, 1984), concerns his condition of conservativeness and the consequence relation used in that condition. Field's contention is that nothing "additional" over and above what follows from the nominalistically accepted base theory N (above) alone should be a consequence of adding the mathematical theory. Shapiro challenges this point and argues that assuming the consequence relation is proof-theoretic, we can – in analogy to the proof of incompleteness results by Gödel – construct a sentence, call it G, in the nominalistically acceptable base theory that is not a theorem in that theory but that can be proved in that theory once one adds some mathematical resources. Hence, if the notion of logical consequence involved in the characterisation of conservativeness is considered proof-theoretically, then mathematics does not seem to be conservative. Alternatively, if the consequence relation is model-theoretic, then Field's position seems unstable. For he will presumably need to appeal to set-theory as model-theory which is not a nominalistically acceptable theory, in order to account for the consequence relation. This dilemma can surely be challenged and further rounds of the Field-Shapiro debate are to be found in (Field, 1985). Here Field suggests that G does not follow from the nominalistically acceptable theory.

would resolve Benacerraf's dilemma, assuming the more specific criticisms are resolved.

In some respects, Field's resolution of the dilemma is simple but radical: We are misled in thinking that mathematics is true – quite the opposite, it is false. Still it is valuable because it is conservative (or at least parts of it are). Just like the two types of platonism above, the *error-theoretic* nominalist adopts the standard view to address the *semantic constraint.* He adopts a Tarskian theory of truth and he respects the surface grammar – but denies the truth of the mathematical discourse, while granting its usefulness cashed out in terms of *conservativeness.* So, provided *error-theoretic* nominalism can overcome various technical difficulties in order to account for enough science, consistent with maintaining that mathematics is false but conservative, this position aims to overcome Benacerraf's dilemma by a radical route: denying mathematical knowledge altogether. Moreover, there is no need to account for a thinker's reference to and object-directed thought about mathematical entities.

However, is this a promising strategy to resolve Benacerraf's dilemma? I don't think so. A satisfying solution to the dilemma should not consist in giving up the basic assumption that we have mathematical knowledge. In a similar vein to my criticism to *naturalised platonism*, I think what is needed is a direct solution – taking Benacerraf's challenge head on – and integrate mathematical knowledge in general. Error-theoretic nominalism, in contrast, merely acknowledges defeat by dropping Benacerraf's main assumption that mathematical knowledge "is no less knowledge for being mathematical" (p.409). In turn, it seeks to explain why, in spite of there being no mathematical knowledge, we can pursue mathematics without a bad conscience. So, I think, in the context of a resolution to Benacerraf's dilemma, Field's approach is a mere last resort. I will therefore, leave aside *error-theoretic* nominalism and continue to explore the possibility for a direct resolution of Benacerraf's dilemma. *Reconstructive* nominalism is an alternative version of nominalism that shares the general scruples about abstract objects, but, less radically than Field, aims to retain the truth of mathematics, even if it is not truth in virtue of the properties of numbers, sets, functions, etc. This strategy will be discussed in the next section.

1.2.4 *Reconstructive* nominalism

The *reconstructive* nominalist view can also be regarded as a response to *naturalised* platonism. Yet, it adopts a different strategy to avoid the problematic ontology, which seems to pose a serious problem for our knowledge of mathematics and so for a satisfying solution to Benacerraf's dilemma. In contrast to *error-theoretic* nominalism that rejects the problematic ontology and the *indispensability* of mathematics, the *reconstructive* nominalist trades ontology for ideology. To explain, he rejects the following crucial move in the *indispensability* argument:

Intermediate Conclusion 2 The mathematics involved in scientific theorising is true.

Premise 4 If Mathematics is true, then there are the abstract entities to which it purportedly refers, such as numbers, functions, sets.

Conclusion Abstract entities, such as numbers, functions and sets that are appealed to in mathematical theories which are involved in scientific theorising, exist.

The *reconstructive* nominalist challenges premise four that the truth of mathematics commits one to the objects purported to be referred to by the mathematical terms. And to underwrite the motivation for rejecting this conditional, the *reconstructive* nominalist discards the second requirement of the semantic constraint, which involves the demand to take the surface-grammar of the mathematical discourse at face-value.

Rejecting this assumption opens up the possibility of *reconstructing* mathematical discourse in various ways. One proponent of this strategy is Hellmann who adopts a version of *Modal Structuralism*[30]. The idea is to trade mathematical ontology – reference to numbers, sets, etc – for added ideology, namely the use of modal discourse. Mathematics is now conceived as concering *possible* structures (and objects).

This conception is indeed a type of nominalism, as it refrains from reference to and quantification over existing abstract objects and instead merely commits one to possible entities. In the case of arithmetic, the commitment is to a possible ω-structure that exhibits the properties normally assigned to numbers and so makes the Peano axioms true.[31] Various formal details need to be attended to, to make this reconstructive strategy work.[32] Here, however, I will leave these formal issues aside and, just as above, note various difficulties with this approach and assess how it fares with regard to overcoming Benacerraf's dilemma.

Reconstructive nominalism typically adopts Tarski's theory of truth but denies the need to respect the surface grammar. Hence, such a conception denies the basic presumption that what seems to be referred to, or what seems to be thought about when doing mathematics, *is* what is referred to or thought about. According to reconstructive nominalism, there are no such objects as numbers underlying our mathematical thought and talk. We are systematically misguided in thinking that the surface grammar represents reality. To some, this might seem like a hard bullet to bite and, in general, I believe that an account which

[30](Hellman, 1989)

[31]On one reading of nominalism, namely Goodman's and the early Quine's, this approach would not be considered nominalistically acceptable. Compare: "Goodman and I got what we could get in the way of mathematics on the basis of a nominalist ontology and without assuming an infinite universe. We could not get enough to satisfy us. But we would not for a moment have considered enlisting the aid of modalities. The cure would have been far worse than the disease." (reply to Charles Parson in (Hahn, 1986) as quoted from (Burgess and Rosen, 1997), p. 248. I won't here enter the dispute what is distinctive of nominalism, and continue to regard Hellman as a nominalist.

[32]For example providing the right type of translation from standard mathematics to modal statements, the problem of trivial conditionals, etc. For details see (Hellman, 1989) and (Burgess and Rosen, 1997).

does respect the surface-grammar has advantages over a reconstructive account.

The important issue, however, is the epistemic constraint. The basic idea is that by avoiding the critical ontology (abstract objects) by appeal to possible structures, this type of *reconstructive* nominalism can overcome Benacerraf's dilemma. So, the thought had better be that knowledge of merely *possible* structures is easier explained and justified than knowledge of actual abstract objects. I think, however, that exactly this claim can be challenged.[33] First, to clarify the modal structuralist position, it should be noted that the possible structures are structures of objects. For there to possibly be an ω-structure to make the Peano axioms true, there have to be possibly infinitely many objects. Otherwise structures *per se* seem every bit as abstract as numbers, sets, etc. As a nominalist, Hellman should not be committed to the possible existence of infinitely many *abstract* objects because it would follow that it is contingent whether there are abstract objects – a curious contingency that needs to be explained.[34] However, Hellman is aware of these difficulties and commits himself to the possibility of a *concrete* ω-structure, i.e. that there could be infinitely many concrete objects (making up an ω-structure).

But, then how can one know this modal claim? What explains our knowledge, if we have it, that there could be infinitely many concrete objects? One option is to argue from the mere conceivability of there being infinitely many concrete objects that it is possible that there are – but the extent to which conceivability is a good guide to possibility is a further issue. In general we can say that, just as in the case of abstract objects, a variant of Benacerraf's dilemma can be raised for the modal realm and the possible existence of certain objects.[35] Unless it is clear that the latter type of knowledge – modal knowledge – is easier to explain and that the general scruples about abstract objects should be upheld, there are

[33]See especially (Hale, 1996b).

[34]This is a point made in Hale (Hale, 1996b).

[35]See for an illuminating discussion (Stalnaker, 1996).

no advantages (but merely disadvantages based on added complexity and the rejection of surface-grammar) to such a reconstructive approach.

I will here have to leave it to the reader to decide whether the view I put forward in this thesis is, in the end, less tractable than their favourite epistemology of modality and therefore will leave this option open for further discussion.[36] In addition, I believe this brief overview of the four conceptions should help to locate a *fundamental assumption* that those strategies share. In the following section I outline this assumption and by challenging it I outline an alternative strategy: *Fregean* platonism.[37]

1.2.5 The fundamental assumption and Frege's claim

We have seen that apart from intuitive platonism, the other three strategies either truncate the knowledgeable part of mathematics, deny any mathematical knowledge, or turn mathematical knowledge into modal knowledge. Hence, these three strategies have not taken the dilemma head on; rather, they give up on basic components that gives rise to the dilemma.

[36]It is noteworthy that there are various other types of reconstructive strategies, such as Hodes' conception which could be regarded as a *reconstructive platonist* conception. The main paper outlining his position is (Hodes, 1984) and further technical work is to be found in (Hodes, 1990a) and (Hodes, 1990b). Hodes adopts a conception in which number-terms do not function as genuine singular terms referring to objects, but rather to so-called Fregean concepts (or just properties for that matter). He is able to provide a formal framework (postulating an explicit axiom of infinity) that allows a reconstruction of ordinary mathematics within his framework. As with Hellman's approach, Hodes view seems – in the context of Benacerraf's dilemma – of no great advantage, since surely Fregean concepts are abstract entities and thus inherit the standard problems facing any type of platonism.
Interestingly, note that (the later) Frege suggests this view, writing: "Since a statement of number based on counting contains an assertion about a concept, in a logically perfect language a sentence used to make such a statement must contain two parts, first a sign for the concept about which the statement is made, and secondly a sign for a second-level concept. These second-level concepts form a series and there is a rule in accordance with which, if one of these concepts is given, we can specify the next. But still we do not have in them the numbers of arithmetic; we do not have objects, but concepts. How can we get from these concepts to the numbers of arithmetic in a way that cannot be faulted? Or are there simply no numbers in arithmetic? Could the numbers help to form signs for these second-level concepts, and yet not be signs in their own right?" (Notes for Ludwig Darstädter, p.366-7)

[37]I have have to forgo discussion of other structuralist views, such as those of (Shapiro, 1997) and (Resnik, 1997), and the more pessimistic position which claims that there is no, or not just one solution, to the Benacerraf challenge, defended in (Azzouni, 1994) and (Balaguer, 1998) respectively.

Considering that neither these indirect responses nor the one direct response (intuitive platonism) offer a satisfying account to resolve Benacerraf epistemic constraints, I here want to explore whether there is an assumption that is shared by these four strategies. By challenging it, we might be able to arrive at an alternative conception that holds the key to resolving Benacerraf's challenge. I think, that each strategy is committed to the following conditional, which I will label the *fundamental assumption*:

> If there is a priori mathematical knowledge and the mathematical discourse is construed at face-value, then there has to be some form of acquaintance with the objects involved that underwrites this knowledge.

This assumption can reasonably explain why the intuitive platonist postulates a "perception-like" faculty – intuition – which should be able to underwrite a priori knowledge of abstract objects. Naturalised platonism also accepts this conditional but, in contrast to intuitive platonism, denies that there can be a perception-like mathematical faculty or any other form of acquaintance with abstract objects, and accordingly denies that there can be any type of a priori knowledge of mathematics. As a result, there is a need to resort to broadly empirical and scientific considerations to explain mathematical knowledge. Both error-theoretic and reconstructive nominalism challenge the sufficiency of a broadly empirical epistemology of mathematics and both nominalist positions can also be interpreted as adopting the fundamental assumption. Error-theoretic nominalism adopts an even stronger assumption, namely that any type of knowledge has to involve some form of interaction or acquaintance, however indirect. Since there can be no such interaction with mathematical entities there is no mathematical knowledge in general. Reconstructive nominalism in contrast also accepts the fundamental assumption but then denies the component of the antecedent which claims that mathematics is to be construed at face-value.

Reflecting on this assumption offers the opportunity to rethink the platonist strategy. What exactly would be involved in rejecting the fundamental assumption? If it were rejected, it should be possible to have a priori knowledge of mathematics without having an acquaintance with the objects that mathematics is about. So, there is no need to account for an initial interaction with the subject-matter in order to found knowledge of that subject-matter. It is this *subject-matter first* idea that I will challenge, such that a subject can justifiably claim knowledge of numbers without direct acquaintance as *intuitive* platonism demands and without resorting to purely pragmatic considerations which threaten a genuine explanation of mathematical knowledge.

Instead, so the suggestion would go, we can have knowledge of numbers by reflecting upon statements about the objects in question – that is we can, in some sense, gain knowledge by linguistic competence and by mastering of the mathematical discourse. I believe that this basic and, at this stage, crude idea, which involves a *linguistic turn* for our knowledge of mathematics, was first expounded by Frege in his *Grundlagen.* Here, he aimed to pursue this strategy against physicalism (a forerunner of contemporary nominalism) and psychologism by embracing the *context principle.* By adopting this principle, which is first and foremost a *linguistic principle* about the meaning of word, the fundamental assumption can be avoided.

The following crucial passage, in the later sections of the *Grundlagen der Arithmetik* where Frege summarises what he takes to be his main achievements, supports the idea of such a pivotal role for the context principle:

> "We then set up the principle, that the meaning of a word is not to be explained in isolation, rather in the context of a sentence; we can, *solely* by our compliance to it, so I think, avoid a physicalist conception of number without slipping into a psychological conception." (*Grundlagen*, p.116, my translation and italics)

I interpret Frege here as saying that the context principle is all that is needed to overcome physicalist or psychologistic conceptions and it is all that is needed to establish his own platonist views which he has outlined previously in the *Grundlagen.* As we shall see later, Frege even assigned an epistemic dimension to the context principle. So, taken together, we can interpret Frege's position as one in which the context principle underwrites a form of platonism which does not appeal to any form of intuition (a claim Frege repeatedly makes) but nevertheless aims to explain a priori knowledge of mathematics (at least arithmetic and analysis) on the basis of this principle and so on the basis of linguistic competence. This interpretation of Frege's context principle underwrites the idea that his principle involves a rethinking and rejection of the fundamental assumption.

I will take Frege's intended use of the context principle as my starting point and briefly investigate various interpretations of that principle. This pivotal role of the context principle has generally been overlooked in the literature, and it is only the so-called *Neo-Fregean* conception defended by Bob Hale and Crispin Wright which has assigned the context principle such a role.

A possible reason that interpreters have neglected this interpretation is that Austin's official translation differs crucially from the passage as I have rendered it. Austin's translation reads as follows:

> "We next laid down the fundamental principle that we must never try to define the meaning of a word in isolation, but only as it is used in the context of a proposition: *only by adhering to this* can we, as I believe, avoid a physical view of number without slipping into a psychological view of it." (my italics)

It seems to me that instead of translating Frege as saying that it is *solely* or *merely* or *just* by adopting the context principle we can avoid these other views, Austin's translation makes it seem as if Frege uses the context principle merely as a necessary condition for establishing the Fregean position. As noted above, I believe that Frege took the context principle as the *one* principle that is needed to avoid these other positions which all subscribe to the fundamental assumption; thus it should be considered as a sufficient condition.[38]

The plan for the following section will therefore be to first trace back the idea that Frege's context principle can help to explain our knowledge of mathematical objects. To this end, I will offer a brief *historical interlude*, where I discuss Frege's appeal to the context principle, followed by different interpretations of it. In section 1.4, I will turn more specifically to the Neo-Fregean use of the context principle showing how invoking the principle involves a rejection of the fundamental assumption and how exactly it promises to underwrite a resolution to Benacerraf's dilemma. The Neo-Fregean claim is that the context principle will in effect help to overcome the dilemma while at the same time respecting all four requirements as put forth by Benacerraf – a challenging undertaking!

1.3 Historical Interlude: Frege and the context principle

Frege's context principle in its crudest form – i.e. as presented by Frege in the opening sections of the *Grundlagen* – is the instruction "never to ask for the meaning of a word in isolation, but only in the context of a proposition"(*Grundlagen*, p.x). Frege refers to the context principle four times: Once as quoted in the Introduction (*Grundlagen*, p.x), on two occasions in section IV – at the end of the subsection entitled "Each individual number is a self-subsistent object" (*Grundlagen*, §60), as well as at the beginning of the next

[38]Defending this reading would involve further Frege exegesis which I won't be able to pursue here. The next section, however, will provide some more evidence of this understanding of the context principle.

subsection entitled "To gain the concept number, one has to fix the sense of a numerical identity" (*Grundlagen*, §62) – and lastly in his concluding remarks near the end of his book in (*Grundlagen*, §106).

To show how the context principle might contribute to undermining the fundamental assumption and answer Benacerraf's dilemma, I will discuss these passages. The aim is not to directly outline how the principle is to be understood – on *my* interpretation – but rather to comprehend what role Frege assigned to it and how he intended to use it. On this basis I will introduce three adequacy-constraints on how the context principle should be understood, which will help us to evaluate the impact of the principle and the contribution it makes to *Fregean* platonism.

In the Introduction to the *Grundlagen*, the principle is used by Frege to caution us against falling prey to psychologism. He claims that if one doesn't respect the principle, then "one is *almost* forced to take as meanings of words mental pictures" (*Grundlagen*, p.x my italics) which would thereby violate another guiding principle Frege puts forth in the Introduction, namely "to separate sharply the psychological from the logical, the subjective from the objective." (*Grundlagen*, p.x) Thus, it seems that the principle is here utilised by Frege as a *heuristic device.* We can summarise this in the following first adequacy constraint for an interpretation of the context principle (CP).

> *(i) CP as a heuristic device*
>
> An interpretation of the context principle must explain how – as a methodological directive – it is supposed to prevent us from adopting a psychologicistic or physicalist view of numbers.

Note that this use of the context principle is negative in that applying the principle should prevent us from taking the "wrong" view. Its first occurrence is therefore not very substantive and does not yet involve a more positive directive. Nevertheless, the final reference to the context principle in the concluding section of the *Grundlagen* points to a more significant role. Here, the context principle is assigned an important and more fundamental function in establishing Frege's view on numbers. It is in some sense constitutive of Fregean platonism. To quote this passage again:

> "We then set up the principle, that the meaning of a word is not to be explained in isolation, rather in the context of a sentence; we can, *solely* by our compliance to it, so I think, avoid a physicalist conception of number without slipping into a psychological conception." (*Grundlagen*, p.116, my translation and italics)

This quotation points to a stronger role for the context principle as envisaged by Frege and so gives rise to the following adequacy-constraint:

> *(ii) CP as a constitutive principle*
>
> An interpretation of the context principle must explain a positive and unique role this principle plays, not only in avoiding other positions, but in establishing Frege's own view.

These first two adequacy-constraints are clearly intertwined and exhibit the negative and positive aspect of the adoption of the principle. While, on the one hand, it is to foreclose the adoption of alternative views, it is, on the other hand, constitutive of Frege's platonism. Nevertheless, I want to clearly separate these two roles, as most interpreters accept the first heuristic principle in outlining the content of the context principle, while overlooking the constitutive role which Frege assigns to it.

In addition to these two constraints, I think we can propose a third – crucial – interpretative condition, if we look at Frege's usage of the principle in §62. In this important passage, Frege refers to the context principle immediately after he poses – in a Kantian fashion – the question: "How, then, could a number be given to us, if we can't have an idea (Vorstellung) or intuition of it?" Frege's answer invokes the context principle. He writes: "It is only in the context of a sentence that words mean something. Therefore, it will depend upon the *explanation of the sense of a sentence* in which the number word occurs." (*Grundlagen*, p.73, my translation and italics)[39]

Thus, next to the first two adequacy constraints, the context principle's function is also to resolve an epistemological challenge. It is to play a role in explaining *how we can have knowledge* of the numbers that, at this stage of the *Grundlagen*, Frege already considers to be self-subsistent[40] and non-actual objects.[41] Thus the following third and last condition:

> *(iii) CP as an epistemological principle*
>
> An interpretation of the context principle should explain how it contributes to the epistemology of abstract objects, such as numbers.

These three constraints flow from a face-value reading of the *Grundlagen* and reflect three key roles that Frege explicitly assigns to his principle. In the following, I will adopt them as a framework for the assessment of various interpretations of the context principle as they occur in the literature. This ensuing discussion is not intended to provide a comprehensive overview of different interpretations of Frege's philosophy, but merely to illustrate how

[39] Austin's translation is: "How, then, are numbers given to us, if we cannot have any ideas of intuitions of them? Since it is only in the context of a proposition that words have any meaning, our problem becomes this: To define the sense of a proposition in which a number word occurs." I will leave further comments about his translation aside.

[40] There is an issue here of how to understand the notion of self-subsistence. The original uses the notion "selbstständig" which can also be understood as corresponding to the English term "self-standing". As Frege uses the expression in this context, it could be interpreted as anticipating his later distinction between predicates and proper names – the former being unsaturated and incomplete and the latter saturated and complete or 'self-standing'. Here, I don't have space to further discuss these conjectures.

[41] Frege reminds us of this claim immediately after he refers to the context principle, as playing an integral part in his answer to the Kantian question.

leading interpreters have either underestimated the role of the context principle or failed to provide a satisfying account of its content in the light of the three adequacy constraints.

1.3.1 Interpreting Frege's context principle

Before discussing various interpretations of the context principle in detail, we will first have to clarify an important terminological point about Frege's use of the term *Bedeutung*. Frege's initial formulation of the context principle is: "never to ask for the meaning [*Bedeutung*] of a word in isolation, but only in the context of a proposition" (*Grundlagen*, p.x). Crucial here is how Frege's use of the term *Bedeutung* should be understood at the time of the *Grundlagen*. As is well-known from his paper "On sense and meaning" [*Ueber Sinn und Bedeutung*], Frege identified the *Bedeutung* of a term with its denotation or referent (in the case of singular terms) and distinguished from it the *Sinn* – sense or mode of presentation – that leads to the referent of a term. Importantly, this paper was published in 1891 and so seven years after the publication of the *Grundlagen*. Unfortunately, Frege never explicitly clarified how the context principle was to be understood in the light of his later introduced distinction.

Most interpreters have taken the context principle to be a thesis concerning Fregean sense and not as a thesis concerning reference as such. However, a few others regard the context principle as implying two claims: firstly, a not very contentious claim about the sense of an expression and secondly, the more substantial claim about the reference of an expression. Lastly, there are also some, who, based on a rather specific understanding of the notion of *Bedeutung*, regard the context principle as not being a semantic principle of any kind.

In addition to these three options, there has emerged a further epistemic dimension of interpretation of the content of the principle. Remember that the third adequacy constraint hinted towards an epistemic component to the context principle. As we will see some – but only a few – interpreters regard it as a significant epistemic principle.

In the following sections, I will examine various examples of these readings of the context principle. The aim is to evaluate how well these interpretations fare in regard to the three adequacy constraints outlined before. Therefore, the discussion is less concerned with the coherence of the various interpretations or whether they correctly reflect other parts of Frege's writing. Rather, assuming that there is a principle fulfilling these three constraints, the issue is what the content of such principle has to be.[42]

Non-semantic interpretations of the context principle

Gregory Currie is one proponent of the non-semantic interpretation of the context principle. He does not regard the principle as a thesis concerning the sense or the reference of an expression. Instead he claims that Frege's *Grundlagen* usage of "Bedeutung" is to be understood as akin to the notion of "significance"[43], but – as he explicitly claims – in a non-semantic sense of significance. Thus what Frege actually means when he asks us to consider "'the *significance* of a word only in a context of a sentence', he is urging us to give an explanation of a word (to analyse the concept which it expresses) which will be faithful to the *intuitively correct* judgements in which that word occurs." ((Currie, 1982) p.151, my italics).

As a result of this non-semantic reading, Currie takes the context principle to be a constraint on how formal notions should be explicated: Their explication has to be in accordance

[42]Hence, I aim to avoid questions as to what Frege really meant considering various other aspects of his philosophy. Also, and although this is a genuine issue, I will avoid discussions of the adequacy constraints themselves. For example, Ricketts (in conversation) rejects the epistemic constraint on the basis of a re-interpretation of the dialectical structure of the *Grundlagen*. These are interesting matters, but discussing them in detail would lead us deeper into Frege exegesis, instead of using Frege's idea to overcome Benacerraf's problem.

[43](Tugendhat, 1970) is the first – to my knowledge – to suggest such a translation of the German notion "Bedeutung". "Significance" here may be understood as synonymous to "importance".

with our intuitive (pre-)conceptions of them.[44] Having this constraint in place, Currie then shows how psychological and empirical accounts have consequences that do not coincide with our intuitive understanding, and even have consequences that are inconsistent with our ordinary view of numbers.

Such an interpretation of the context principle, it seems to me, has at least the prospects of complying with the first adequacy constraint, i.e. to provide a methodological device that helps to prevent the adoption of psychological or physical views. Thus, no matter whether the arguments Currie suggests against the opposing (physicalist or psychologistic) views are any good – and I am not particularly impressed by them – the interpretation fulfils the first purpose, by assigning the context principle a role that steers a path between physicalism and psychologism.

However, if the gist of the context principle is a constraint on how to explain formal notions, then it is hard to see how it could play a positive role in establishing Fregean platonism. Currie does not invoke the principle in his explanation of why we should conceive of numbers as objects and not as certain kind of higher-order concept, for example. As a result, he falls short at least in respect to the second constraint – the need to explain the context principle as a *constitutive* principle as claimed by Frege in §106 of the *Grundlagen*.

Interestingly, Currie does take his interpretation of the context principle to have an epistemological component, which might help to fit his interpretation with the third adequacy constraint, writing that the context principle aims to "ease his [Frege's] epistemological problems" ((Currie, 1982), p.155). This epistemic gain is achieved by interpreting the context principle as making redundant the postulation of a special faculty for perceiving non-spatial objects such as numbers, because "if we can give an analysis of the concept number which

[44] Note that such a restriction to formal notions is very substantial. Frege never held the context principle to only be applicable to formal notions.

adequately explains the sense of sentences which express judgements about number, then we have nothing more to show concerning the adequacy of our grasp of numbers." (op.cit.) Thus, Currie thinks that providing such analyses – in compliance with the constraint imposed by the context principle – suffices for grasp (and possibly knowledge of) of the object falling under the concept concerned.

Unfortunately, Currie does not clarify further the conditions under which grasping the sense of an expression suffices for grasping the object falling under it. Prima facie, it looks as if this transition can't in general be correct, unless further constraints on the sentences in which the target concept occurs are imposed. For example, one can easily explain the function (significance) of the concept "golden mountain" in certain sentences, but this should not yet compel one to the claim that grasping the significance is enough for grasping of golden mountains since there are no such objects and so it is not possible to have any grasp of them.

Furthermore, there seems to be an (at least prima facie) incompatibility within Currie's view between regarding the principle as non-semantic on the one hand and assigning it the described epistemic role on the other. The worry is that, if the context principle is to pave the way to knowledge of objects from the grasp of the senses of the expressions, then there has to be an explanation of how the sense and crucially the reference of an expression is fixed as well as under what conditions a grasp of the sense suffices for grasp of the referent. This, as we will see in 1.4.1, can indeed be explained by appeal to the context principle. However, since Currie's interpretation is non-semantic and so not intended to be about the sense or reference of a term, he is unable to adequately explain the epistemic dimension – at least by appealing to the context principle – since such an explanation will have to appeal to these semantic notions.

Putting to one side this difficulty in Currie's reading, it remains that although his interpretation complies with the first adequacy-constraint and provides at least a starting point for the third adequacy constraint, a more sophisticated and persuasive interpretation of the context principle is desirable. Before moving on to semantic interpretations of the principle, let me briefly note another interpretation that could be regarded as non-semantic.

In recent years Ricketts[45] has put forth an interpretation of Frege's philosophy that can be referred to as the "no-metaperspective" conception. Very briefly the idea is that Frege – according to Ricketts – takes as the "given", i.e. as something which "needs no securing and admits of no deeper explanation"[46], the objectivity of judgements. From this he infers the no-metaperspective view, according to which

> "the conception of judgement commits Frege to taking the statements of language more or less at face value. There is no standpoint from which to ask whether the thoughts expressed by a statement of language really represent reality, whether they are really true or false. Similarly there is no standpoint from which to ask whether the statements of language really do express thoughts." ((Ricketts, 1985), p. 8)

Consequently, the semantic notions of "reference", "object" and "truth" neither are nor can be explained from this perspective, and, according to Ricketts, are not Frege's real concern. From this background, it seems that a non-semantic interpretation of the context principle is the only reading available. The context principle, therefore, is not concerned with and explicatory of the sense or reference of expressions since these are semantic notions about which nothing can be said. Consequently, it is not surprising that Ricketts plays down the role of the context principle.[47] He rejects what I presented as a face-value reading of the

[45](Ricketts, 1985), (Ricketts, 1986b), (Ricketts, 1986a) (Ricketts, 1996). Rather similar to Ricketts interpretation is (Weiner, 1990). Similar worries will apply to her interpretation. I won't have space to discuss her here.

[46](Ricketts, 1986b), p. 72

[47]Ricketts provides only a short reference to the context principle in (Ricketts, 1986b) on page 86 and on

passages in the *Grundlagen* that give rise to the three adequacy constraints and denies any epistemic or metaphysical role for the context principle.[48]

Ricketts provides a very challenging interpretation of Frege's ideas.[49] Here, however, my focus is on a possible understanding of the context principle that aims to provide for a resolution of Benacerraf's dilemma. This discussion crucially requires considerations of notions such as "sense", "reference", "object" etc. However, a discussion of this type is foreclosed from the "no-metaperspective"- perspective and although I don't propose to refute this interpretation of what Frege "really" meant – I, from hereon, leave aside this reading since it cannot help with the main issues raised at the beginning of this chapter.

Sense-only interpretations of the context principle

In this section, I will examine semantic interpretations that regard the context principle as a thesis about sense and only about sense. A main proponent of such an interpretation is Hans Sluga.[50] His reading is semantic in so far as what the context principle involves is that the sense of an expression is fixed by its occurrences in the context of sentences. Thus the phrase "never to ask for the meaning of a word in isolation, but only in the context of a proposition", should here be understood as saying that sentences are the primary semantic unit in fixing the sense of an expression. Sluga also argues that, so understood, the principle is akin to (a derivative of) the earlier Kantian maxim concerning a priority of judgements over concepts.[51]

Broadly speaking, an interpretation of the context principle as a thesis concerning sense

page 180 in (Ricketts, 1986a). Weiner only provides one reference to the context principle in her influential book – see (Weiner, 1990).

[48]Ricketts in conversation.

[49]For an excellent and much deeper discussion and rejection of Rickett's idea consult (Sullivan, 2004), especially, p. 726ff

[50]See especially (Sluga, 1980), as well as (Sluga, 1975), (Sluga, 1976) and (Sluga, 1977)

[51]See (Sluga, 1980), p.94-95

can be understood as setting up a restriction, viz. that the meaning should be fixed within the context of a sentence, thus forestalling taking mental images or "Vorstellungen" into our account of the meaning of expressions. Thus, on Sluga's interpretation of the context principle, it is superfluous to consider any mental images or ideas when considering the sense of the number-term. The term will be adequately explained by its use in appropriate statements. Hence, the principle so understood, provides a *heuristic* not to appeal to mental images and the like, and thus fulfils the first adequacy constraint as previously laid down.

But apart from this point, it is hard to discern in what way Sluga assigns any clear role to the context principle which might fit Frege's other ideas. In the following paragraph Sluga acknowledges it as an important principle in the derivation of Frege's account of number. He writes:

> "Frege agreed with Kant's claim that judgements possess an original unity and he reasoned, with Kant, that if there were logical objects they would have to be defined in terms of that pure formal unity. Empirical, psychological, and formalist conceptions of numbers have their origin in the separation of the number term from the propositional context. The characterization of numbers as logical objects had to proceed in another way. *Since logical objects cannot be given in intuition, the meaning of the term referring to them must be completely determined by the sentential context in which they occur.* These considerations led Frege to emphasize again and again in the *Foundations* [Grundlagen] the importance of the context principle of meaning." ((Sluga, 1980) p.123ff, my italics)

Whether or not Frege is in agreement with Kant about the role of the "pure formal unity" (whatever that is), Sluga seems to constrain the context principle in a very important way. First note, however, that Sluga is correct to emphasise that psychological and other accounts do not obey the context principle and thus he gives it some weight in Frege's rebuttal of the other accounts. Curiously though, as highlighted in the italicized sentence, Sluga seems to assume that numbers as logical objects exist, and that *because* they are *not* given in intuition, the expression that denotes these objects must be *completely* defined through its occurrence in whole sentences. So, the context principle is restricted and only applies to expressions whose objects are not given in intuition.[52] Yet, Frege never took the context principle to be restricted to non-intuitable objects. In its first occurrence in the introduction to the *Grundlagen*, he clearly takes it to apply to expressions of any sort.

Sluga's account of the context principle is unfortunately silent as to whether it contributes to the metaphysical or epistemological demand as outlined in our second and third adequacy constraints. These questions remain open within the context of Sluga's interpretation of the principle, because he believes that nothing substantial hinges on this principle apart from the above mentioned priority of judgement over concept. Moreover, these questions have to remain open because the context principle, according to Sluga, is merely a thesis concerning sense – the *Sinn* – of an expression. He does not make any connection between proper names and reference or under what circumstances we can assume that terms, such as numerical terms, do refer to an object.[53] Thus, I believe that Sluga's austere reading does not account for the role that Frege ascribes to this principle in the *Grundlagen*.[54] So I will move

[52]This is in some respect similar to Currie's reading.

[53]There is a hint of this idea in the following paragraph: "Frege characterizes objects by the properties of the expression that stand for them. In other words, he transforms what looks like a material and ontological problem into a formal semantic one. An expression standing for an object is called a proper name by Frege. That terminology is perhaps not altogether happy, since it turns out that the term has an essentially wider use for Frege than in everyday life. But nothing hangs on this usage. Instead of using the two terms 'object' and 'proper names' he could have used two newly coined terms A and B and expressed his views completely in terms of them." ((Sluga, 1980), p.122) This latter statement is very confusing. How can 'A' and 'B' express substantial metaphysical claims, if these terms are not properly explicated?

[54]This does not imply that the whole methodology of Sluga, who interprets Frege in a historical perspective – as a Kantian – is wrong. My point is merely that his discussion of the context principle is insufficient, regardless of whether Frege is or is not a transcendental Idealist.

on to another interpretation that understands the principle as entailing more substantial claims. One reading suggested by Resnik[55] will be briefly discussed before moving on to the Neo-Fregean interpretation.

Referential interpretations of the context principle

There are various accounts that interpret the context principle as a thesis concerning the reference of an expression – most notably the Neo-Fregeans whose interpretation will be the main focus of the next section. Here, I want to briefly review an interesting reading of the context principle offered by Michael Resnik. His account takes the context principle as a metaphysical doctrine and ties Frege's view of numbers *as objects* directly to this doctrine.

Resnik claims that the context principle explains how Frege could "retain his ontological platonism"[56] and regards the use of the principle as motivated through the rationalist tradition. Although his interpretation is not backed by much detail – only a few pages in his book are concerned with the context principle – some support for his reading stems from the explication of the following quotation of Frege:

> "In arithmetic we are concerned with objects that we come to know not as something alien, from without through the medium of the sense, rather they are directly posited to reason, which, as its nearest kin, it can completely grasp." ((Frege, 1884), §105 my translation)

To say that numbers are given to reason – Resnik's thought goes – is to say that because judgements are the main function of reason, it is through articulated judgements, and so through sentences expressing such judgements, that the *nature* of numbers is recognised.

[55]See especially (Resnik, 1980)

[56]To be an ontological platonist in this context means to believe in the existence of numbers as being on a par, albeit abstract, with other "ordinary" objects; this coincides with what I previously called "platonism".

Subsequently, to ask what numbers are outside their mention in such sentences is, according to Resnik's interpretation, "to undertake to say what things are like independent of reason, [which] would be as much to judge without judging, or to wash the fur without wetting it." (*Grundlagen* §26)

On this reading, the context principle plays a role in *exhibiting* what the nature of numbers is and it is used to explain how we can come to know of them a priori. As Resnik points out, the existence of numbers as objects can be known – according to Frege – by appeal to a faculty of reason that illustrates the remnants of the rationalist tradition within Frege, but also allows us to interpret Frege as an ontological platonist.[57]

The idea that it is through reason and through judgement that a subject can *know a priori* about the nature of abstract objects such as numbers – without some kind of acquaintance or a perception-like faculty – underlies a position that rejects the fundamental assumption. Unfortunately, however, it is hard to discern a clear and transparent understanding of what the context principle, according to Resnik, contributes to a Fregean platonism that rejects this assumption. His discussion, although intriguing, does not provide sufficient material and content to regard it as fulfilling the three adequacy constraints and so Resnik's interpretation provides merely the first steps towards a genuine *Fregean* platonist position.[58]

In contrast the Neo-Fregean conception, as defended by Hale & Wright, provides far more detail and content to the role that the context principle plays in founding Fregean platonism, and it is this view I will discuss next.

[57]See ((Resnik, 1980), p. 166). The section also describes Frege's position not only as an ontological platonist but also of an objective idealist, which is confusing!

[58]So, for example he also claims that all the arguments Frege provides against the psychologistic and physicalist conceptions can be interpreted without appealing to the context principle. This claim is outlined in (Resnik, 1976).

1.4 The Neo-Fregean interpretation of the context principle

In the following, I will outline the interpretation and use of the context principle adopted in the Neo-Fregean programme as defended by Bob Hale and Crispin Wright.[59] Importantly, and in contrast to the other views previously outlined, the context principle now plays a significant part in overcoming Benacerraf's challenge. Here, it is primarily my aim to show how the context principle is understood within the Neo-Fregean framework and not to defend it. Unfortunately, however, it is hard to find one clear description of the content and the role of the context principle within the Neo-Fregean project. As a result, the following investigation will adopt a piecemeal approach and in order to guide this presentation, I will separate the following three questions and discuss them in turn.

1. What is the content of the context principle according to the Neo-Fregean?
2. What role can it play according to the Neo-Fregean in answering Benacerraf's challenge?
3. What is the relation between Frege and the Neo-Fregean?

The first two questions will receive what I take to be answers from within the Neo-Fregean perspective. Then, I will briefly review connections and differences between the Fregean and Neo-Fregean approach. A critical discussion of different arguments against the Neo-Fregean interpretation of the context principle will be pursued in the next chapter.

[59]See (Wright, 1983), chapter 1 and section x, (Wright, 1995), (Hale, 1987), and as recent as (Hale and Wright, 2002) – early hints of this interpretation can be found in (Dummett, 1956).

1.4.1 The content of the Neo-Fregean context principle

The Neo-Fregeans interpret the context principle primarily as a thesis concerning the reference of singular terms. Recall Frege's first enunciation of the principle in the *Grundlagen*:[60] "one has to ask for the meaning of words in the context of sentences, not in their isolation." (*Grundlagen*, p. x). Crudely speaking, the Neo-Fregean version of the context principle is gained by replacing "words" with "singular terms" and crucially by identifying Frege's use of "meaning" [Bedeutung] with reference. Furthermore, what is distinctive about the occurrence of singular terms in sentences, as opposed to those same terms in isolation, is the truth value of the sentence. So, reference for a singular term is established, provided there is a true sentence – of the appropriate type, i.e. an atomic sentence – in which the singular term occurs.

This idea has various consequences as we will see in the following, but first let me make out one point of affinity between Frege's initial formulation of the context principle and the Neo-Fregean interpretation. Frege claims that it is part of accepting the context principle that one should not ask for the meaning of a term in *isolation*. On the Neo-Fregean interpretation, this claim can be understood as preventing us from asking the additional question as to whether the term has a referent, even if all linguistic criteria for singular termhood are fulfilled and the truth of the statement is established. Doing otherwise, is to *isolate* questions of existence and thus to think that further investigation would yield an answer. On this reading this additional question is superfluous.[61]

Hence, so far, the context principle – as understood by the Neo-Fregeans – entails that if there is a singular term occurring in a true sentence (of the appropriate type) then the term refers to an object. As such, this claim would not be of great interest unless the additional

[60] The following paragraph is partly based on (Cook and Ebert, 2004)

[61] This is similar to a point made by Resnik in section 1.3.1.

claim is made that the criteria of singular termhood can be established without appeal to the object presumed to be denoted by that term.

Exactly this thought is an important component in the Neo-Fregean interpretation of the context principle, which is missing from other interpretations. The general idea is that there is a priority of linguistic questions or categories over ontological questions or categories. So, the idea is that linguistic criteria should suffice for recognising the referential potential of a term (whether it does refer obviously depends on the truth of the sentence). This type of "priority" of linguistic over ontological categories is further explicated in the following passage from Wright:

> By 'take priority' I mean simply this: questions of reference are not to have the independence that would make it possible to determine that a class of expressions have no genuine reference when, by the best syntactic criteria, these expressions function as singular terms in a range of statements [...] which we have every reason to suppose to be true." ((Wright, 1983), p.25)

Hence, whether a term has referential potential (i.e. is to be regarded as a singular term which does refer if it figures in a true atomic sentence) can be established in virtue of certain syntactic criteria a term has to fulfil in order to be regarded as singular term. This results in the idea that "the notion of an object is posterior in the order of explanation to that of a singular term".[62] Crucially, this thesis which is, according to the Neo-Fregean, tied to the context principle, puts the onus on its proponent to outline specific *syntactic criteria* independent from the notions of reference, or object, which suffice to identify singular terms.

[62](Wright, 1983), p. 24. Note that just before this passage, Wright identifies this thesis explicitly with the context principle. These views are still held by Wright. So, for example he claims one should "... treat syntactic categories, *singular term* and *predicate*, as primary in the order of explanation and the ontological categories *object* and *concept/property* as derivative.[...] But as far I can see [...] the general notion of an object remains: *referent of a (possible) singular term.*" (Wright, 1998b), p. 263

In the following, however, I will forgo any discussion of the syntactic criteria of singular termhood, which have received much discussion in recent years,[63] and simply assume that these syntactic criteria can be made out and that number terms thereby feature as singular terms – just as they intuitively seem to.

Interestingly, note that the proclaimed priority of linguistic categories over ontological categories seems to oppose (or at least provides the beginning of an opposition to) the *fundamental assumption.* Remember that the fundamental assumption gives rise to a *subject-matter first* view – hence in general ontological categories such as objecthood have an epistemological priority over any other category. In contrast, the context principle, as understood by the Neo-Fregeans, breaks with this type of epistemic priority and inverts it. Here it is linguistic categories such as issues about singular termhood and issues about the truth of judgements that are prior to ontological questions about objecthood.

However, this *Copernican revolution* is sometimes thought to reflect on the status of the objects involved. The syntactic priority thesis has been taken to entail that numbers conceived in accordance with it, are in some sense mind or language-dependent. In various passages even the Neo-Fregeans hint towards such consequences about the status of the objects involved. So, for example, in the following quotations Wright can be interpreted as endorsing this mind dependence.[64]

> "If we endorse the syntactic priority thesis [the context principle], we abandon the view, characteristic of the *Tractatus* that the structure of a state of affairs is somehow determined independently of the syntactic structure of any statement

[63]See for example, as proponents (Dummett, 1981a), p.54-80, (Hale, 1979), (Hale, 1995) (Hale, 1996a) (Wright, 1983) p.10-12 and 53-64. For an overview of criticism see (MacBride, 2003) and more specific criticism (Williamson, 1988) and (Wetzel, 1990).

[64]MacBride's survey article provides this impression of the Neo-Fregean view, see his (MacBride, 2003). In contrast, the other Neo-Fregean, Bob Hale, has argued strongly against such dependency – see his (Hale, 1987), chapter 7.

of it." ((Wright, 1983), p. 129, my addition in square brackets)

> "The irresistible metaphor is that pure abstract objects [such as numbers], conceived as by Fregean platonism, and the states of affairs to which, in accordance with the Correspondence Platitude, merely minimally true sentences correspond, are no more than *shadows* cast by the syntax of our discourse. And the aptness of the metaphor is merely enhanced by the reflection that shadows are, after their own fashion, real." (Wright, 1992),p. 181-2, my addition in square brackets)

These two quotations can be interpreted as suggesting a reading of the Neo-Fregean version of the context principle in which objects are mind or language dependent. Here, I won't dwell on whether these passages *should* be understood in this way or not. However, in chapter 2 we will focus on various arguments put forth by Dummett and others according to which it is a *consequence* of adopting the context principle that the objects referred to are mind or language dependent.

In the light of the various (presumed) consequences that are directly tied to the context principle, I will distinguish very clearly two crucial components which are constitutive of the Neo-Fregean interpretation of the context principle and one third component which I don't regard as constitutive of the principle. The first component can be captured by the following conditional:

> *Minimal conception of the context principle*
>
> If there is a singular term occurring in an appropriate true sentence, then this term refers and there is an object which is the referent of this term.

Note that this claim is rather minimal; no priority of linguistic categories is mentioned and no constraint on the criteria for singular termhood is in place. As such, this *minimal* reading is a claim that seems compatible even with *error-theoretic* or even *reconstructive* nominalism.[65]

In addition to the minimal claim, we noted that for the context principle to be of interest, a further claim about singular terms has to be in place, motivated by the claimed priority of linguistic over ontological categories.

> *Syntactic Priority Thesis*
>
> Linguistic categories have priority over ontological ones and thus the criteria for singular termhood can be established without prerequisite appeal to the objects referred to by singular terms.

This, as previously noted, is a more important and bolder Neo-Fregean claim and one which underwrites part of the challenge to the *fundamental assumption.*

Lastly, there is the most contentious and for many unwelcome consequence of adopting the context principle, viz. that such priority of linguistic over ontological categories affects the status of the objects involved. Thus the context principle – contrary to what the Neo-Fregeans claim – is inconsistent with platonism. The claim is:

> *Unwelcome Consequence: Metaphysical status of objects*
>
> At least some objects – such as pure abstract objects, like numbers, sets, etc. – whose reference is established by the context principle, exist mind/language dependently.

To summarise, I believe that the content of the context principle as understood – or at least as it should be understood – by the Neo-Fregean can be characterised by the first two claims, the *minimal* component and the *syntactic priority thesis.* The third issue – the *unwelcome*

[65]The former just rejects the truth of the mathematical sentence involved and the latter rejects the claim that number-terms do function as singular terms.

consequence – is a possible threat to the Neo-Fregean version of the context principle and as such is not intended to be part of the content of this principle.[66] Having this understanding of the context principle in place, it remains to be seen how it can help to overcome Benacerraf's challenge and provide a foundation for the Neo-Fregean platonist position.

1.4.2 The Neo-Fregean context principle and Benacerraf's challenge

To provide a first indication of the intended solution, the following rather simple but elegant way to arrive at knowledge of numbers as objects has been suggested by the Neo-Fregean. It is also know as the *Fregean argument*:[67]

Premise 1 If a range of expressions function as singular terms in true statements (of the appropriate type), then there are objects denoted by the expressions belonging to that range.

Premise 2 Number terms, and many other numerical expressions, do so function in statements of mathematics.

Premise 3 The statements of mathematics are true.

Conclusion There exist objects denoted by those numerical expressions (i.e. there are numbers).

Premise one is just a formulation of the minimal version of the context principle. Premise two is the claim that number terms do fulfil the syntactic criteria of singular termhood. Premise

[66]Frege would also regard it as an unwelcome consequence to his position if the context principle would entail the rejection of platonism.

[67]See (Hale, 1987), p.11 for this argument – I've made some changes to the presentation of the argument.

three is the claim that the sentences in which the terms function are true and the resulting conclusion is that there are numbers, existing as objects – which if anything (although it does not follow from the argument) should be regarded as abstract and mind-independent.

This sketch of reasoning indicates the various problematic steps that have to be argued for in order to meet Benacerraf's challenge. Also, an important and at the moment missing component from this proposal, is the issue of how a subject can be warranted in holding true the three premises and thereby acquire a warrant for the conclusion! That is: what warrant can we have for regarding the context principle to be correct and for thinking that the number term is a singular term? Equally, further discussion is needed concerning a subject's grounds for holding true the statements of mathematics, as well as grounds for thinking that the objects involved are really mind-independent. So, what is apparent is that although the context principle does play an integral part within the Neo-Fregean solution, various other theses need to be added in order to achieve a resolution of Benacerraf's dilemma.

The aim of the rest of this section is to outline the place of the context principle within the Neo-Fregean programme. The Neo-Fregean adopts three broad claims – one of which is the context principle – that together are aimed to address both the semantic and the epistemic constraint of Benacerraf's dilemma. They are:

1. The context principle;
2. A theory of abstraction principles backed by a broader theory of implicit definitions;
3. Second-order logic

I will explain how the context principle (i.e. the minimal reading combined with the syntactic priority thesis) contributes to upholding the standard view to comply with the semantic constraint – a Tarskian theory – and how it may resolve one ingredient of the epistemic constraint.

As mentioned in section 1.1, the context principle, even in its *minimal* reading, is the driving force behind the adoption of the standard view to comply with the semantic constraint as outlined by Benacerraf. According to this minimal reading of the principle, we can take the surface grammar at face value – provided number terms fulfil intuitive constraints for singular termhood. In addition, a broadly Tarskian account for truth in terms of satisfaction and reference is built into this version of the context principle. So, a proponent of the context principle will fulfil the first two semantic requirements – *homogeneous semantic theory* and *surface grammar* – of Benacerraf's dilemma.

However, if that were all there is to the context principle, then it hardly makes any headway in providing a resolution of Benacerraf's dilemma. For one thing, it only concerns the *semantic constraint* of the dilemma, and so far does nothing to address the *epistemic constraint*. Yet, the thesis of *syntactic priority*, which the Neo-Fregean regards as the second main component of his understanding of the context principle, might address at least one ingredient of the epistemic constraint. In general, I can discern at least two important features of the syntactic priority thesis in the Neo-Fregean account. Firstly, consider again the *Fregean argument*, and in particular premise 2 which said:

Premise 2 Number terms, and many other numerical expressions, do so function in statements of mathematics.

Assume for a moment that syntactic priority is not in place and that to *know* whether a term functions as a singular term, the subject requires previous knowledge or acquaintance with the object the terms purports to refer to. If that were the case, the *Fregean argument* would suffer from a *transmission of warrant-failure*. To explain, it seems that no warrant from the premise could be transferred to the conclusion, i.e.:

Conclusion There exist objects denoted by those numerical expressions (i.e. there are numbers).

since the truth of the conclusion would have to be appreciated already in order to know the premises. So, and this is the first important feature of the *syntactic priority thesis*, only by adopting it, can the Fregean argument avoid this type of *epistemic circularity*.[68]

The second important feature the *syntactic priority thesis* involves, has to do with meeting Benacerraf's first requirement of the epistemic constraint. To explain this feature, recall the two ingredients of the epistemic constraint earlier distinguished (section 1.1):

1. *Reference and object-directed thought*

   The demand to explain how the objects posited by the semantic theory can, in principle, be in the range of directed thought and talk of the subjects.

2. *Knowledge*

   The 'integration challenge': The demand to reconcile the truths of the subject matter with what can be known by ordinary human thinkers. Crucial here is to provide an explanation of how a subject can have mathematical knowledge and on what basis the subject can claim such knowledge.

Now, note that the context principle (minimal reading with syntactic priority thesis) can't by itself address the second of these. For, in order to be able to explain how we can know mathematical truths, assurance is needed that mathematical sentences are true – assuming factivity of knowledge. Surely though, such assurance can't stem from an adoption of any plausible version of the context principle. Rather, I think, the context principle should only be understood in contributing to an answer to the first requirement of the epistemic constraint; and I think that the thesis of syntactic priority – the crucial component of the

[68]The notion of transmission of warrant-failure will feature prominently in chapter 5. A more thorough characterisation of this concept can be found in 5.2.

Neo-Fregean understanding of the context principle – is intended to resolve the first step of how a subject can – in thought and talk – have access to abstract objects.

I believe we can distinguish a stronger and weaker reading of the idea of 'object-directed thought'. The stronger reading involves having *singular identifying thought* about an object. So, to grasp a statement (of the appropriate kind) involving an object, carries with it the ability of identifying, being able to single out, the object. In contrast, on the weaker reading, object-directed thought to abstract objects need not involve any identifying thought about a specific object, but mere object-directed thought in general.

With this distinction in place, let me outline how the Neo-Fregean claims to provide an answer to the first requirement and clarify whether he aims to answer to the stronger reading or the weaker reading of object-directedness. To explain the Neo-Fregean answer, remember what the syntactic priority thesis entails, namely: "no better explanation of the notion of an object can be given than in terms of the notions of singular term and reference."[69]. Consequently, the Neo-Fregean argues that grasping the sense of a singular term and knowing its role in a range of statements suffices for *object-directed thought* concerning the objects involved. So, on this view, to have object-directed thought about an object need *not* consist in any prerequisite acquaintance with the object itself (for example via some causal interaction). Rather, the linguistic priority thesis implies that understanding the terms, which purport to refer to the object in question, by grasping the truth-conditions of suitable kinds of statements involving these terms, suffices for object-directed thought in general.

In addition, the Neo-Fregean has argued that understanding the *right type* of sentences will even guarantee a stronger reading of object-directedness. So, for example, the strong conception is explicitly endorsed in the following quotation: "Singular thought, and object-directed thought in general is, on this view, *enabled and fully realised* in an understanding

[69](Wright, 1983), p.24

of suitable kinds of statements."[70] So what is achieved by grasping "suitable kinds of statements" is singular identifying thought about the objects they concern.

Here, I won't dwell on whether this strong reading is in general necessary or whether a weaker reading of access suffices to conform to the first epistemic requirement.[71] Whatever conception is adopted however, the *syntactic priority thesis* is akin to a kind of *minimalism about object-directed thought and reference*: All there is to intelligible thought *about* objects (in either the strong or weak sense of access) is grasping singular terms referring to those objects in sentences of the appropriate kind. This conception, if correct, will enable the Neo-Fregean to comply with the first requirement of the *epistemic constraint* by appeal to the context principle, provided that it can be made clear what counts as an *appropriate kind of statement*. More specifically, the question now arises what kinds of statements are such that understanding them suffices for an understanding of number terms.[72]

It is (in part) to address this last issue that the Neo-Fregean theory of abstraction principles comes into play. But before explaining how abstraction principles contribute to a full answer to the epistemic constraint, let us summarise the role the context principle plays within the Neo-Fregean reply to Benacerraf's challenge:

1. The context principle involves the adoption of the standard view to comply with the semantic constraint.

2. The context principle (especially the syntactic priority thesis) plays an essential part

[70](Hale and Wright, 2002), p.115

[71]This also depends on what the "suitable kinds of statements" are supposed to be.

[72]Note that if the strong conception about object-directedness is adopted by the Neo-Fregean, a recent criticism by (MacFarlane, 2004) that there seems to be no reason to treat "the number of F's" (once instantiated) as a singular term, instead of regarding the resulting expression as a quantifier could be addressed: Namely, it can be argued that it is only singular terms that can underwrite singular thought, while conceiving of "the number of F's" as a quantifier can – at best – underwrite object-directed thought in general, yet not singular identifying thought. This is why a Russellian treatment of mathematical expressions is inappropriate.

in addressing the first requirement of the *epistemic constraint.*

3. The context principle is unable to assist addressing the second requirement of the epistemic constraint.

Let me now proceed to outline how the theory of abstraction principles ties in with the context principle and how this theory addresses the second requirement of the *epistemic constraint.*

The Neo-Fregean answer to the question of what the appropriate sentences are to have a sufficient understanding of number terms and thereby object-directed thought about numbers, goes as follows: It suffices for an understanding of a number-term to have a grasp of the sense of some *identity statements* involving it – that is, by grasping the truth-conditions of such identity statements. The Neo-Fregean offers a characterisation of the truth-conditions of identity statement by appealing to the theory of *abstraction principles.* Abstraction principles are statements that introduce a term-forming operator, such as "the number of" by connecting the truth-conditions for identity statements involving this operator, with the truth-conditions of sentences involving previously understood terms and relations.

More specifically and to illustrate the general pattern, consider the following abstraction principle:[73]

Direction principle

(DP) The direction of line a is equal to the direction of line b if and only if line a is parallel to line b

[73]For this expository purpose it won't matter that the following is an objectual abstraction principle, while abstraction principles used for arithmetic and more generally mathematics are conceptual abstraction principles, abstracting upon concepts and not objects.

Here, the newly introduced term-forming operator is "the direction of"; the truth-conditions of statements involving it are fixed by the obtaining of parallelism of the relevant lines.[74]

Importantly, the Neo-Fregean claims that such biconditionals can be *stipulated* to be true and thereby provide a first grasp of the operator "the direction of". A term formed by means of the operator, provided it fulfils the syntactic criteria, is then regarded as a singular term. Assuming that the biconditional holds (that is, the stipulation succeeds) and there is a reliable method to grasp the truth of instances on the right hand side – i.e. parallelism of specific lines – we can then arrive at knowledge of the identity of specific directions.

Interestingly, what (DP) provides is knowledge of abstract objects, i.e. directions, in virtue of knowledge of certain statements involving the parallelism of lines which, arguably, can be regarded as concrete objects. Thus, abstraction principles are sometimes considered as bridge principles between the concrete and abstract world.[75] So, the theory of abstraction principles has at least two components: First, it is meant to provide a grasp of a new expression, such as "the direction of" in the above type of case. This achievement, in conjunction with the context principle, provides access to abstract objects (directions). Second, provided abstraction principles can be stipulated to be true – an issue that will inform much of the later parts of this thesis – they can underwrite knowledge of the abstract objects they are purportedly about.

In the following, I will offer some details concerning how this general idea can be applied in the case of knowledge of arithmetic. Here, the Neo-Fregean defends an abstraction principle for number-terms, which – by providing a similar story as above – would then explain

[74]This example exhibits another important feature of abstraction-principles, namely that the right hand side of the biconditional involves an *equivalence relation*, that is a reflexive, symmetric and transitive relation which divides the domain into distinct equivalence classes. So, for example, the direction of a specific line (say line z) can be arrived at or we might say, abstracted from the equivalence class of parallel lines to line z.

[75]This – one might think – only applies to objectual and not conceptual abstraction. For an elaboration of such a characterisation, see especially (Rosen, 1993).

knowledge of numbers. The abstraction principle in question is known as *Hume's Principle* and can be formulated as follows:

Hume's Principle (HP)

$$\forall F \forall G((Nx : Fx = Nx : Gx) \leftrightarrow (F \approx G))$$

where "$Nx : Fx$" stands for "the (cardinal) number of F's" and "$\approx$" expresses the equivalence relation of equinumerosity.[76] Thus, the principle claims that the cardinal number belonging to the concept F is identical to the cardinal number belonging to the concept G, if and only if there is a one to one correspondence between the objects falling under F and G respectively.

Assuming that we can put forward this abstraction principle as true – stipulatively so – the question arises how we can acquire knowledge of the right-hand side. And this is exactly where the third broad issue (mentioned on p.57) gets its grip and is why the Neo-Fregean, as did Frege, considers himself a *Logicist.* For, it is a matter of logic that there are true instances of the right-hand side.[77] To explain this, just note that it is a logical truth that the instances of the concept "being non-self-identical" can trivially be put in a one-one correlation with themselves. This true instance of the right-hand side of HP, suffices – assuming that HP is true – to yield a true identity statement about numbers on the left hand side. More formally this can be expressed as follows:

Step 1

$$(Nx : x \neq x = Nx : x \neq x) \leftrightarrow (x : x \neq x) \approx (x : x \neq x)$$

[76]The equivalence relation of equinumerosity can be formulated in purely logical vocabulary. In full detail Hume's Principle is the following statement:
$\forall F \forall G\Big((NxFx = NxGx) \leftrightarrow \exists R\big(\forall x[Fx \supset \exists y(Gy \wedge Rxy \wedge \forall z(Gz \wedge Rxz \supset z = y))] \wedge \forall y[Gy \supset \exists x(Fx \wedge Rxy \wedge \forall z(Fz \wedge Rzy \supset z = x))]\big)\Big)$

[77]In chapter 4, I will further discuss the role of these 'truths of logic'. See especially 4.5.

The right hand side of step 1 is a logical truth. Assuming the truth of the biconditional we can discharge the right hand side of HP to arrive at:

Step 2

$(Nx : x \neq x = Nx : x \neq x)$

Assuming both that number-terms are singular terms and step 2 expresses a truth, we can, by assuming the context principle, infer the claim that there is an object to which the singular term refers; and so we can quantify into this formula, reading the existential quantifier at face value:

Step 3

$\exists y(y = Nx : x \neq x)$

In addition, having the technical result in place that the Peano axioms for arithmetic can be deduced from Hume's Principle and second-order logic – a result which is called *Frege's theorem*[78] – the Neo-Fregean can justifiably claim that knowledge of logic leads, merely through the stipulation of Hume's Principle, to knowledge about numbers as objects and so to knowledge of arithmetic as such. Benacerraf's worries are met by appeal to the trinity of the context principle, the theory of abstraction and the unique status of second-order logic.

There is much to say about this "simple" reasoning and much of chapter 3 and beyond will be concerned with the theory of abstraction principles or, more broadly, the theory of implicit definitions and the role assigned to the notion of stipulation. Therefore I will postpone detailed discussions of various contentious points I assumed in this rough characterisation

[78]This theorem was first explicitly noted by Parsons, in his (Parsons, 1965) and later independently "rediscovered" in (Wright, 1983), p. 158-69. More recent presentations of the proof can be found in (Boolos, 1987a) (discursive), (Boolos, 1990) (rigourous),(Boolos, 1995) and (Boolos, 1996). Note that even a weaker version of Hume – Finite Hume – suffices for this derivation, see (Heck, 1997).

of the Neo-Fregean train of thought. By way of summarising, I will briefly put the main claims of the Neo-Fregean position in the context of the above *Fregean argument* and note how according to the Neo-Fregean we can come to know each premise:

Premise 1 If a range of expressions function as singular terms in true statements (of the appropriate type), then there are objects denoted by the expressions belonging to that range.

Premise 2 Number terms, and many other numerical expressions, do so function in statements of mathematics.

Premise 3 The statements of mathematics are true.

Conclusion There exist objects denoted by those numerical expressions (i.e. there are numbers).

As discussed above premise 1 just is the minimal reading of the context principle and, if correct, can be known by philosophical reflection. Premise 2 ties in with the linguistic priority thesis in that we can know that number terms do function as singular terms independent of a prerequisite acquaintance with the objects the term purportedly refers to. However, as already mentioned, the fact that number-terms do so function won't be discussed here any further and will be one of my basic assumptions in this thesis. Premise 3 is connected with the theory of abstraction. Here, as we saw above, the idea is that we can stipulate certain abstraction principles to be true. In virtue of these stipulations we can arrive at knowledge of the abstraction principles and combined with knowledge of logic (to instantiate Hume's Principle) this gives rise to our knowledge of the truth of the statements of mathematics, and so to knowledge of the existence of numbers. Finally, the Neo-Fregean regards knowledge of arithmetic as a priori and so he embraces the additional claim that all three premises can be known a priori and that reasoning within second-order logic (which for example is needed to establish Frege's Theorem) preserves the a priori status of the (a priori knowable) abstraction principle.

There, in rough outline, is the answer of the Neo-Fregean programme to Benacerraf's challenge. It involves various contentious claims about stipulation of abstraction principles and their ability to found knowledge. These issues and others will be the subject of the next chapters. Before this more detailed discussion of the proponent of *Fregean* platonism, or as we might also call it, *semantic* platonism, let me offer some final observation about the relationship between Frege and the Neo-Fregean.

1.4.3 Frege and the Neo-Fregeans

There are enough similarities between Frege and the Neo-Fregean to justify the attribute "Neo" within their position. In terms of the context principle, I think the role it is assigned by Frege and extracted in terms of the three adequacy constraints, fits nicely with the Neo-Fregean interpretation. It provides a clear heuristic to avoid physicalism and psychologism by pointing out that no prerequisite direct acquaintance, in virtue of intuition, causal constraints, etc, is needed, in order to know about numbers. It provides a first step in answering the epistemological challenge and so fulfils an epistemological role. Finally, it also explains why we should conceive of numbers as objects, since number terms function as singular terms.

Also, the theory of abstraction principles adopted by the Neo-Fregeans and regarding second-order logic as logic (and not set-theory) is already suggested in Frege's writing, although it is far from clear whether Frege would accept the Neo-Fregean conception of abstraction principles as foundational principles.

In contrast to Frege, who based his theory of numbers on Basic Law V, an abstraction principle which introduced the term "the extension of", the Neo-Fregean does without this principle (which, of course, Russell showed to be inconsistent). Instead the Neo-Fregean adopts the above-mentioned Hume's Principle as that on which knowledge of arithmetic is based. This is in contrast to Frege, who despite using Hume's Principle (subsequently 'justified' by deriving it from Basic Law V) in his derivations in the *Grundlagen* and the *Grundgesetze*[79] rejected Hume's Principle as a *foundation* for arithmetic since it does not provide a sufficient characterisation of the concept of number. According to Frege, Hume's Principle is subject to what is known as the *Caesar Problem*. The Caesar Problem is a problem arising from the fact that abstraction principles only provide the necessary and sufficient conditions for abstracts of the same kind to be identical. That is, Hume's Principle fixes the truth-conditions of identity statements involving number terms. But, as Frege noted, this principle fails to fix the truth-conditions knowledge-conferringof mixed identity statements:

> "we can never – to give a crude example – decide by means of our definitions, whether a concept has the number *Julius Caesar* belonging to it, whether this famous conqueror of Gaul is a number or not." (*Grundlagen*, §56)[80]"

There is an extensive literature on the Caesar Problem which, however, I won't be able to discuss in this thesis.[81]

There are many more similarities and dissimilarities between Frege's initial position and

[79]The *Grundgesetze* vol I and II (Frege, 1903) is Frege's more formal work in which he showed how arithmetic can be reduced to logic.

[80]Austin's translation: "we can never – to take a crude example – decide by means of our definitions whether any concept has the number *Julius Caesar* belonging to it, or whether that same familiar conqueror of Gaul is a number or is not.

[81]See (Wright, 1983) (Dummett, 1981a) and (Dummett, 1991a), (Potter and Sullivan, 1997) and (Hale and Wright, 2001b) and a special issue of *Dialectica* on the *Caesar Problem*, vol. 59 (2) with numerous articles. Why exactly Basic Law V evades the Caesar Problem is curiously not discussed by Frege himself.

Neo-Fregeanism which would need further careful study. But, since my main interest is whether the Neo-Fregean programme as outlined promises to be a viable position to overcome Benacerraf's dilemma, I will leave further discussion of this exegetical question to one side.

1.5 Conclusions and outlook

The main aim in this introductory chapter has been to outline Benacerraf's dilemma and to review distinct strategies to overcome his challenges, a proper response to which is necessary for any viable philosophy of mathematics. I do not claim to have fully undermined the four alternative conceptions I outlined earlier. Nonetheless, I hope my rather discursive account of the strategy of the *Fregean* or *semantic* platonist, in form of the Neo-Fregean project, has provided sufficient motivation and indicated recognisable prospects in pursuing and further investigating this strategy.

There are various issues I have mentioned that won't be discussed here any further: I already postponed any further discussion about the syntactic criteria for singular termhood – I will just assume that number-terms are singular terms. Also, I won't be discussing in more detail the issue about the two readings of *objected-directed* thought and which of the two views is needed for a successful account of knowledge of numbers. Also, I here merely outlined and did not argue for how the context principle resolves this first ingredient of the epistemic constraint and again, I will leave further discussion for another occasion.

I also mentioned the Caesar Problem. Elsewhere, in (Cook and Ebert, 2005), I (and Roy Cook) have discussed a version of this problem and argued that a certain strategy to resolve this version of the Caesar Problem is unsuccessful and in (Cook and Ebert, prep) we suggest a resolution of the problem. This discussion however won't be taken up here.

Finally, there is of course the crucial issue about second-order logic. (Quine, 1986) has famously claimed that second-order logic is set-theory "in sheep's clothing" which would undermine any prospects of the Neo-Fregean project. A full discussion of why his view is incorrect and how properly to conceive of second-order logic is a project worthy of another thesis.[82]

Despite all this there are many themes that *will* be picked up in the following chapters. The next chapter will be concerned more closely with the context principle and what I called the *unwelcome consequence* as well as other arguments against the context principle put forth by Michael Dummett, Kit Fine and Robert Brandom.

[82]Marcus Rossberg is currently writing such a thesis with the title *In Defence of Second-order Logic.*

Chapter 2

Defending the context principle

2.1 Two strands of internal criticism of the context principle

In this chapter I will be concerned with two strands of internal criticism of the context principle. The criticisms are internal in the sense that what is at issue is whether the principle is coherent and not whether it begs the question against, say, a nominalist or someone who wants to adopt causal constraints on reference.[1] Here, I won't be concerned with convincing the unconvinced of the cogency of the context principle. Rather, I aim to rebut various arguments that claim to show that the principle is either incoherent or else leads to unwelcome consequences. The facts that the principle has great intuitive appeal, that it helps to overcome Benacerraf's dilemma, and that it is a coherent principle, are the only positive reasons I will be able to give to the neutral philosopher for adopting the principle. Rejecting the various arguments provides at least an internal justification for the

[1] A good overview of various rounds of the broad discussion between the nominalist and platonist can be found in the first chapter of (Burgess and Rosen, 1997).

context principle.[2]

I will here distinguish two broad strands of internal criticism. The first strand of criticism assumes that the context principle is generally applicable but argues that, if it were correct, it would lead to problems since it either justifies false or contradictory statements, or leads to problematic consequences in other specific areas of discourse. So, for example, the following passage from Dummett fits nicely this first line of criticism:

> "the contradiction was a catastrophe for Frege, not particularly because it exploded the notions of class and value range, but because it showed that justification [i.e. the context principle] to be unsound. It refuted the context principle, as Frege has used it." ((Dummett, 1991a), p.225)

In the next section I will discuss in detail the crudest version of this criticism as put forward by Michael Dummett[3] and Kit Fine[4]. In section 2.4 I will briefly review more refined versions of the criticism given by Kit Fine[5] and Robert Brandom[6].

The second strand of internal criticism concerns what I previously called the *unwelcome consequence*. Here, Dummett argues that the minimal reading of the context principle plus the syntactic priority thesis entail the truth of the *unwelcome consequence*, which states that "At least some objects – such as pure abstract objects, as numbers, sets etc – whose reference is established by the context principle, exist mind/language dependently" (see p. 55 of this thesis). Consequently, one can't interpret the first two theses as justifying "real"

[2]Having made this concession, I think that alternative positions hardly provide any more motivation for their basic and constitutive claims. For example, the main proponents of nominalism have said this much in favour of their rejection of abstract entities: "Why do we refuse to admit the abstract objects that mathematics needs? Fundamentally this refusal is based on a philosophical intuition that cannot be justified by appeal to anything more ultimate."(Goodman and Quine, 1947), p. 107

[3]See mainly his (Dummett, 1981a) and (Dummett, 1991a) other reference below

[4]See his (Fine, 2002), chapter 2

[5]op.cit

[6]See his (Brandom, 1996).

reference to *pure abstract objects* such as numbers or sets. Rather, according to Dummett, we can only adopt some kind of "thin" reference, which will be insufficient for platonistic purposes. In effect, Dummett aims to show that there is an irresolvable tension between adopting the context principle on the one hand and aiming to meet Benacerraf's challenge via the platonist route on the other. If Dummett is correct, the Neo-Fregean is misguided in thinking that the context principle is of any help for his purposes.

2.2 The first strand of internal criticism: The legitimating role of the context principle

Dummett has claimed on various occasions that the context principle should be rejected because of the inconsistency of Basic Law V. His criticism is illustrated by the following two passages:

> "What mattered philosophically, however, was not the definition in terms of classes, but the elimination of appeals to intuition, a condition for which was the justification of a general means of introducing abstract terms, as genuinely referring to non-actual objects, by determining the truth-conditions of sentences containing them. The contradiction was a catastrophe for Frege, not particularly it showed that justification to be unsound. It refuted the context principle, as Frege had used it."((Dummett, 1991a), p.225)

> "For Frege's method of introducing the abstraction operator – that is, of introducing value ranges – was, notoriously *not* in order. It rendered his system inconsistent; and that inconsistency forced him eventually to acknowledge that his entire enterprise had failed. **If the context principle, as stated by Wright,**

were sound, there could have been no inconsistency." ((Dummett, 1991a), p. 188, italics by Dummett – my bolding of last sentence.)

To explain this quote briefly, Frege's method of introducing the abstraction operator is the idea that abstraction principles introduce functions which, in the case of conceptual abstraction principles, are functions from properties to objects. Frege's infamous Basic Law V, is aimed to introduce the extension (or value-range) function which maps co-extensive properties (or functions that map every object to the same value) to the same extension (or value-range). As previously noted in section 1.4.3 Basic Law V is inconsistent.

The main thought in these two quoted passages is that the inconsistent abstraction principle raises problems for the general application of the context principle. Disregarding for the moment possible differences in the context principle *as Frege used it* and *as stated by Wright*, Dummett's thought seems to be that the context principle justifies or legitimates abstraction principles in general. So, if on the basis of such justification an instance of such abstraction principles turns out to be inconsistent, then there is something wrong with its initial justification, namely the context principle. Hence, roughly we have the following short piece of reasoning:

Premise 1 The context principle legitimates abstraction principles.

Premise 2 Some instances of abstraction principles (e.g. Basic Law V), despite such legitimation, are inconsistent.

Conclusion The context principle is defective as a legitimation for abstraction principles.

As it stands, this reasoning needs clarification and disambiguation. Let me outline two ways in which the idea of justification or legitimation can be understood and analyse why Dummett might think that premise 1 is correct.

On the first reading, the Neo-Fregean uses the context principle within the context of abstraction principles, such that to be credited with an *understanding* of the new term it suffices to have this term figuring in an abstraction principle. If the abstraction principle is successful, it fixes truth-conditions of the identity statements in which the new expression figures, by means of already understood statements that occur on the right-hand side. Now, if legimitation is understood in just this sense – i.e. what the context principle *legitimates* is that to be credited with an understanding of a new term, grasping an abstraction principle is sufficient – then this type of legitimation is surely insufficient to provide a *guarantee* that abstraction principles, in general, are consistent. This reading of legitimation is concerned with the role of abstraction principles for a subject's understanding of a new term but not with whether the abstraction principle itself is "in good order", that is whether it is true or consistent.

Hence, this type of legimitation is not intended to effect the truth of the underlying abstraction principle. Nor is it prima facie necessary that the abstraction principle has to be consistent in order to provide an understanding of the new term involved. It might well be claimed – as I do in chapter 3 – that even inconsistent abstraction principles can provide an understanding of a new term. If correct, the issue of consistency is independent of this type of legitimation and therefore premise 2 above is confused since it is not the aim of the intended legimitation to *guarantee* the consistency of an abstraction principle.[7]

On the second reading, we might understand the type of justification and legitimation of the context principle as providing a *guarantee* that new terms introduced in virtue of abstraction principles figure in true/consistent sentences. This understanding of the context

[7]There is an idea in the literature often referred to as epistemic analyticity which connects understanding of a principle with having a warrant for the truth of that principle. See especially (Boghossian, 1997) and (Boghossian, 1996) for a defence of this idea, which would provide a link here between understanding and truth/consistency. In my chapter 5, I will exclusively examine the notion of epistemic analyticity.

principle would make the occurrence of inconsistent abstraction principles very problematic indeed and so, the above reasoning would make a genuine and rather worrying point. Yet, I can find no reason to believe why the content of the context principle as previously outlined, i.e. the minimal reading and the syntactic priority thesis, should entail this idea of legimitation!

Let me shortly repeat the content of the context principle, as I understand it, to make this very important and often misunderstood point more transparent. The *minimal reading* of the context principle is just the conditional that '*if* there is a singular term occurring in a true sentence (of the appropriate type) *then* there is an object as the referent of that term'. Clearly, this mere conditional statement can't provide a guarantee that the abstraction principle is true. The second component of the context principle, the syntactic priority thesis, which gives rise to syntactic criteria of singular termhood cannot – as far as I can see – provide such a guarantee either. It merely claims that linguistic criteria have priority over ontological ones. Subsequently, the context principle has no bearing on whether the abstraction principles are consistent or not. So, although the context principle provides a suitable framework in which to place abstraction principles, it does not itself say anything about the truth of abstraction principles; the context principle only has an effect (in that it guarantees reference and access to abstract objects), once it is established that the principle is true. Again, the context principle presupposes the truth of the abstraction principle rather than guaranteeing it. So, on this second understanding of legimitation, the first premise of the above reasoning is to be rejected since the context principle does not legitimate abstraction principles in the strong sense of guaranteeing every abstraction principle's truth/consistency.

This is not to say, however, that no problem is posed by inconsistent abstraction principles. Quite to the contrary, inconsistent abstraction principles are very problematic! But these problems have to be addressed within a theory of abstraction, or more broadly, within a theory of implicit definitions. What has to take centre stage in the discussion of consistency is the notion of 'stipulating an implicit definition (or abstraction principle) to be true'. Such a stipulation can only be successful once certain conditions are met; I will examine these in greater detail in chapter 3. What should be clear, therefore, is that the problem posed by inconsistent principles and the conditions for successful stipulations is internal to the theory of abstraction principles and therefore can't underwrite a reductio on the general applicability of the context principle.

Interestingly, the connection between the context principle and the theory of abstraction has often been misunderstood. For instance, in a similar fashion to Dummett, Kit Fine argues against the context principle. In his recent book *Limits of Abstraction* he dedicates a chapter discussing this principle. Therein, he supposes that "the 'context principle' can be regarded as an attempt to vindicate such contextual definitions."[8] Just as in the case of Dummett, it is unclear in Fine's writing to what extent the context principle is supposed to justify or vindicate such definitions. On various occasions, however, Fine seems to have a rather strong reading in mind, and regards the context principle as providing a justification that the abstraction principle so introduced are true.

Repeating what I said above against Dummett, I believe there is no reason to suppose that the context principle is intended to 'vindicate' such definitions in the sense of providing a *guarantee* that these definitions are successful. Rather, all the context principle (in its minimal reading and syntactic priority thesis) provides is: (a) that to be credited with an understanding of the term in question, it suffices to grasp the abstraction principle which aims to introduce the new term and (b) that the context principle secures reference of a singular term arrived at on the basis of an abstraction principle, *provided* it is a successful

[8](Fine, 2002), p. 55 Contextual definitions can be understood as abstraction principles in this quotation.

stipulation.[9] So, for these reasons, I reject this first internal criticism of the context principle, which attempts a reductio on this principle.

Let me now turn to the second type of criticism against the context principle. In contrast to rejecting the application of the context principle altogether, Dummett also suggests that one could maintain the principle "but declare that it does not vindicate the procedure Wright has in mind." ((Dummett, 1991a), p. 205). And here, what "Wright has in mind" is the thought that the context principle suffices to impose 'robust' or 'real' reference to abstract objects such as numbers. In opposition to Wright, Dummett believes that the context principle fails to guarantee a genuine reference relation to something existing independently. To this end, he argues that if the minimal reading and the syntactic priority thesis are accepted, they together entail the unwelcome consequence, i.e. that some objects (i.e. pure abstract objects) exist mind-dependently. In the next section, I examine various arguments occurring in different passages of Dummett, but I argue that they all fail to establish Dummett's unwelcome conclusion.

2.3 The second strand of internal criticism

In this section I outline and discuss three separate lines of reasoning in Dummett's writings that aim to undermine a 'realistic' or robust interpretation of the reference relation for abstract terms as introduced by the context principle.[10] The first argument is based on Dummett's very subtle elaboration of a disanalogy between understanding (or better manifesting an understanding) of concrete terms and manifesting such understanding in the case of terms standing for abstract objects. Based on this disanalogy he draws conclusions about

[9] For a similar point against Fine, see (Cook and Ebert, 2004), p. 798ff.

[10] The relevant sections of Dummett are ((Dummett, 1981a), chapter 7 and more importantly 14), ((Dummett, 1981b), chapter 18, 19 and 20) and ((Dummett, 1991a), chapters 15 - 18).

the ontological status of the objects purportedly referred to by the terms in question. The second argument, which is not yet thoroughly discussed in the literature, is based upon a distinction between abstract objects and *pure* abstract objects, from which Dummett argues for the mind-dependent status of the latter kind of objects. The third argument, which I call the *argument from conceptualization* is rather intricate, and in examining it, I also consider Bob Hale's reply to it.

Before delving into the details of the various arguments, a point on my terminology should be noted. I propose to regard the following group of terms, which Dummett often uses to capture reference, as synonymous: robust, real or realistic, external or substantial reference. I interpret them as conveying the idea of a reference relation to an object existing *mind-independently*. Also, I regard 'constituent of an external reality' as synonymous with an object existing mind-independently; therefore an object that is 'not a constituent of an external reality' is to be understood as being a mind or language-dependent object. Furthermore, Dummett's use of 'thin reference' or even 'reference being internal to language' is best understood as a reference relation to something that does exist merely mind- or language-dependently. The notion of 'thin' reference will be discussed further in section 2.3.1. Lastly, whenever Dummett talks of a 'contextual definition', this can be understood, for our purposes, as a definition by abstraction, i.e. an abstraction principle.

2.3.1 The argument from disanalogy

In this section I will first present Dummett's argument from disanalogy and then provide various lines of reply to guard against the conclusions drawn from it.

Dummett[11] motivates a disanalogy between how we normally learn the use of singular

[11]Dummett notes the disanalogy in (Dummett, 1981a), p. 494 in the context of Goodman's nominalism. The use of the disanalogy as an argument is on p.499 and most explicitly in p.505 middle paragraph.

terms purporting to refer to *concrete* objects and those that "seem" to refer to *abstract* objects. The idea is roughly the following: In the concrete object case, we establish the reference of a term by the so-called *name-bearer prototype*, in which intuition – that is a perception of an object or a cognitive encounter with the object – plays a direct role in the identification and re-identification of the referent of the term. So, according to Dummett, to manifest an understanding of a newly introduced term purporting to refer to a concrete object is to exhibit the ability to identify the referent of this new expression – for example by understanding demonstrative phrases in which the new term occurs.[12]

In clear disanalogy, Dummett contends, this kind of manifestation does not seem to be possible for expressions purporting to refer to abstract objects. The role of the referent is minimized and it seems impossible – just because the object is abstract – to use the name-bearer prototype to establish a reference relation for the term. So, to manifest an understanding of this new term does not involve an appeal to the object purportedly referred to.

Based on this disanalogy, Dummett concludes that abstract terms do not have genuine reference (i.e. reference to something externally existing). The inference to this conclusion is supported by the claim that intuition has to play a certain role in the *identification* of the referent for there to be a genuine reference relation to something external; and if it doesn't, the term has no genuine referential relation to something external. Differently put, for the case of abstract terms the role of reference is *semantically idle* as the referent does not play a role in determining the truth-value of statements in which it figures, i.e. no intuition is involved in determining the truth-value of statements involving the new term. Therefore,

It reappears using slightly different vocabulary (such as the notion of being semantically idle vs. being semantically operative) in (Dummett, 1998), p.384ff.

[12]Dummett is aware of Wittgensteinean consideration that mere ostensive definitions could not establish this relationship. However, Wittgensteinean worries will be left untouched in my discussion.

neither the prototype of a name-bearer relation is used to introduce the term, nor the alleged bearer plays any role in manifesting an understanding of the term.

As a result, Dummett introduces a less substantial "thin" notion of reference on grounds that there is a clear disanalogy between prototype cases of how reference functions in concrete cases and cases where abstract objects are involved. Compare the following quotation of Dummett concerning the reference of new terms introduced in virtue of contextual definitions:

> "In such a case [of contextual definition], no view stronger than an intermediate one could be taken of a claim that a reference had thereby been conferred upon them [the new terms]; the reference so conferred would be reference only in the *thin* sense [. . .], since the notion would play no role in the semantic account of how the truth-values of sentences containing the terms are determined." ((Dummett, 1991a), p.236)[13]

As it stands, I believe that Dummett's *argument from disanalogy* can be challenged on at least three grounds:

1. Grant the disanalogy but show that something else can be regarded as fulfilling the role of the name-bearer prototype in the abstract case, which thereby guarantees a real reference relation for abstract terms.

2. Challenge Dummett's crucial transition from the claim that intuition is not necessary for manifesting our understanding in the case of abstract singular terms to the claim that therefore they don't refer realistically.

[13]See also the following section: "The proponent of the intermediate view of terms introduced by contextual definition – the view for which I have here argued – maintains that the thin notion of reference will not bear the weight of a realistic interpretation of those terms;" ((Dummett, 1991a), p.198)

3. Challenge the coherence of Dummett's own proposal concerning "thin" reference.

The first line of response is very interesting and has been extensively pursued by Wright[14], but I will leave a detailed discussion of this strategy aside. I do believe that this discussion ends in a stand-off between Dummett and Wright and as such won't be sufficient to undermine Dummett's argument. Instead, I will focus on what I think is wrong with Dummett's argument by focusing on the second strategy and then raise some doubts about the coherence of his position by pursuing strategy three.

First rejoinder: Dummett's illicit inference

In order for Dummett's argument from disanalogy to have force he requires a strong assumption, which consists in moving from the role an object plays in our *understanding* of the term referring to it, to conclusions about its *ontological* status. More precisely, Dummett's argument starts out with the intuitive claim that in the case of concrete objects where we do have real reference, intuition plays a crucial role in manifesting our understanding and thereby settling the reference of a term. In the case of abstract terms, intuition does not play any such role in our *understanding* of the term, a point that again is intuitive. But from this, Dummett draws the *ontological* conclusion that the object referred to does not exist externally, i.e. mind-independently. However, what is it that underwrites such a transition? I think that the transition cannot be upheld in general. In order to highlight my worries consider the following two questions:[15]

1. Do we need intuition to justify or manifest that we understand new terms figuring in

[14]See, for example in ((Wright, 1983), p.78) when he writes that: "Knowledge of the reference of an abstract singular term, if taken to be a recognitional ability at all, thus has to be construed as the ability to recognize the truth or falsity of identity statements involving that term." Section x of (Wright, 1983) pursues this strategy in detail.

[15]The distinction I am about to make is similar to one made by (Hale, 1987), p.165ff.

certain sentences (say sentence S)?

2. Do the objects referred to by the new terms involved in the sentence S exist mind-independently?

If I understand Dummett correctly, he is committed to answering the second question with "no" provided that he returns a negative answer to the first. But I find it hard to detect any compelling reason to think this being correct. Consider, for example, sentences in modern theoretical physics involving mainly theoretical terms (such as "strings", "neutrinos", etc.). In this case manifesting our understanding of such terms can hardly involve any intuition. Still it does not seem to have any bearing on the ontological status of the objects involved. I therefore regard the general application of this transition as questionable at best; and so the onus is on Dummett to provide some reasons why there is such a strong link between these two issues. Failing to do so, we are entitled to reject the argument as it stands.

However, there may be a weaker reading of the argument, namely, that based on the disanalogy, the *default reason* to think that *abstract* terms refer realistically is lacking. Hence, unless there is independent reason to think otherwise – and arguably in the physics case there is – we should conclude that such terms don't genuinely refer to mind-independent objects.[16]

Even granting this much, I think that, dialectically, some reason for making the transition from understanding a term to the ontological status of the object referred to is required. Otherwise it is unclear why the name-prototype is a default reason for mind-independent existence. Furthermore, the platonist might well be credited with having independent reasons for thinking that numbers and sets exist mind-independently. So, for example the belief that mathematical truths are "eternal" could give rise to the belief that the objects referred to are mind-independent. Either way, therefore, Dummett's argument seems to involve an

[16] I knowingly pass over agnosticism as an alternative conclusion. It seems to me that, at best, Dummett's arguments only give rise to agnosticism.

unmotivated assumption and even the more favourable version of the argument (i.e. that the disanalogy shows that the default reason for real reference is lacking) can easily be challenged.

I now turn to the conclusion of Dummett's *argument from disanalogy* and challenge the coherency of his own conception of "thin" reference.

Second rejoinder: Some scruples about "thin" reference

Dummett claims that there is a type of reference – "thin" reference – which is a genuine reference relation, even though it lacks the realistic underpinnings of robust reference. Unfortunately, Dummett says very little about the characteristics of this notion of reference. The aim of this section is, firstly, to raise doubts as to whether Dummett is entitled to claim that "thin" reference is a reference *relation* and, secondly, to suggest two ways of interpreting this notion and then examine how they might play a role in the current dispute.

First, let us investigate why Dummett thinks he is entitled – within his framework – to assume that abstract terms actually do refer, but merely "thinly". As noted, Dummett regards robust reference as a relation that only holds if it is embedded in intuition and in this way the term becomes semantically operative. But then, if there is no intuition and nothing else in which the reference relation is embedded, on what grounds can Dummett maintain that in the case where the terms are not semantically operative (i.e. they are semantically idle), there still is a genuine, yet only "thin" reference relation? Why isn't it the case that these are not referring terms?

To highlight this problem, let me put forth the following dilemma for the Dummettian position: Either it offers an alternative to intuition which underwrites the reference relation in case of abstract terms; but then it could be argued – along the first line of response which I did not discuss in any detail – that these grounds suffice to ensure "real" reference and not merely "thin" reference. Or, alternatively, Dummett refuses to provide an alternative to intuition, which would then open up the option to reject any type of reference relation – even "thin" reference. Hence, it seems Dummett's position is on shaky grounds and I hope this dilemma highlights this worry about the coherence of the notion of "thin" reference.

Leaving aside this dilemma as an open challenge, I will turn to the second issue and disambiguate various interpretations of Dummett's notion of "thin" reference. The aim is to clarify what exactly is it that is supposedly "thin" about this sort of reference. Two possible interpretations can be considered.

1. That the relation of reference is somehow itself "thin" or deflated.

2. The referent, i.e. the object referred to, is "thin" in some sense.

Concerning the first interpretation the question is what it is for the relation itself to be "thin". I will suggest two ways this could be made sense of. One possible interpretation is that "thin" reference for abstract terms means that the relation holds at best indeterminately.[17] However, if that is the intended view, then it is not immediately obvious why the notion of "thin" reference doesn't also apply to concrete objects. In addition, usually the idea of indeterminate reference is not regarded as in any way less "substantial" than the normal one. It is normally regarded as still picking out (albeit indeterminately so) independently existing objects. So, it seems that Dummett's views about "thin" reference have to be understood differently.

A further suggestion could be extracted from earlier writings of Dummett. In his *Frege: Philosophy of Language* he distinguished two types of reference relations, normal or *name-bearer prototype* reference and reference as *semantic role*. Might the latter be of help interpreting his later notion of "thin" reference – this despite the fact that Dummett never

[17]See for example (Fine, 1975) where he provides a formal framework, i.e. supervaluationist semantics for indeterminate reference.

explicitly identified them?[18] The difference between the two notions is that the normal (or prototype) reference is regarded as a relation between a term and something extra-linguistic, whereas reference as semantic role is defined by the contribution the expression makes in the determination of the truth-value of sentences in which it occurs.[19]

The main difficulty with this interpretation, however, is that if "thin" reference is just the latter semantic role notion of reference, then it seems highly questionable whether "thin" reference is not just a denial of a genuine reference *relation* altogether. Dummett's notion of reference as semantic role is dangerously close to the notion of Fregean Sense which does not involve a relation to an object. Thus this reading would undermine the idea of a genuine but "thin" reference relation; it is therefore generally unclear what the interpretation of "thin" reference in the sense of a "thin" relation of reference could amount to.

Alternatively, and more intuitively, we might interpret the notion of "thin" as applying to the referent instead of the reference relation. If that is the case, what Dummett should have been talking about are "thin" objects or referents and not "thin" reference! Consequently, to invoke the idea of "thin" reference is just another way to claim that abstract terms do refer *as usual*, but that the objects referred to are mind-dependent (as I initially explained this notion in section 2.3).

If that is the right way of looking at Dummett's notion of "thin" reference, and if, as I argued, the argument from disanalogy fails (and even raises doubts about the internal consistency of Dummett's position), then an alternative argument is needed to establish the *ontological* claim that abstract terms refer to mere "thin" objects. In the next section I discuss such an argument, but, to forestall any expectations, my conclusion will again be

[18]His discussion of reference as semantic role took place 20 years earlier in (Dummett, 1981a) chapter 7, p.204-45. Hale (Hale, 1987) and Wright (Wright, 1983) often used the idea of semantic role as an interpretation of "thin" reference.

[19]See (Dummett, 1981a), p.210f

that Dummett's argument fails to convince.

2.3.2 Pure abstract objects and the argument from analyticity

Dummett's second argument is based on a distinction concerning the nature of objects. He distinguishes abstract objects in general from what he calls *pure* abstract objects. The former are objects to which terms refer that either have been introduced contextually, i.e. through abstraction principles involving concrete objects on the right-hand side (such as in the case of the direction abstraction principle), or terms that are introduced by a demonstrative combined with a so-called functional expression.[20] Pure abstract objects, in contrast, are characterised as those "whose existence *may be recognised* independently of any concrete object, and therefore independently of any observation of the world."[21] And for Dummett these are objects to which terms refer that are introduced in virtue of second-order abstraction principles, where one does not abstract on objects but on concepts, as is the case in Hume's Principle. On another occasion, Dummett characterises pure abstract objects as those "whose existence is analytic."((Dummett, 1981a), p.505).[22]

Granting Dummett the tenability of such a distinction between abstract and pure abstract objects, Dummett aims to show that pure abstract objects can't be "constituents of an external reality". One argument in which the notion of a *pure* abstract object seems to play a role is in the following quotation:

"..., but for, say, shapes of physical bodies the sequences of concrete objects, the

[20]An example of a functional expression is the notion "shape", which can be introduced with the statement "the shape of this figure" accompanied with an appropriate pointing gesture. See Dummett on "functional" expressions in (Dummett, 1981a), p.176-9. This characterisation of abstract objects is merely suggestive and not intended as strict conditions. More exact conditions need not concern us here.

[21](Dummett, 1981a), p.504, my italics.

[22]I assume what Dummett here means is that an existential statement involving such objects is an analytic truth, since the notion of analyticity normally applies to statements and not objects.

use of these terms is still clearly related to processes of observation of the external world and identification of constituents of it. For that reason, therefore, it is still possible to apply to such terms the notion of reference, construed realistically as a relation to something external; although, indeed, the further we travel along the scale, the more stretched becomes the analogy with the prototypical case. *It is only when we reach terms for pure abstract objects, however, that the thread snaps completely, and we are concerned with the use of terms which have no external reference at all.*" (Dummett, 1981a), p.510

There are at least two readings of this remark. The first is one in which the "snapping of the thread" is due to the fact that the prototypical case of reference does not apply and so we are back at the earlier *argument from disanalogy.* The second reading, however, assigns the notion of a *pure abstract object* a distinct role in arriving at the conclusion that pure abstract singular terms don't have external reference. In a previous section, Dummett mentions the following additional consideration about pure abstract objects which might support this reading:

> "But the picture does seem to require that what may be called a 'constituent of reality' is something which can be encountered; and, if the existence of something is an analytic truth, a recognition of its existence can hardly be held to constitute an encounter." ((Dummett, 1981a), p. 503)

Combining this thought with the first quotation, we can reconstruct the following *argument from analyticity*:

Premise 1 In order for a term to have external reference, minimally, the object it purports to refer to has to be a 'constituent of reality'.

Premise 2 For an object to be a constituent of reality it has to be encounterable in some sense.

Premise 3 Objects whose existential statement is an analytic truth are not encounterable in any relevant sense.[23]

Conclusion "The thread snaps" and pure abstract terms do not have reference to an external object.

Now, since the argument is aimed at a platonist, we need to be careful in interpreting load-bearing notions involved such as "constituent of reality" and "encounterable". It would be question-begging to take "constituent of reality" in premise 1 and 2 to require that such a constituent has to be spatio-temporally located. Furthermore, Dummett's notion of "encounterable" can't just be understood to mean "perceivable by a subject" or "capable of causal interaction". Consider, for example, premise 3 and assume with Dummett that numbers are objects whose existence is analytic. That is, according to Dummett, we recognise their existence without encountering them. Certainly, if "encounterable" is here understood as "capable of causal interaction" then premise 3 could well be regarded as true (even by the platonist), yet premise 2 then blatantly begs the question. So crucially, in order to run the argument without begging the question, what is needed is an interpretation of the notions "constituent of reality" and "encounterable" which avoids these problems. In addition, we need an understanding of these notions which allows abstract objects but not pure abstract objects to be encounterable or to be constituents of reality. This latter constraint is needed in order to respect Dummett's rejection of nominalism.

Subsequently, I find it hard to offer any suitable understanding of these notions for

[23] I simplified this point. Dummett talks about "a *recognition* of its existence" not involving an encounter; but I assume that this entails, in general, that recognising their existence cannot involve an encounter.

Dummett's purpose which respects the above constraints. Lacking this (for what it's worth: I don't think that there is such an understanding), I can only but deny the compellingness of the argument. Dummett's argument is *crucially* unclear when it comes to the important concepts and so, pending further clarification, I conclude that the argument from analyticity fails to impress the opponent.

2.3.3 The argument from conceptualization

A further argument, that I will label the *argument from conceptualization*, can be found in Dummett's work.[24] The following quotation provides the gist of the argument:

> "When we apply the conceptual apparatus with which language supplies us to reality, this results in the discernment of a variety of objects, concrete and abstract: but the apparatus is such that certain objects will be recognized however the reality is constituted to which we apply it; these are pure abstract objects, like the natural numbers, whose existence is analytic. This is incomprehensible if we think of the world as composed of objects, as coming to us already segmented into objects: in that case, how could there be a whole plurality of eternally existing, uncreated objects? But, once we realize that *our apprehension of reality as decomposable into discrete objects is the product of our application to an unarticulated reality of the conceptual apparatus embodied in our language* [1], it should not be particularly surprising that *certain objects should result from this operation no matter what the reality is like*[2] to which it is applied.
>
> Perhaps not: yet for that reason it appears *impossible to regard the pure abstract*

[24]Interestingly, it disappears in his later writing although the argument plays a figurative role in the chapter entitled *Abstract Objects* in (Dummett, 1981a).

objects as constituent of an external reality.[3] ((Dummett, 1981a), p. 504-5 my italics and additions)

Again, I will try to transform this reasoning into a more perspicuous argument-form, followed by a brief discussion of Hale's reply to the argument[25] and my own evaluation of the various points raised.

The first premise [1] is Dummett's claim about how the structure of language shapes our apprehension of reality. It is what I call the *Objects qua Conceptualization Thesis (OCT)*:

Premise 1 (OCT) The apprehension of reality as having distinct objects is dependent upon the usage of the conceptual apparatus supplied by language.[26]

This premise is somehow vague. It is not clear what exactly this dependence amounts to. Furthermore, note that what is claimed is not that the *existence* of objects is somehow dependent upon our language – a precariously strong claim – but rather that our *apprehension* of reality as having objects depends upon our conceptual scheme. Also, Dummett thinks this premise follows directly from the adoption of the context principle (more on this below).

The second point [2] Dummett makes is a restriction to *pure abstract objects.*

Premise 2 Certain objects may "result" from the operation of conceptualization alone, i.e. no matter what reality is like.

Again, this premise needs some sharpening. I shall abandon Dummett's modal talk and assume that his point in premise 2 is that, in fact, certain objects, such as pure abstract objects, *do* "result" *solely* from operations of conceptualization. Interestingly, this premise seems not

[25] See (Hale, 1987), p. 157

[26] Similar expressions of the same thought are "what objects we recognize the world as containing depends on the structure of our language" ((Dummett, 1981a), p.503), "it is we who, by the use of language,[...] impose structure on it."(op.cit, p.504) and "it is only because we employ a language [...] that we learn to slice the world up, conceptually, into discrete objects." (op.cit, p.407)

to concern our apprehension of such objects as resulting from conceptualization but suggests that it is the objects themselves that result from these operations of conceptualization.

Lastly, putting the two premisses together, Dummett arrives at the conclusion:

Conclusion Pure abstract objects are not constituents of an external reality.

There is much to say about the argument. On first sight, it seems, at best, as if various steps are missing in order to arrive at the conclusion. One thought might be that there is something similar in play as in the previous *argument from analyticity*: Because certain objects result from conceptualisation and their existence is "analytic", it follows that they can't be a constituent of an external reality. However, there is another interpretation of Dummett's argument that is independent of the above *argument from analyticity* and that has been suggested by Hale.[27]

Hale's strategy is to show that Dummett relies on an implicit assumption for the argument to go through. By highlighting it, he offers a new interpretation of the structure of Dummett's argument. To reveal this assumption he questions what exactly restricts the denial of an external reality to pure abstract objects and why it doesn't apply to impure abstracts or even concrete objects. The reply on behalf of Dummett is stressing premise 2, namely that pure abstract objects result solely on the basis of our conceptual apparatus and *no matter what the reality to which we apply it is like.* However, Hale questions why this should constitutes a relevant difference and concludes that there is a relevant difference only if the following additional claim is adopted. Hale writes:

> "It would be so [i.e. that there would be a relevant difference], if it were assumed that, for objects of some kind to be constituents of an external reality, their existence must be a contingent matter – that is, it must depend upon just how the

[27]See (Hale, 1987), p.157

world is, as a matter of fact, in respects to which it could have been otherwise." ((Hale, 1987), p.157)

So, premise 2 makes a relevant difference only if the additional thesis, call it *Mind-independent Existence is Contingent (MEC)*, is assumed,

> "For objects to exist externally, i.e. mind-independently, their existence must be a contingent matter, i.e. it is dependent on how the world is and thus could have been otherwise." (op.cit.)

and only then Dummett's argument would go through. However, Hale does not stop here, but continues to argue that this thesis (MEC) is implausible because it relies in turn on another, even more problematic assumption. He writes:

> But why make this assumption [MEC]? Well, that would, it seems, be a fair assumption, if, but only if, it were the case that all necessity is of our making - the inevitable, but metaphysically innocuous by-product, as it were, of our efforts of conceptualization. In short, if this diagnosis of the underpinnings of Dummett's argument is correct, its capacity to undermine platonism depends, after all, upon the tenability of some sort of conventionalist reduction of necessity. (op.cit.)

Hale's interpretation is revealing but some further questions remain. First and foremost, what role exactly is OCT (premise 1) playing in the original argument? The following reconstruction based on Hale's interpretation seems valid as a self-contained argument:

Premise 1* - MEC For objects to exist externally, i.e. mind-independently, their existence must be a contingent matter.

Premise 2* Certain objects, such as pure abstract objects "result" from the operation of conceptualization alone, i.e. no matter how the reality is like and thus exist necessarily.

Conclusion Pure abstract objects do not exist externally and so do not exist mind- independently.

So, MEC, plus the thought that pure abstract objects exist necessarily, are already sufficient for the conclusion and thus premise 1 – i.e. OCT – is irrelevant.

Second, it is not clear that adopting a conventionalist view on necessity is necessary and sufficient for MEC. It seems possible to adopt MEC and reject the conventional character of necessity and, conversely, it seems equally possible to adopt a conventionalist account of necessity and reject MEC. Only if the additional thesis is adopted that the metaphysical status of a kind of object (as being mind-dependent or mind-independent) just is a question about the metaphysical status of truths about these objects, can a link be established between the two claims. It is noteworthy that Hale explicitly adopts this additional thesis throughout his discussion. He writes, "what is required for those objects to be 'external' or constituents of a *mind-independent* reality [...] is best seen as, at bottom, a question about the (metaphysical) status of the *truths* of the corresponding sort - viz. do these *truths* hold independently of us/of our thought and talk?" ((Hale, 1987), p.165). Here, I don't want to deny this link but rather point out that there is logical space for alternative views.

More importantly, let me offer some evaluation of Dummett's argument. First, if it relies on MEC, then I think it should be rejected. MEC is highly counter-intuitive. So, for example, according to MEC, it is a necessary condition for an object to exist mind-independently that its existence is a contingent matter. So, to take the example of God: Assume that if God were to exist, he would exist necessarily, but then, according to MEC (contraposed), he would exist mind-dependently, which is rather counter-intuitive. Also on this conception, pure sets, i.e. sets without individuals as their members would exist mind-dependently. Yet, it remains open whether objects such as impure sets, i.e. objects such as the singelton of Dummett, which, at least prima facie is a *contingently* existing abstract object, exists also mind-independently. This seems to me is a rather problematic situation.

A friendly suggestion which might help to provide more motivation and an alternative explanation of Dummett's hard-to-follow reasoning might come from the "artefact of the model" metaphor.[28] Take for example first-order logic. In this case, one might say that it is an "artefact of the model" or an artefact of the semantic-theory that at least one object has to exist.[29] Here, we tend to say that first-order logic does not "really" have ontological commitments to at least one object, but rather claim that the existence of this object is a mere artefact of doing semantic theory. Obviously, it is very hard to decide in which cases we can speak of artefacts rather than genuine commitments; but this need not concern us here. For Dummett, it seems that certain objects, such as pure abstract objects, that come about "no matter how reality is like" are in this respect mere "artefacts of our conceptualization". They come about through conceptualising and thus can't be regarded as existing genuinely mind-independently. Maybe it is a consideration like this that underlies Dummett's *argument from conceptualisation.*

However, even if that is so, Dummett has not yet explained why, adopting this talk of artefacts in the case of pure abstract objects, these kinds of objects would qualify as genuine artefacts. More precisely, he has failed to explain away the alternative explanation of why pure abstract objects always exist in any conceptualisation: It is just *because* objects such as pure abstract objects do exist necessarily and mind-independently! Therefore, a correct conceptualization of the world will always involve such objects. So, on this (platonist) view,

[28]This interpretation of Dummett, however, still does not rely on OCT.

[29]A more controversial example might be exact borderline cases for "vague" predicates or even higher-order vagueness!

the order of determination is reversed and it is not because of our conceptualization that certain objects exist or result (that is the artefact view); rather, it is because certain objects exist necessarily and mind-independently that we end up always conceptualising these object (provided it is a correct conceptualisation).

So, to conclude, I believe that Dummett's argument from conceptualisation also fails to impress his opponent. This is because, firstly, if Hale's interpretation is correct and Dummett's argument relies on MEC, then he needs to provide a genuine motivation for adopting MEC, as well as explain away the various counter-intuitive consequences which I have put forth above. Secondly, we have seen that Dummett's own conception could be better explained by taking pure abstract objects as a kind of artefact. This interpretation of the argument, however, makes apparent that there is an alternative explanation of the phenomenon of pure abstract objects, one that is both well motivated and clearly accepted by his opponent. Hence, from this perspective of evaluation, Dummett's consideration is not convincing. Lastly, it is worth noting that if, alternatively, the argument relies on a conventional account of necessity, then too this is problematic not only because it seems to lead to MEC (if the above link that Hale adopts is accepted) but due to various convincing arguments against such an account of necessity, which I can't cover here.[30]

One remaining question is whether there is a further, alternative interpretation that makes use of Dummett's initial premise 1. In the following section, I will suggest a reading which should not be understood as a direct argument against pure abstract singular terms and their referential ability, yet it does exhibit a potential tension. As Dummett writes: "There is, undisputably, a considerable tension between Frege's realism and the doctrine of meaning only in context [context principle]: the question is whether it is a head-on

[30]See especially (Quine, 1954) and (Quine, 1935).

collision."[31]

An alternative interpretation of Dummett's argument from conceptualization

The aim is to try to develop an at least *prima facie* tension within the Neo-Fregean view between the adoption of the context principle on the one hand and platonism on the other. The rough tension can be characterised in the following way by reconsidering premise 1. As Dummett writes:

> "The apprehension of reality as having distinct objects is dependent upon the usage of the conceptual apparatus" and thus that "what objects we recognize the world as containing depends on the structure of our language" ((Dummett, 1981a), p.503).

According to Dummett, this claim follows from the context principle. The reason is that language-dependent apprehension flows from the *syntactic priority* component of the context principle.[32] Consequently, one can arrive at premise 1 of Dummett's argument because it is through singular terms and linguistic categories in general that we explain and grasp things as objects – our apprehension of objects depends (in parts) on our linguistic framework.

Why might this idea be regarded as in tension with platonism, i.e. the thought that pure abstract objects exist mind-independently? One might argue that if certain abstract or pure abstract objects do exist mind-independently, any reference to these objects is a relation to something whose existence and character is viewed as independent of our modes of conceiving of them and our modes of talking about them. This, so one might think, is in

[31](Dummett, 1981a), p. 499

[32]As a reminder: this comprises the thought that linguistic categories have priority over ontological ones. So, the criteria for singular termhood can be established without prerequisite appeal to the objects referred to by singular terms.

tension with premise 1 that our apprehension of objects crucially depends on our linguistic framework.

Maybe, but only maybe, it is this alleged basic tension that Dummett discerns within the position of Neo-Fregeanism. He thinks that such a tension can be avoided for concrete objects, but that it is problematic in the case of pure abstract objects. In the following, I want to show that even this worry can be avoided but that it nevertheless raises important questions that will receive answers in the following chapters.

There is an important and crucial constraint on whether we can apprehend or grasp objects through singular terms as entailed by the context principle: Namely, singular terms have to figure in *true* sentences. And, it is perfectly compatible with the context principle that our apprehension of reality is, in certain circumstances, incorrect. This can happen, if, to speak crudely, the world does not co-operate and so the sentences in virtue of which "apprehension" apparently takes place actually turn out to be false, despite initially good reasons for thinking otherwise. Consequently, we can only legitimately say that there is an object as referent of an abstract singular term if we have sufficient reasons to hold such sentences true – and even then such reasons may be considered defeasible. Furthermore, if the notion of *truth* for the relevant sentences is sufficiently robust, then there is no tension between our apprehension of the objects in question being language-dependent on the one hand, but only appropriate, if those sentences correctly reflect how the world is (the mind-independent truths) on the other.

Nevertheless, a remaining worry in the context of Neo-Fregeanism might be that *stipulations* of certain abstraction principles that "automatically" make these statements true either imply a conventional element or – broadly speaking – can't involve a notion of truth "robust" enough. So, one might think – and this may be Dummett's tension – that if we arrive at an apprehension of objects purely in virtue of stipulations whose truth is in some sense "man-made" then these objects can't also be regarded as mind-independent.[33]

Importantly, however, if that is the issue that gives rise to Dummett's tension between the context principle and platonism, then the tension is wrongly located: It is an issue for a defender of abstraction principles – or implicit definitions in general – involving stipulations and platonism; not for a defender of the context principle and platonism.[34] It is perfectly consistent to adopt both the context principle and platonism yet reject the idea of stipulation as a method for introducing true sentences. Hence, if there is a genuine tension as suggested in the previous paragraph, it has to be resolved within a theory of implicit definitions and abstraction principles which appeals to the idea of stipulation – a topic we will pick up in the third and fifth chapters.[35] First, though, I want to review some other (selected) criticisms of the context principle.

2.4 A brief overview of other criticisms of the context principle

Let me finally note and briefly discuss two other criticisms of the context principle. The two arguments won't be discussed in full detail since both of them rely on additional assumptions which would take me too far afield to review. Nevertheless, I believe it is worth noting these additional worries (especially the second) as they have received much attention. However, I believe, I can make a prima facie case that neither is very threatening to the proponent of the context principle.

[33]This is a very informal way of putting this tension, but making the tension clearer would quickly lead to a resolution of it.

[34]The discussion of the first internal criticism highlighted a similar point.

[35]See 3.7 and more importantly in 5.8.

2.4.1 Fine's argument from the Caesar problem

In the *Limits of Abstraction*[36] Kit Fine argues on various grounds against the context principle. Although, Fine is not very clear on what he takes the content of the context principle to be – and so it is hard to evaluate the success of his arguments – I will note one general line of argument against the principle.[37] On p.68ff of *Limits of Abstraction* he provides the following reasoning against the context principle.

> "I wish in this section to propose a schematic solution to the Caesar Problem [...]. We shall see that although the proposal applies in a relatively straighforward manner to definitions by abstraction, its application to other forms of contextual definition sanctioned by the context principle is somewhat problematic." ((Fine, 2002), p. 68)

Fine's underlying thought seems to be the following: Because certain contextual definitions based on the application of the context principle are incompatible with his solution to the Caesar Problem, we should be sceptical either about such contextual definitions or the context principle in general.

Here, remember an important point I made earlier when discussing Dummett's criticisms. The context principle (at least as I outlined it) and the theory of abstraction, although connected in spirit, should be separated as distinct theses. From this perspective it is all but clear what the distinction is that Fine draws between definitions by abstraction and other contextual definitions *sanctioned* by the context principle. Also, just as in section 2.2 we need to be careful how the notion of *sanctioning* is understood. It seems reminiscent of Dummett's thought that the context principle somehow legitimates or, as Fine says,

[36](Fine, 2002)

[37]One type of argument that Fine suggests is already discussed and rejected in section 2.2

vindicates contextual definitions. As argued before, it isn't part of a proper understanding of the context principle that it guarantee an abstraction principle to be "in good standing", i.e. to be consistent.

Leaving aside a discussion of Fine's resolution of the Caesar problem, the general structure of Fine's argument limits its dialectical strength. First of all, it is hard to comprehend that a resolution to the Caesar problem is only applicable to definitions by abstraction and not to other contextual definitions[38], unless trivially so, since the Caesar problem arises due to a specific feature of abstraction principle which is not shared by "other forms" of contextual definitions. So, no resolution applies to the latter, since no resolution is needed for them. But then, why should this lack of solution constitute a problem for these "other forms of contextual definitions as sanctioned by the context principle"?

Secondly, various proponents of the context principle have suggested solutions to the Caesar Problem[39] and Fine has not yet shown that these fail. Overall, I don't regard Fine's general reasoning as too worrying for the proponent of the context principle, though on the proviso that further research and development of these ideas might well lead to a more succinct criticism.[40] Let me therefore turn to a more widely discussed criticism of the context principle.

2.4.2 Brandom's argument from complex analysis

Like Dummett's, Brandom's argument is intended to exhibit a tension for any position that combines the context principle with a version of platonism. However, in contrast to Dummett, Brandom does not think that the tension applies generally; rather, the tension

[38]Fine (obviously) adopts here a broader conception of contextual definition and does not confine them to abstraction principles.

[39]See the most recent version of a solution by (Hale and Wright, 2001b)

[40]The above criticism was noted in (Cook and Ebert, 2004).

emerges only once we consider the special case of complex numbers. In his argument, Brandom doesn't worry about whether the context principle leads to an absurd conclusion or whether it entails a specific view about the metaphysical status of objects referred to. Instead, he is concerned with the uniqueness of the object picked out by a singular term. He writes:

> Here is my claim: In the case of complex numbers, one cannot satisfy the uniqueness condition for the referents of number terms (and so cannot be entitled to use such terms) because of the existence of a certain kind of symmetry (duality) in the complex plane.((Brandom, 1996), p.298)

In the following I will provide a brief reconstruction of his argument, followed by various suggestions for how the proponent of the context principle (and platonism) can evade his criticism. His argument goes as follows:

Context principle If a singular term occurs in a true sentence then the term refers to an object

Uniqueness condition For a term to refer to an object entails that the term determinately picks out an object

Assumption The complex number term "i" is a singular term and complex analysis is true.

Intermediate Conclusion I The complex number term "i" determinately picks out an object

Symmetry of complex field The complex field gives rise to a specific automorphism whose upshot is that "swapping each complex number for its complex conjugate (that is i with -i) leaves intact all the properties of the real numbers, all the properties of the complex numbers, and all the relations between the two sorts of numbers." ((Brandom, 1996) p.300)

Indermediate conclusion II It is indeterminate what the complex number term "i" refers to.

Conclusion Reject the [context principle] to resolve the dilemma between [intermediate conclusion I] and [intermediate conclusion II].

There is much to say about this argument and a chapter could be spent properly discussing it, so I will confine myself to merely suggesting various solutions that seem acceptable to the proponent of the context principle. Interestingly, Brandom's argument only relies on what I called the *minimal reading* of the context principle which seems a rather benign principle.

First to note is that the conclusion as presented should involve a disjunction and so it should instead read as follows:

Conclusion* Reject either the [context principle], [uniqueness condition], [assumption] or [symmetry of complex field] to resolve the dilemma between [intermediate conclusion I] and [intermediate conclusion II].

The [symmetry of complex field] one might say is a mathematical fact and hardly challengeable. Furthermore, Brandom takes [assumption] for granted without further elaboration and regards [uniqueness condition] as being part of the claim involved in the [context principle] which is why he quickly arrives at a rejection of this premise.

Various options, however, are available for defending the context principle against this reductio and I believe that challenging [assumption], namely that complex number terms are singular terms, might well be an option. However, I think that the [uniqueness condition] is the one claim that seems the most obvious to challenge.

As Brandom himself suggests in the final part of his paper, the demand that reference has to be uniquely specifiable can be loosened. What the symmetry consideration shows is that for complex numbers we can't provide a *categorical specification* of the referents; but that doesn't mean that there can't still be a *hypothetical specification*, one that provides a specification in relation to other complex numbers. This might be one way to argue for a resolution of the argument from complex analysis.

Connected to Brandom's initial suggestion, one could alternatively adopt a form of epistemicism concerning complex numbers and argue on such grounds that the argument above does not present a genuine dilemma. All that the symmetry considerations for complex numbers show is that it is impossible to determinately specify the properties of i and -i respectively. We can, however, specify them in relation to each other, but then whether "i" refers to i or -i can't be settled due to the existence of an automorphism. This much can be granted without thereby contradicting the [uniqueness] conditions since it merely says that if a singular term refers, then it refers determinately. There is no contradiction because what the symmetry consideration shows is that *we* cannot properly *specify* or *individuate* i and -i; but that is not yet sufficient to undermine the claim that the term "i" refers determinately to one or the other. The epistemicist will hold exactly this: the singular term does determinately refer, yet we cannot determinately specify which object it refers to. Hence to summarise, while the uniqueness constraint is about the reference relation, the symmetry considerations concern the specifiability or individuation of the referent. Not being able to do the latter need not involve a denial of the former.

Again, I think much more could be said about this argument. However, I believe this discussion shows that Brandom's argument from complex numbers against the context principle is surmountable and so does not provide a knock-down argument against the proponent of the principle.

2.5 Conclusions and outlook

My proviso at the beginning of the chapter was that it is mainly negative in character: I have not provided new positive reasons for adopting the context principle, but instead I have examined and rejected the most common criticisms of the context principle. To this end, I have separated six different arguments. The first four arguments (from Dummett), I regard as flawed. I have also shown that there is little reason to be worried about the final two arguments. As a result, I think that the context principle neither leads to unacceptable consequences[41] nor is incompatible with platonism. Rather, I hope to have shown that the context principle – the minimal reading and the syntactic priority thesis – is still alive and kicking even after this six-pronged attack.

I here end this discussion of the context principle and, based on the results, regard this principle as a viable and well-motivated thesis which successfully underwrites, in general, reference to, and object-directed thought about, abstract objects. The following chapters will be concerned with the challenge of meeting Benacerraf's second *epistemic* constraint: To explain how we can know or even claim to know truths of mathematics (see section 1.1). I previously argued that an answer to this question can't be made by merely appealing to the context principle, but it lies within the theory of abstraction principles or implicit definitions in general. Crucially the issue here concerns the notion of stipulation and how to provide an explanation of how a subject grasping a certain mathematical statement can arrive at knowledge of the statement.[42] According to the *Fregean platonist*, no additional faculty or any type of pragmatic considerations have to be appealed to in order to know a mathematical statement – a daunting claim that I will discuss, and defend, in the following chapters.

[41] Remember it is a *feature* and not a problem that it justifies reference to abstract objects!

[42] This will also provide further insights by which to explain away Dummett's tension as outlined in 2.3.3.

Chapter 3

A framework for implicit definitions

Introduction

In the previous chapter I defended a version of the context principle against various arguments and showed that it is compatible with a form of platonism. Also, I argued that some criticisms of the context principle are better understood as criticism or scepticism about implicit definitions and the notion of stipulation involved therein.

In this chapter, I will focus on the idea of implicit definitions and the notion of stipulation. I regard abstraction principles as a special case of implicit definitions. In the following, I will first introduce the idea of an explicit definition and then, in contrast, outline various features of implicit definitions and abstraction principles. Here, I will also explain what role the notion of a stipulation plays.

Having distinguished these two forms of definition, I will distinguish three aims, or *dimensions of achievement* for implicit definitions which have been confused in current literature. Following this explanation, I will state and assess various necessary conditions one should impose on an implicit definition. The conditions that have to be fulfilled depend upon what the achievement of the definition is intended to be.

Having provided this framework for implicit definitions, I will finally outline the so-called *traditional connection*: the view that implicit definitions can found a priori knowledge of first principles of mathematics and logic. To this end, I will explain a further epistemic dimension for implicit definitions which, according to the proponent of the traditional connection, will underwrite our knowledge of logic and mathematics.

The traditional connection is upheld by the Neo-Fregean programme in order to provide a fuller explanation of a subject's knowledge of mathematical theories in virtue of abstraction principles. The details how the traditional connection can be justified and what use it is for the Neo-Fregeans, or a Fregean platonist in general, will be assessed in the following chapters. I will end this chapter by distinguishing various challenges to the traditional connection which will structure the subsequent debate.

3.1 Explicit versus Implicit Definitions

Standardly, there are two kinds of definitions which introduce new terms into a language: Explicit definitions on the one hand, which are usually taken to be unproblematic and, on the other, contextual or implicit definitions, which are, by some, taken to be somehow inferior to explicit definitions. They are considered inferior or are even rejected as genuine definitions, because many philosophers and mathematicians consider them to be either insufficient to provide a determinate meaning of a term or questionable because the notion of stipulation involved in an implicit definition is suspect. I will start by outlining the standard features of explicit definitions and then turn to various features of implicit definitions.

3.1.1 Explicit definitions – some general features

In mathematical logic text books, explicit definitions are normally regarded as mere abbreviations of already understood terms or sentences (depending on what semantic category of expression is being defined) in that they fulfil two conditions: they are eliminable and they are non-creative. *Eliminability* in this context means that on the basis of the definitions, the newly defined symbol is avoidable in any formula of the theory.[1] More precisely following Suppes, this condition can be characterised as follows:

> *Eliminability*
>
> A formula S introducing a new symbol of a theory satisfies the criterion of eliminability if and only if: whenever S_1 is a formula in which the new symbol occurs, then there is a formula S_2 in which the new symbol does not occur such that $S \rightarrow (S_1 \leftrightarrow S_2)$ is derivable from the axioms and preceding definitions of the theory. (Suppes, 1957), p.154

So, a definition which fulfils this criterion is such that – theoretically – we could do without the definition and therefore without the new term, without compromise of the expressive power of the theory.

The condition of *non-creativity* on the other hand provides a guarantee that even if the defined term is eliminable, once the new term is introduced, no new theorems in which the new term does not occur, are provable from adding the new term involved. More formally (and precisely) – again following Suppes[2]:

[1] Eliminability is what Quine took to be the main feature of a definition, compare his statement: "to define a term is to know how to avoid it" (Quine, 1940), p.47).

[2] This constraint is similar to Field's conservativeness condition (compare section 1.2.3). Fields notion however is concerned with mathematical theories and not a specific formula. In section 3.5.2, we will come across a further notion of conservativeness which differs crucially from Field's conception and this understanding of non-creativity.

Non-creativity

> A formula S introducing a new symbol to a theory satisfies the criterion of non-creativity if and only if: there is no formula T in which the new symbol does not occur such that $S \rightarrow T$ is derivable from the axioms and preceding definitions of the theory but T is not so derivable. (Suppes, 1957), p.154

We can take these two constraints as necessary and sufficient conditions for acceptable explicit definitions. A formal characterisation of explicit definitions can be given in the following way:[3]

> α is explicitly definable in terms of $\beta_1, \ldots, \beta_n$ in T if a definition of α from $\beta_1, \ldots, \beta_n$ is one of the sentences of T.

While what it is to have a definition of α in terms of $\beta_1, \ldots, \beta_n$ is to have the following statement:

$$\forall x_1, \ldots \forall x_k(- - \alpha - - \leftrightarrow -\beta_1, \ldots, \beta_n-),$$

where all non-logical symbols occurring in $-\beta_1, \ldots, \beta_n-$ belong to $\{\beta_1, \ldots, \beta_n\}$; all variables occurring free in $-\beta_1, \ldots, \beta_n-$ belong to $\{x_1, \ldots, x_k\}$. The form of $- - \alpha - -$ depends on what is defined. That is either it is the formula $x_1 = \alpha$ if α is a name; or just a sentence α, in which case k= 0 and the definition is a biconditional. The relevant adjustments will have to be made for predicates and functions.

This purely formal characterisation also entails a grammatical constraint, namely that the definiens has to provide a *semantically equivalent* expression for the definiendum. It should therefore be substitutable in all occurrences without changing either the truth-value or the grammaticalness of the sentence. Generally, such explicit definitions are not stated

[3] I follow Boolos & Jeffrey's characterisation – see (Boolos and Jeffrey, 1989).

in the object-language, but only in the meta-language – so conceived "$=_{def}$" or "$\leftrightarrow_{def}$" is a metalinguistic symbol, and therefore the definiens merely provides a fruitful abbreviation which is available within the meta-language. So, in the light of these constraints and features of explicit definitions, it might seem odd, or at least unhelpful, to talk about the truth of an explicit definition. Rather, as for example, Frege has noted, the "correctness" of such definitions is governed by their *fruitfulness*, i.e. whether the abbreviation they provide is useful or makes reasoning more surveyable.

3.1.2 Implicit definitions – some general features

In contrast to explicit definition, the characterisation of implicit definitions and the way they are used varies in the literature. In order to avoid confusion, I will first introduce a purely formal characterisation of implicit definitions and contrast them with explicit definitions and then later introduce a more informal and broader conception of implicit definitions.

Standardly, the formal characteristic of implicit definitions is just the following requirement:[4]

> α is implictly definable from $\beta_1, \ldots, \beta_n$ in T, if any two models of T with the same domain which agree in what they assign to $\beta_1, \ldots, \beta_n$ also agree in what they assign to α.

Thus, roughly the idea is that by having the interpretation for all terms but the new term fixed, the relations among these concepts and the newly introduced concept suffice to fix the new term's interpretation. For a while, this type of definition was regarded as genuinely distinct from explicit definition in the sense that if α is implicitly definable, it does not hold (in general) that α is also explicitly definable. Surprisingly, however, *Beth's theorem*

[4] I am here following (Boolos and Jeffrey, 1989)

shows that in first-order languages at least, every term which is implicitly definable in the above sense, is also explicitly definable (the converse is less surprising).[5] As such there is no significant difference between the two types of definitions, at least within first-order logic.

Despite this result for first-order languages, the formal characterisation of implicit definition makes for a different type of definitions, once we are within higher-order languages. Here, the Beth-definability theorem does not hold in general.[6]

As an example, take the definition of the expression "the number of" in virtue of Hume's Principle and second-order logic. It can be shown that it uniquely characterises numbers (on a domain of a particular size at least and leaving aside any Caesar-type worries for the moment that might undermine uniqueness) and fulfils the formal constraint on implicit definitions. Nevertheless, the definiendum is not *eliminable* and thus the definition can't be regarded as an explicit definition.

A possibly more contentious case that is nevertheless worth noting, is the implicit definition of the membership-symbol '$\in$' in virtue of its role in the second-order set-theoretic axioms (ZF). On the assumption that there is exactly one inaccessible (and no urelemente) one can provide a categorical characterisation of ZF and so implicitly define '$\in$' within that system, but there again there is no explicit definition of '$\in$' available.[7]

[5]See (Boolos and Jeffrey, 1989), chapter 24

[6]It fails in cases where in a second-order language we allow statements to have higher-order terminology, i.e. the non-logical terminology is of higher-order than the variables. So, for example Hume's Principle (within second-order logic) which involves an abstraction operator and so a function from properties to objects is one such case. For a detailed account on Beth's definability properties in second-order language, see (Shapiro, 1991) p. 164ff.

[7]Confusingly, (Antonelli, 1998) provides an example of an implicit definition by appealing to the recursive definitions of the symbol '+'. This definition is given in virtue of the two equations: (1) $x + 0 = x$ and (2) $x + y' = (x + y)'$. Antonelli notes that this definition of '+' is unable to eliminate just this symbol in the following occurence: $\forall x \forall y(x + y = y + x)$. This is confusing, because in a first-order language this can't be a genuine implicit definition (due to Beth's theorem and the lack of eliminability) and in second-order languages this definition is eliminable and as such can be rendered as an explicit definition. See (Shapiro, 1991). Primitive recursive definitions are a special case of definitions that go beyond pure first-order logic and should better not be characterised as an implicit definitions. I will thus postpone a more thorough discussion of recursive definitions.

So, to summarise, only once we adopt second-order logic as the background framework is there a genuine formal difference (that extensionally-speaking makes a difference) between explicit and implicit definitions. Here, one of the main differences – as can be deduced from the above examples – is that implicitly defined terms are not always *eliminable*, in contrast to expressions introduced by explicit definitions.

One important issue, next to the condition of eliminability, as outlined above, is the condition of non-creativity. In what respect and whether an implicit definition can be successful despite failing the non-creativity constraint, will be discussed in more detail in section 3.5.2.

In the following, I want to consider a broader notion of implicit definition or, as it is sometimes called *contextual* definition, which is not necessarily confined to formal language and its formal characterisation. The feature of this broader notion is that the content of the implicit definition is more flexible than that of an explicit definition. In an explicit definition we define a single occurrence of an expression by connecting it to an expression of the same semantic category (for example consider the following definition: brother of x $\leftrightarrow$ male sibling of x). The way I will conceive of implicit definitions in the broader sense (and this, I think, is the notion often appealed to in recent literature) is the idea that a definiendum can acquire a meaning by merely fixing a certain context for the use of the new expression by *stipulating* sentences in which the terms occur as true. As we will see below the meaning-fixing context can take on different forms, and I will distinguish four models of how an expression can be implicitly defined in this broader sense.[8] As a result of this characterisation of implicit or rather contextual definitions, the definiendum of an implicit definition is usually *not* of the same semantic category as the definiens.

[8]A definite description of the from "The ϕ is ψ" is the conjunction of the three sentences: (1) There is at least one ϕ (2) There is at most one ϕ, and (3) Whatever is a ϕ is a ψ. In this context it can be regarded as an implicit definition in the broader sense. Namely, in this case there is no explicit definition of an expression, rather the truth-conditions of sentences in which ϕ and ψ occur are specified in virtue of other sentences.

This feature of implicit definitions points towards a further and important epistemic/ semantic feature of such definitions. Namely, implicit definitions are – at least sometimes – directed towards introducing a *new concept* and not merely a new term by abbreviating a known expression or concept (as in the case of explicit definitions). In this sense, implicit definitions are often regarded as "meaning-constitutive" as they explain how one can come to acquire a new concept in virtue of understanding a certain context or sentence in which the new term figures. As an example, often Hume's Principle is regarded as "meaning-constitutive" of the notion "(cardinal) natural number", or alternatively one might argue that the Peano axioms can be considered as "meaning-constitutive" of the notion "(ordinal) natural number". Equally, although again slightly more contentiously, the axioms of ZF can be regarded as meaning-constitutive of the notion of "set". This is in contrast, again, to an explicit definition where a mere abbreviation of *already understood* expressions obviously cannot be meaning-constitutive of a new concept.[9]

Before I outline the four different models of how implicit definitions can fix the meaning of a term in a context, the notion of stipulation should be briefly elucidated. First of all, all definitions have in common an element of stipulation. Even in the case of explicit definitions we stipulate that a certain expression is to be used as an abbreviation for another. In the case of implicit definitions, the element of stipulating the *truth* of a sentence or sentences plays an important role in fixing the meaning of a definiendum. The very basic idea is that by stipulating that a sentence in which the new term figures is to be true, the sentence thereby fixes a meaning for the term, since the latter will acquire the meaning required for this sentence to be true.[10] The scope of the stipulation can either be simple sentences,

[9]Here I won't discuss under what conditions – exactly – new concepts are introduced but I hope the general idea of a contrast between introducing a new concept through an implicit definitions versus introducing a new (abreviatory) expression for an already understood expression is clear enough.

[10]I will slightly modify this idea of stipulation later – see section 3.2 and discuss and even further modify it in the light of a general problem for this conception of stipulation, in chapter 5. However for present

conjunctions of sentences, special types of sentences for instance biconditionals or rules. In the following, I will outline four models.

The most general model is the following:

Model for implicit definitions by stipulation of truth

"$\ldots\xi\ldots$" is true

A common example in the current literature[11] is the implicit definition of "Jack the Ripper" through the sentence: "Jack the Ripper committed such and such murders in the 19th century in London." The semantic effect of this definition is not to provide a mere synonym by providing the name "Jack the Ripper", rather the aim is that "Jack the Ripper" is to rigidly stand for the object that committed these murders.[12]

A second type of stipulation is one where instead of stipulating a sentence to be true simpliciter, the truth-condition of a type of sentence in which the new term figures is fixed by connecting it to the truth-conditions of another already understood sentence or collection of sentences. Thus, the more general model for this type of implicit definition, which is a special case of the above *model for implicit definitions by stipulation of truth*, might be this:

Model for implicit definitions by stipulation of truth-conditions

"$\ldots\xi\ldots \leftrightarrow A$" is true

where 'A' is previously understood. One example of this model is Russell's account of definite descriptions where the truth-conditions for the sentence "The ϕ is ψ" are spelled out in virtue of the conjunction of the three sentences (1) There is at least one ϕ (2) There is at most one ϕ, and (3) Whatever is a ϕ is a ψ. Another common example of this type of stipulation

purposes this rough idea of stipulation suffices.

[11]See for example (Hale and Wright, 2000)

[12]So the idea is a type of reference-fixing model for names.

are abstraction principles: here truth-conditions for identity statements containing the new expression are connected with the holding of an equivalence-relation on items of a previously understood kind. Thus, there are two occurrences of the newly introduced expression on the left-hand side flanking an identity-sign; and by so fixing the truth-conditions of such identity statements we introduce – according to the proponents of abstraction principles a new concept.[13]

Abstraction Principles

"$\forall\alpha\forall\beta(\xi(\alpha) = \xi(\beta) \leftrightarrow \alpha \approx \beta)$" is true[14]

If α and β stand for singular terms, the resulting abstraction principle is first-order as in the case of Direction Abstraction. If they stand for first-order predicates the resulting abstraction principle is second-order as in the case of Hume's Principle.

Contrasting to the above case where only one new type of term is introduced by providing the truth-conditions for a certain context in which it is used, there are what might be called Hilbert-style or *structural* definitions. Here, a collection of sentences in which various new expressions figure and different relations between these terms are supposed to hold, are stipulated. So, for example in his "Grundlagen der Geometrie"[15] David Hilbert aimed to define the terms "point", "line" and "plane" in virtue of how they are related to each other

[13] Historical note: Frege's view of definitions (in the *Grundgesetze* (Frege, 1903) §33) is that definitions have to be simple. This, I believe, entails that there must only be one occurrence of the definiendum in the definition. In the light of his acceptance of Basic Law V, where there are two occurrences of the newly introduced term-forming operator "the extension/value-range of ", one is forced to conclude that Frege did not regard Basic Law V as a definition. However, in the following, I disregard Frege's contention and I will continue to regard Basic Law V as an effective implicit definition.

[14] ξ is a term-forming operator applicable to expression of the type of α and β and $\approx$ is an equivalence relation on entities denoted by expressions of that type.

[15] See (Hilbert, 1899). Hilbert's idea of implicit definition was fiercely rejected by Frege. See (Frege, 1976) for the Frege-Hilbert controversy which due to a lack of space won't be discussed here. But various issues raised therein will come up in later discussions.

within different sentences (or a context).[16] Thus, it seems, a general model like the following captures Hilbert's idea:

Model of structural definition

"$\ldots \xi R \pi \ldots \ \& \ldots \pi S \tau \ldots \ \& \ldots \tau T \phi \ldots$ " are true

where either terms such as ξ, π, τ and ϕ or predicates (R, S or T) are introduced as new expressions.

A final model of implicit definitions, prima facie contrasting with all the above, is one in which instead of stipulating the truth of sentences, we put forward rules. One rule introduces the new term – the Introduction rule, and one which shows how to eliminate the term – Elimination rule. The initial idea – due to Gentzen – was that "the introduction represents, as it were, the 'definitions' of the symbol concerned, and the elimination are no more, in the final analysis, than the consequences of these definitions."[17] So, in this case we have the following model:

Introduction and Elimination Rule

$$\frac{\Gamma \vdash A}{\Delta \vdash \ldots \xi \ldots} \text{ Introduction Rule } \quad \frac{\Delta \vdash \ldots \xi \ldots}{\Gamma \vdash A} \text{ Elimination Rule}^{18}$$

What all these models have in common is that – standardly regarded – the initial definitions are put forth in the metalanguage and it is only by disquoting on the definition, provided it is successful, that a new term has been introduced into the object-language.[19] The difference

[16]Hilbert did not stipulate the truth of the sentence, rather claimed that if they are consistent stipulations, then they are also true.

[17](Gentzen, 1969), p.80

[18]Examples such as conjunction are obvious and are later discussed in chapter 5. I take '$\Gamma \vdash$' to mean that there is a proof from Γ that leads to A. More specific examples will follow below.

[19]I will come back to this assumption in chapter 5 and 6.

in the latter case of rules in contrast to sentences, is while sentences are stipulated to be true, the rules are stipulated to be valid.[20] Lastly, it is often claimed that I-E rules are very suitable to introduce "new concepts" – as we noted, a feature specific to implicit definitions – since they determine the role a concept plays in reasoning, which can be considered sufficient for concept-determination and grasp of which constitutes concept-possession.[21]

Before turning to a discussion of the various constraints that should be imposed and the various aims implicit definitions can fulfil, I want to note some relationships between these four models for implicit definition. Above I remarked that the most general model for implicit definitions is the first model which involves a stipulation of truth and that the second and third are merely special cases of the former (the fourth case involves the stipulation of validity rather than truth, but if we generalise and talk of stipulations of truth/validity of sentences/rules then the first and fourth both fall under this generalisation (just as the second and third)). Interestingly, however, one could assign priority to the stipulation of truth-conditions and consequently modify the first model. For example one might consider the laying down of a certain axiom not as a simple stipulation of its truth but invoking a tacit biconditional with a necessary truth figuring on one side and the axiom on the other; in this way, one could always view stipulations of truth as stipulations of truth-conditions. In addition, one might – in certain circumstances – also assimilate the model of stipulation of truth-conditions to the idea of introduction and elimination rules (and even vice versa). That is, one could take the one side of the biconditional in which the new term does not figure as the "top" part of the introduction rule, which introduces us to a new term, and then use the other direction as the elimination rule. Equally, one could turn the introduction

[20]As (Prior, 1961) showed various constraints have to put on such rules. I will turn to this in the next section.

[21]Gentzen as well as (Dummett, 1991b) defended this claim, but imposed in addition the constraint of harmony on the rules, in order to rule out problematic cases and guarantee that such rules fix a concept. I will come back to this issue on p.128.

and elimination rules into the statement of the rule and so involving a biconditional which is then stipulated to be true.[22]

Hence, there seems to be an interesting link between a truth-conditional account of meaning appealed to in the biconditional account and the conceptual role semantics which is assumed by using Introduction and Elimination rules. Here, I don't want to commit myself to any special conception or priority of one model over another, but rather merely note these interrelationships.[23]

Let us now turn to explore further features of the theory of implicit definition and, in particular, under what conditions an implicit definition is successful. I noted that implicit definitions can be regarded as involving stipulations of the truth of the sentence (or sentences).[24] What is needed, therefore, is an account of the circumstances under which we are justified in making such stipulations and justified in regarding stipulations to have succeeded.

It seems that an implicit definition can simply fail, for example in the "Jack the Ripper"-case, if there was no such murderer, or if there had been a number of people committing these crimes. Furthermore, we need to ask: what happens if the stipulation fails? Does this entail that the definition also fails to fix a meaning or introduce a new concept? In the following, I will distinguish three aims, or dimensions of achievements for an implicit definitions and argue that these issues need to be dealt with separately.

[22]This works in the case Hume's Principle or other abstraction principles, since they can be regarded either as a biconditional or a pair of rules. It does not work in general however. Consider for example the I-E rules for the conditional or disjunction that can't be parsed (at least in no obvious way) into a biconditional.

[23]In their recent paper (Hale and Wright, 2000) prefer the rule formulation of Hume's Principle, exactly why this is of advantage however is not made clear.

[24]For simplicity I will here merely talk of the truth of a sentence. The same applies to the stipulations of the validity of a rule. I will use this simplification throughout the rest of the chapter.

3.2 Three dimensions of achievements for implicit definitions

The "standard view"[25] of implicit definitions ties together the success of an implicit definition with the definiendum's acquiring meaning. The idea is that an implicit definition endows the new term with a meaning, *in virtue of the fact that the term will mean whatever it takes to make the sentence, in which it figures, true.* So considering the different models for implicit definitions above, therein 'ξ' will mean whatever makes the sentence true.

However, the term "mean" is inherently ambiguous and we need in particular – although not uncontroversially – to differentiate between implicit definitions that are merely meant to fix the reference of a term, i.e. where "meaning" is understood as the reference (Bedeutung) of terms, and cases where a new term is implicitly defined by endowing it with a sense. So, on the latter view, the definition either introduces a new concept or furthers our understanding of an already accepted but previously unclear concept.[26]

So, with this background we can differentiate various aims an implicit definition can have. What seems the most important aim is what I call *success* – success involves the *truth* of the sentence (or *validity* of the rule) and, in addition, a successful implicit definition not only fixes the reference of the term but also imparts an understanding or fixes the sense of the new term. Hence, we can say the following:

Success

An implicit definition is *successful*, if and only if the sentence/rule, in virtue of which the term is introduced is both true/valid and imparts an understanding of the new term.

[25]See (Horwich, 1997) and (Hale and Wright, 2000) for such characterisation

[26]This is an idea that underlies Carnap's idea of "explication" – more on this below.

However, one might wonder whether in order to introduce a new concept or – less controversially – to impart an understanding of the new term, the statement (or statements) that provide the vehicle of the implicit definition *has* to be true. The "standard view" entails that it is in virtue of the statements' being true that the term gets its meaning. If 'meaning' here is understood as referent and the idea is to fix a referent, then it seems clear that the statement has to be true. Nonetheless, I want to suggest that, even if it fails to be true, an implicit definition can be *effective* in providing an understanding of the term and so might help to at least partially determine a concept. Hence, I believe that the standard view – taking on a suggestion by Horwich[27] – isn't unqualifiedly correct. Rather I think the model for implicit definitions should be understood such that an implicit definition endows the new term with a meaning, in virtue of thinkers' *regarding* the sentence, in which it figures, *as true*. This allows that even false implicit definition can provide an understanding of a new term and, in section 3.3, I will argue that certain implicit definitions – such as Basic Law V – would not be done justice, if the aim of *success* embedded in the standard view of implicit definition is the only candidate.

Accordingly, I want to distinguish from the aim of *success* a different achievement an implicit definition can have, which is merely concerned with what I will call the *effectiveness* of implicit definitions. Being an *effective* implicit definition can be characterised as follow:

> *Effectiveness*
>
> An implicit definition can be regarded as *effective*, just in case the sentence/rule in which the term to be introduced figures, endows the term with sense and thereby imparts an understanding of the term.

Having distinguished the two achievements, we need to further develop and motivate

[27](Horwich, 1997)

them in the next two sections. Nonetheless, we can already note an obvious asymmetry, namely, that while a definition can't be successful and not be effective, it can be effective but fail to be successful. As a result, there seems to be logical space for a third achievement of implicit definition, which involves the truth of the sentence (or sentences) without fulfilling the constraint of effectiveness.

Let us call an implicit definition *productive* in the case where the statement is true but does not yet provide the term with a (definite) understanding.

> *Productiveness*
>
> An implicit definition can be regarded as *productive*, just in case the sentence in which the term to be introduced figures is true, without imparting the new term with a (definite) understanding.

This will surely seem odd. How can a sentence/rule be true/valid while the new term fails to be properly defined and so fails to be properly understood? One example is the following case of the connective "tunk" which uses the following I-E rules:[28]

$$\frac{P,\ Q}{P\ tunk\ Q}\ \text{I-Rule} \qquad \frac{P\ tunk\ Q,\ P \vdash S,\ Q \vdash S}{S}\ \text{E-Rule}$$

This connective uses the introduction rule of conjunction and the elimination rule of disjunction. These rules are valid, but arguably, the connective "tunk" fails to acquire a sense. The pair of rules are somehow 'not in harmony' and thus they fail to establish a genuine connective, despite the validity of this pair of introduction and elimination rules.[29] The aim of productiveness, however, won't be very important for the subsequent discussion

[28] The idea of "tunk" – which uses the inverse of the "tonk" rules (see next section) – is due to Crispin Wright.

[29] In the next section I will say a little bit more about the notion of harmony which seems necessary for an implicit definition to be effective.

and so I won't further motivate or defend this third achievement of implicit definitions.

Instead, I will focus on the following issues: I will explain and further motivate the criteria of *effectiveness* and *success*. The aim of *success* is widely discussed in contemporary literature[30] and it will be reviewed in the next but one section. First, I will offer a more detailed discussion of the notion of *effectiveness*. Later, I will explain how the traditional connection uses these different achievements of an implicit definition by assigning an implicit definition an additional epistemic dimension, in order to explain and found knowledge of arithmetic and logic. This crucial additional component will be further discussed in section 3.7.

3.3 Conditions for Effectiveness

In this section I won't be able to offer a clear-cut criterion for effectiveness, instead I will suggest various ways this notion can be elucidated. Then, I will offer some motivation as to why we should adopt the aim of effectiveness and thus propose a revised version of the standard view of implicit definitions which does not entail that the definition actually has to be true in order to introduce us to a concept. Previously, I suggested the following characterisation of the aim of effectiveness.

> *Effectiveness*
>
> An implicit definition can be regarded as *effective*, just in case the sentence/rule in which the term to be introduced figures, endows the term with sense and thereby imparts an understanding of the term.

In order to be completely clear what this aim amounts to we would need to decide upon a general account of what it is to possess a concept or for an expression to have sense

[30] See (Boghossian, 1997) , (Horwich, 1997), (Horwich, 2000) and (Hale and Wright, 2000)

which, surely, I won't be able to settle here! Nevertheless, there are broadly two options for understanding the above characterisation. On the one hand one could hold that an implicit definition is effective only if it is concept-constitutive, i.e. if it introduces a new concept. Consequently, one might adopt various views as to what constitutes a concept, for example whether providing a certain *pattern of use* is sufficient.[31] On the other hand, one could hold the view that an implicit definition is effective if it provides an understanding of a certain concept without necessarily being itself concept-constitutive. Hence, if the implicit definition provides a good grasp, or proper understanding, of the concept involved then the definition is effective. What exactly counts as providing a good grasp or proper understanding of a term is notoriously hard to answer. A minimal constraint should be Evan's *generality constraint.*[32] The idea roughly is that in order to have a good grasp of a term (say a singular term) t in a sentence like 't is F', one should be able to understand its occurrence in a wide range of other (appropriate) contexts, such as 't is G', etc. (*mutatis mutandis* for predicates). Let us regard this last requirement as a *minimal* condition for effectiveness of an implicit definition.

The aim of *effectiveness* is also in the vicinity of Carnap's conception of definitions as "explications". Here, the aim is to clarify and specify a previously indeterminate concept, and so either to provide a better grasp of an already accepted concept or to introduce a new concept on the basis of a hitherto indeterminate understanding of a term.[33] Interestingly, Carnap himself regards Frege's Basic Law V as a prototype of such explications and thus seems to hold that, despite its inconsistency, it is effective in providing a better understanding of the antecedent concept of extension (or value range). In a similar spirit, by showing that there are cases of unsuccessful but effective definitions, I will hope to establish the achievement

[31]See (Horwich, 1998) for a defence of this view.

[32](Evans, 1982), p.100ff.

[33]See (Carnap, 1967) chapter 1, for a characterisation of such explications. To note, even for Carnap (and (Hempel, 1952)) sometimes explications are understood as actually introducing a new concept based on an old one and sometimes merely as specifications, one that does not introduce a "new" concept.

of effectiveness as a genuine achievement for implicit definition and thereby argue that the standard view of implicit definitions, which requires the *truth* of the stipulated sentences is not correct. In order to underwrite this view, I will offer two general considerations.

The first consideration is mainly based on the intuition that there are cases, such as Basic Law V or the unrestricted T-schema ("S" is true iff S), which are inconsistent as definitions, but still at least prima facie impart an understanding of the defined term. If the achievements of implicit definition were confined to the *success* of an implicit definition, this intuition could not be accounted for. A definition like Basic Law V (and others) would have to be regarded as unsuccessful, as it fails to be true, and so could not endow the target term with meaning. However, a general account of implicit definitions should account for the above intuition concerning inconsistent definitions and subsequently recognise the project of effectiveness in order to accommodate inconsistent definitions. Let me try to substantiate this intuition with the following points – I will focus for simplicity on Basic Law V.

First to be noted is the fact that Basic Law V is still used today, at least as a *heuristic device*, for the notion of *extension* or *set*. This observation might be vindicated by looking at various introductions to set-theory where Basic Law V is used to introduce the reader to the notion of set. It indicates that some understanding of this notion is provided by the inconsistent "axiom". Accordingly, it can be regarded as assisting in a grasp of the term "set" and a general account of implicit definitions should account for the fact that inconsistent axioms can still contribute to imparting an understanding of a concept.

Supporting this view of Basic Law V as providing an understanding of the term "extension" is that this principle provides the basis for further *improved* versions of the notion of "extension" as introduced by it. Consider for example recent work of George Boolos and the idea of founding set theory on what he calls "New V".[34] I am tempted to think that this principle (New V) tries to render more precisely the intuitive idea behind the term "extension" as first given in Basic Law V, while avoiding the inconsistent instantiations of the axiom. In any case, from this perspective, the idea that Basic Law V is "improvable" points to a previous understanding this axiom confers.

A further consideration to underline this type of motivation is to engage in the thought-experiment that ZF is inconsistent. It is epistemically possible to think that ZF is inconsistent. In that case, would this also result in the system's never having provided a meaning for the term "set", that is, would we then agree that we have used meaningless symbols for the last hundred years in set theory? It seems not, and by appealing to the further weaker category of effectiveness, we can account for this, I presume, ubiquitous intuition that we have a good understanding of the notion "set" through the axioms of ZF (albeit their at least epistemically possible inconsistency). To be sure, this reasoning has its limitations in that strictly speaking it merely shows that it is epistemically possible to think that ZF is effective on the assumption that it is epistemically possible to think that it is inconsistent.[35] Nevertheless, I hope it encourages acceptance of the category of effectiveness as independent of success.

To summarise, this first line of argument aims to bolster the idea that Basic Law V or some other inconsistent principles do provide some understanding of the terms they introduce and that this kind of achievement needs to be accounted for.[36]

[34] See for example (Boolos, 1989), p.99 New V is a principle resembling Basic Law V, but which abandons the requirement that each distinct concept gets a unique extension – this requirement only holds for concepts that are not 'too big' – thereby avoiding the inconsistency.

[35] Also provided we have closure of the epistemic possibility operator.

[36] It should be noted that Basic Law V is not the only such case. Similar motivation can be given for the unrestricted T-schema, which gives rise to Liar sentences. Also consider in analogy the game of Croquet. Its rules or better "laws" are apparently inconsistent. Nevertheless, it is still played and its "laws" are still considered as effectively providing rules for a game. Also, we would say that we have a grasp of the concept "Croquet" which presumably is connected to the inconsistent rules. For this observation see Appendix B of

The second consideration will contrast two inconsistent definitions, which, if we only have the criterion of success, are on the same level (namely both are unsuccessful). Nevertheless, I will try to show that they differ in that the former at least seems to provide an understanding (Basic Law V), while the other is completely useless.

Compare Basic Law V with a different also inconsistent implicit definition that of the connective "tonk". "Tonk" is characterized by the following pair of introduction and elimination rules:

$$\frac{A}{A\ tonk\ B}\ \text{Introduction Rule} \quad \frac{A\ tonk\ B}{B}\ \text{Elimination Rule}$$

Tonk, as is evident, is a combination of the standard disjunction introduction rule and the conjunction elimination rule.[37]

If we were to accept this pair of I-E rules, we would, on the basis of its inferential resources provided, be justified in holding true any claim possible. From a statement A we would be justified in inferring any statement, including not A. Hence, the tonk-rule is (blatantly) inconsistent.

Now, if the truth or validity of the sentence/rules stipulated is a necessary condition to "endow" the term with a meaning, then it would seem that no distinction can be drawn between the case of Basic Law V and the tonk-case, since their acceptance leads in both cases to an inconsistency, and thus both alike have to be regarded as merely unsuccessful implicit definitions. However, the temptation remains that Basic Law V gives us *some* sort of *understanding* of the term introduced. By contrast, the connective tonk seems empty;

(Miller and Thorp, 1966) on the "Possible Inconsistency of the Laws of Croquet".

[37]It was first introduced by (Prior, 1961). Tunk, as discussed on p.122, is the inverse of tonk.

it provides no intelligible use, as it just expounds triviality and no even partly coherent or concept-constituting pattern can be recognised.

It is hard to exactly characterise what the difference here between Basic Law V and the tonk-rules is and I can only provide some rather undeveloped suggestions: First, based on the above discussion one could, at least in an informal way, say that one contrast between Basic Law V and the tonk-rule is that it does not seem possible that the connective tonk can be the basis for the further explication or improvement of what might be regarded as the underlying concept. No partial meaning is rescuable from its inferential pattern and hence we are not in a position to improve on and produce, in analogy to "New V", a notion like "New tonk", which captures or explicates the initial coherent proposal.[38]

Alternatively, one might attempt to make a further distinction between inconsistency and triviality. So, while Basic Law V is inconsistent, triviality only follows in virtue of additional logical inferences (assuming classical logic), while the tonk-rules are immediately trivial. This 'inbuilt' triviality could then be blamed for its ineffectiveness.

A third and last suggestion is to appeal to a notion of harmony. The idea, as it already surfaced in our discussion of tunk (p.122), would be to require that the introduction and elimination rule have to be related in such a way that the content of the introduction rule is properly reflected (carved up?) in the elimination rule. So, in the case of Basic Law V – as a pair of rules – one is tempted to say that the introduction and elimination rule are in harmony in the sense that the top line in the introduction rule appears as the bottom line of the elimination rule and the bottom line of the introduction rule just is the basis on which the elimination rule draws its consequence. This is obviously not the case for either tonk or

[38]Recently (Cook, 2006b) showed that the tonk-rules have some usage in a non-transitive logic, where tonk can be understood as explicating a four-valued connective. However, I find it hard to believe that this is an *improvement* on the initial pattern. It just seems as far fetched as thinking of "New tonk" in terms of self-identity!!

tunk.[39]

Here, I will have to leave various issues open – for example how to exactly draw a line between effective and non-effective principles among inconsistent and among consistent ones. Still, I believe (and have argued for the claim) that the case of Basic Law V shows that despite being inconsistent, it should be considered an example of the aim of *effectiveness*, while the tonk-rules on the other hand, fail to be effective, bestowing no understanding of the operator involved. These considerations, I hope, provide some motivations for regarding the notion of effectiveness as useful and non-empty. Moreover, it shows that the standard view for implicit definitions fails since certain implicit definitions need not be true to convey a meaning. The idea of effectiveness and the revised version of the standard view will figure prominently in section 3.7. So, let us now turn to the achievement of *success*.

3.4 Conditions for Success

Whether an implicit definition is successful or not depends on its effectiveness and the truth of the 'vehicle' of the implicit definition. Here, I want to review various constraints that have to be fulfilled for an implicit definition to be true and so successful (assuming it is effective).

There are at least two constraints that come to mind and which can be easily motivated by means of the "Jack the Ripper" – definition, or any definite description for that matter. The first is that the sentence has to be *satisfiable*, i.e. there has to be someone who committed all those crimes. The second is that there should be only one satisfier, i.e. the constraint of

[39]This very rough idea is unfortunately too strict since other presumably effective rules such as the I-E rules for conditional or disjunction do not fulfil this rough characterisation of harmony. However, I won't be able to pursue further discussion of a stricter criterion for the notion of harmony. Note though, that the idea of harmony hinted towards above is distinct from harmony as understood by others in terms of non-creativity or normalizability (Basic Law V is surely not harmonious on this reading and it remains to be seen why such notion of harmony should have anything to do with the effectiveness of a definition, although that is the way it is usually motivated!). (Read, 2000) has recently argued that the term is inherently ambiguous in contemporary literature (Read focuses mainly on (Dummett, 1991b)).

uniqueness.

The first constraint of satisfiability can be explained rigorously by appeal to so-called *Ramsey* sentences.[40] The Ramsey-sentence of an implicit definition can be obtained by substituting the introduced term by a variable (of the appropriate type) bound by an existential quantifier. So to take a simple example: If the stipulation of "$\ldots\xi\ldots$" is meant to fix the meaning of "ξ" embedded in the sentence-frame "$\ldots\_\ldots$", the Ramsey sentence is simply "$\exists x \ldots x \ldots$".

The idea of a Ramsey sentence is to encapsulate the ontological commitments of a theory. So, in case the term "ξ" is a singular term, the Ramsey sentence will involve a first-order quantifier and thus highlight the fact that the truth of the implicit definitions will involve the existence of (at least) one object. Similarly, the notion of satisfiability also applies to predicates and other kinds of expressions and thus higher-order constructions might be needed to construct the corresponding Ramsey-sentences.[41] Lastly, also note that this constraint is rather harmless as supposedly – *pace* various free logical systems – the corresponding Ramsey-sentence follows using basic logical operations from the truth of the initial definition.

Belnap[42], Horwich[43] and others have suggested *uniqueness* as an additional constraint on the success of implicit definitions. This demands that the corresponding Ramsey sentence is *uniquely* satisfied. Intuitively, this additional constraint has to be in place when we are concerned with *singular* terms, and the initial formal characterisation of implicit definitions – which demands that we uniquely specify the interpretation of the new term (see 3.1.2) –

[40]See (Ramsey, 1929). I will discuss Ramsey sentences in detail in 4.2.1

[41]The objectual commitments of an implicit definitions involving higher-order constructions will be less obvious. Compare, for example the Ramsey-sentence of Hume's Principle. More on this in chapter 4, especially section 4.5.

[42](Belnap, 1962)

[43](Horwich, 1997)

entails the uniqueness condition. However, as I aim to extend the idea of implicit definitions also to non-formal languages, resulting in definitions of predicates of natural language, the uniqueness condition quickly leads into trouble once we are concerned with vague and (if distinct) indeterminate predicates, which I won't be able to discuss. Therefore, for the sake of the present discussion, I will leave aside the issue of uniqueness and focus on existence.

3.5 Further constraints on implicit definitions

Here I will briefly discuss two further constraints on implicit definitions that have been proposed in the literature: That implicit definitions should be *predicative* and that they should be *non-creative.* I will discuss both constraints in turn and clarify which achievement of implicit definition either constraint concerns.

3.5.1 Predicativity

Many philosophers and mathematicians argue that an additional constraint on definitions and especially foundational definitions is that they should be *predicative.*[44] The basic idea of the constraint is that an implicit definition is *impredicative* if the new expressions introduced refer to entities that are within the range of quantifiers in virtue of which the new expressions are characterised within the definition. To take a very salient example, consider Basic Law V:

$$\forall F \forall G(Ext : F = Ext : G \leftrightarrow \forall x(F(x) \leftrightarrow G(x)))$$

Here, the objects referred to by the introduced term "the extension of" are taken to lie within the range of the universal quantifier on the right-hand side, while at the same time

[44] In this tradition are Poincare, Russell, Chihara, Weyl, Dummett and Fine amongst many others.

it is this right-hand side which is supposed to provide the conditions for the identity of such extensions.

One main reason to impose the predicativity requirement is that the impredicativity involved in Basic Law V, for example, is a crucial factor in deriving the inconsistency.[45] But, if the quantifiers of Basic Law V were restricted to predicative versions, then the paradox is blocked. Hence, one motive for adopting predicativity as an additional constraint is tied to the presumed 'inconsistency-causing nature' of impredicative definitions. However, if that is all there is in terms of motivation for this constraint then it is questionable why it should be adopted over and above the already accepted constraint of satisfiability (and so, assuming soundness, consistency). Moreover, there seems to be a wealth of definitions that are (or, at least we have good reason to believe that they are) satisfiable/consistent yet impredicative such as Hume's Principle, the Peano axioms or even more natural language examples such

[45]So consider the following derivation of Russell's paradox: First we define the (impredicative) Russell Predicate (with $\S(Y)$ standing for Ext(Y)):

$$R(x) \leftrightarrow (\exists Y(x = \S Y \wedge \neg Y(x)))$$

The paradox can then be deduced using two crucial impredicative steps:

(1)	1	$\neg R(\S R)$	Assumption
(2)	1	$\neg(\exists Y)(\S R = \S Y \wedge \neg Y(\S R))$	(1), def. of R
(3)	1	$(\forall Y)(\S R = \S Y \rightarrow Y(\S R))$	(2)
(4)	1	$\S R = \S R \rightarrow R(\S R)$	(3) (impredicative step I)
(5)	1	$R(\S R)$	(4)
(6)		$\neg R(\S R) \rightarrow R(\S R)$	(1) - (5)
(7)		$R(\S R)$	(6)
(8)		$(\exists Y)(\S R = \S Y \wedge \neg Y(\S R))$	(7), def of R
(9)	9	$\S R = \S P \wedge \neg P(\S R)$	Assumption
(10)	9	$\S R = \S P$	(9) $\wedge$ - elimination
(11)	9	$\forall z(R(z) \leftrightarrow P(z))$	(10)
(12)	9	$R(\S R) \leftrightarrow P(\S R)$	(11) (impredicative step II)
(13)	9	$\neg R(\S R)$	(9) ($\wedge$ elimination), (12)
(14)	9	$R(\S R)$	(9) – (13)
(15)	9	$\perp$	(7), (14)

The proof relies on two impredicative steps (4) and (11) as "The tallest man in the room".[46] So, an alternative motivation has to be looked for which does not rely solely on the 'inconsistency-causing nature' of impredicativity.

(Dummett, 1991a) and (Fine, 2002) have both argued that there is a vicious circularity involved in impredicative definitions and claimed that therefore it undermines the ability to either *fix a concept* or to provide an *understanding* of the term which is to be introduced. So, for example Dummett argues:

> [Frege] had however fatally overlooked the circularity of the entire procedure: that of specifying the criterion of identity in terms of the truth of the theory, and, more generally, that of attempting simultaneously to specify the domain and the application of the primitive predicates to its elements. ((Dummett, 1991a), p.233)

The circularity Dummett is alluding to here is supposedly generated by the impredicativity involved in such definitions as in Basic Law V or Hume's Principle. How exactly? Dummett writes as follows:

> Impredicative specification of the conditions for the truth of identity-statements involving one or two such terms is *not* sufficient [...]. It fails to fix truth-conditions for all sentences containing the new terms, when these terms are formed by attaching an operator to a predicate or functional expression; and it fails to do so because of the lack of independent specification of the domain, which it attempts, but fails, to circumscribe simultaneously with its determination of the truth-conditions of sentences containing the new terms.(Dummett, 1991a), p.234

[46] One can distinguish further first-order impredicativity from second-order impredicativity (based on the quantifiers involved), but I won't discuss this possible line of reply here, since it seems to me doubtful that it would lead to a more stable position.

This type of worry, based on the impredicativity of definitions is completely different in character to the first motivation above. The point here is that such definitions fail to fix the meaning of the new term, because of the following reasoning (in analogy to (Wright, 1998a), p 352):

1. Any statement involving quantifiers is only determinate if the domain of quantification is fixed in advance.

2. To specify the domain of quantification in advance involves knowing or being able to specify antecedently what types of objects it comprises.

3. Hence for an implicit definition involving quantification to determinately fix the meaning of a new term one has to be able to specify antecedently the types of objects the domain comprises.

4. Therefore, the objects to which reference is purportedly introduced by means of the definition can't be part of the range of quantification, as otherwise the implicit definition can't determinately fix the content of the new terms involved.

Clearly, if that is the worry about impredicative definition it has to do with the *effectiveness* of the definition and is quite consistent with such definitions being satisfiable. Hence, the definition might well be *productive* – i.e. the third achievement of implicit definitions – and so fail to introduce the new term in a satisfactory manner. As a result, it seems to me that whether we can allow impredicative implicit definitions is an issue that needs to be resolved within a theory of meaning. For example one issue is whether, in order to have a determinate statement, the universal quantifier has to be specified in the way characterised (step 1 and 2); that is whether in order to fix the truth-conditions of '$\forall x \phi(x)$' it is a requirement that the domain is independently and antecedently specified. It is an important issue here, whether grasping such specification of the domain is a *prerequisite* for an *understanding* of the quantified sentence. Only if this case can be made would there be a vicious explanatory circularity in even grasping the statement due to the impredicativity involved.

These are substantial issues and it seems that it would be a rather surprising outcome, if, as Dummett seems to suggest, we can't so much as *understand* impredicative definitions such as Hume's Principle or the Peano axioms! Consequently, I think we can confront the proponent of the constraint of predicativity with the following rather basic dilemma: Either, the predicativity constraint is motivated in virtue of the presumed *inconsistency-causing* nature, but then it seems otiose once the standard constraint of consistency (or satisfiability) is in place. Or, the predicativity constraint is located within the achievement of *effectiveness* of implicit definitions. However, then the theorist faces the problem in explaining away the ubiquitous intuition that many impredicative definitions are, at least prima facie, *effective*, in that they do provide an understanding of the new terms involved.

While, I won't here be able to take the issue further, dialectically I believe the onus is on the proponent of predicative-only definitions to lead a way through the above dilemma. It should also be noted that recent discussions of *unrestricted quantification* has tended to challenge various presuppositions of Dummett's view on quantification. A review of this alternative has to be postponed to another occassion.[47] In the sequel, I will assume that impredicative definitions – so long as they fulfil other constraints – are in good order and should not be rejected just because they are impredicative.

3.5.2 Non-creativity

In this section I want to revert to a question previously postponed in section 3.1.2 concerning *non-creativity*. As we have seen before, implicit definitions do not in general comply to

[47]For more discusssion see (Rayo and Uzquiano, 2006).

the requirement of *eliminability*, so the issue is whether some form of requirement of non-creativity should be imposed on these types of definitions. The requirement, as outlined before, is:

> A formula S introducing a new symbol to a theory satisfies the criterion of non-creativity if and only if: there is no formula T in which the new symbol does not occur such that $S \rightarrow T$ is derivable from the axioms and preceding definitions of the theory but T is not so derivable.((Suppes, 1957), p.47f)

This notion of non-creativity, although intuitive on first sight, can become problematic if imposed in full generality. First to be noted is that, so characterised, non-creativity is relative to the already accepted background theory (axioms and definitions). Hence a set-theorist putting forth the Peano axioms or Hume's Principle within his background theory can regard them as complying to the constraint. In contrast however, a theorist, who aims to regard say Hume's Principle as a foundational principle in the sense that he is currently only accepting the axioms of (second-order) logic and aims to put forth Hume's Principle as an implicit definition on which our knowledge, not only of numbers, but also of the existence of infinitely many objects is based, will, if the constraint of non-creativity is applicable in general, have to reject it. Basically any such foundational project is a non-starter if the constraint as formulated is in place.

As a result, we are faced with a dilemma: Either we can't put forth creative implicit definitions in any foundational sense and, accordingly, it seems that no interesting foundational project is available. Or, we allow at least some creative definitions which will then provide for some fruitful theories, but then are faced with the demand of an explanation why, in certain cases, creativeness is acceptable.

Here, I want to explore the foundational perspective. Hence the aim is to accept the second horn and so to argue that there are acceptable, yet creative, definitions. Clearly, this comes with a cost, since whether or not a definition is acceptable despite it being creative becomes a different, epistemological question: Can we ever be justified in believing that the commitments of *creative* definitions exist? This crucial question will underlie the discussion of the next chapter and will re-occur on various occasions throughout the sequel.

In the following, I will discuss a suggestion by Hale and Wright to dispense with the criterion of non-creativity as outlined above and instead adopt a notion of *conservativeness* which would allow for a traditional perspective on foundational principles.[48] Hale and Wright have suggested the following formulation of a conservativeness constraint, involving the following two conditions:

> *Hale & Wright notion of conservativeness*
>
> An implicit definition is conservative if and only if "it [does] not introduce fresh commitments which (i) are expressible in the language as it was prior to the introduction of its definiendum and which (ii) concern the previously recognized ontology of concepts, objects, and functions, etc. whatever in detail they may be."(Hale and Wright, 2000), p 302[49]

According to this formulation, Hale & Wright argue that Hume's Principle can be regarded as conservative, because although Hume's Principle entails the existence of infinitely many

[48]An important note on my use of "non-creativity" and "conservativeness": Usually conservativeness is used interchangeably with the notion of non-creativity as I used it above. In this chapter, however, I will reserve the notion of conservativeness for Hale and Wright's characterisation; it is in important respects different from Field's notion of conservativeness (section 1.2.3).

[49]See also (Wright, 1997), p. 297 (fn 49) where the following formal characterisation is give: "Let Θ be any theory with which Σ is consistent. Then Σ is conservative with respect to Θ just in case for any T expressible in the language of Θ, $\Theta \cup \{\Sigma\}$ entails the Σ-restriction of T, only if Θ entails T." Crucially (to get the second requirement above), a Σ-restricted formula restricts the first-order quantifier to the members of the domain of individuals which are not referents of the abstraction terms.

objects – a statement which can be expressed in the previous language – it does not concern the *types* of objects that have been *previously recognised* since, assuming the Caesar Problem is resolved (more on this below), Hume's Principle concerns numbers and no other kinds of objects.

This result surely suits Hale & Wright's intention to regard Hume's Principle as an acceptable implicit definition; however, I think there is some unclarity in this characterisation which I hope to eliminate. This will give rise to a revised version of the conservativeness constraint that is in the spirit of Hale and Wright's formulation.

I think that there is an ambiguity in the second part of the conservativeness condition – the part which concerns the "previously recognised ontology". One interpretation which gives rise to a relativistic understanding of conservativeness is such that it depends on the subject's recognised ontology when putting forth such definitions whether a definition is conservative or not. Thus, an implicit definition is *relatively conservative* iff it does not introduce fresh commitments which (i) are expressible in the language as it was prior to the introduction of its definiendum and which (ii) concern what is the previously recognised ontology of the subject.[50]

In contrast, a non-relative reading of conservativeness can be characterised as follows: An implicit definition is *non-relatively conservative* (or *conservative* simpliciter), iff it does not introduce fresh commitments which (i) are expressible in the language as it was prior to the introduction of its definiendum and which (ii) concern any kind of previously recognised ontology.

So, whether a definition is conservative is independent of what the subject regards as the previously recognised ontology; rather it is more general and concerns any kind of previously

[50]This relative understanding is broadly similar to the relativity involved in the non-creativity constraint. Whether or not a definition is non-creative depends partly upon the (previously accepted) "axioms and preceding definitions".

recognised ontology.

This clarification, however, exhibits a weakness in the formulation of conservativeness since both readings face some difficulties. According to the characterisation of relative conservativeness, it might seems that as long as there is no previously recognised ontology, which will be the case for the foundationalist in any case, and the previous language is weak enough, then *any* definition with *any* ontological commitments whatsoever is *conservative*! For example, we can easily imagine a case where a foundational theorist introduces the whole set-theoretic hierachy in a conservative manner. This is because his previously accepted ontology does not involve any such objects (after all he is a foundationalist) and so one might be tempted to think that if Hale & Wright's notion of conservativeness is relative, it is also *otiose.*

Considering the non-relative reading of conservativeness, it also gives rise to a difficulty since it is not clear what types of objects *any kind of previously recognised ontology* might involve. Does the non-relative notion of previously recognised ontology just comprise all concrete objects? Or all objects one can recognise without appeal to abstraction principles? For sure, a general explanation is needed what this ontology involves and crucially why, for example, numbers or other mathematical objects can't be part of the (non-relative) previously recognised ontology. For, if they were, the implicit definitions purporting to be about them would not be conservative!

In the following, therefore, I want to reconsider exactly what the aim and motivation for the conservativeness condition might be and then suggest a slightly rephrased version of the condition.

I think one main motivation for the conservativeness constraint is to provide a general basis to reject principles such as the Nuisance Principle[51] – an abstraction principle whose

[51]See (Wright, 1997), p. 191ff

truth entails that the universe *as a whole* is finite. The worry here is that since this principle is not concerned with special types of objects but constrains the size of the universe as a whole, it should be regarded as non-conservative and considered as unacceptable. So, in this respect, I think that the Nuisance Principle does not concern merely the previously recognised ontology but rather makes attribution to any object (previously recognised or not) in the universe. This is, I believe a reproachable feature of any implicit definition, since the principle is not concerned with specific objects, rather with any type of objects in the universe. As a result, we might say that the Nuisance principle is *imperialistic*: it does not concern *sui generis objects* but rather makes attributions (for example concerning cardinality or existence) to any type of object that exists.

In contrast, let us say that a principle is *not imperialistic* if any attribution (such as existence, cardinality, etc) the principle makes is only about *sui generis objects*, i.e. about objects the principle is intended to be about. So, for example, Hume's Principle is in this sense *non-imperalistic*, provided that it concerns sui generis objects, namely numbers.

I think that this might well be the motivation underlying the idea of Hale and Wright's conservativeness requirement. If that is so, then we should first note that this idea crucially depends on a successful solution of the *Caesar* and, more importantly, what in (Cook and Ebert, 2005) is called the *C-R problem*.[52] This specific version of the Caesar Problem, occurs in contexts where we have two distinct abstraction principles:

$$AP_{@_1} \ : \ (\forall X)(\forall Y)[@_1(X) = @_1(Y) \leftrightarrow E_{@_1}(X,Y)]$$

$$AP_{@_2} \ : \ (\forall X)(\forall Y)[@_2(X) = @_2(Y) \leftrightarrow E_{@_2}(X,Y)]$$

[52]I say more importantly since a resolution to the simple Caesar Problem (which concerns the issue whether some person (such as Caesar) is identical to a number) isn't sufficient for our purposes. It has to be clear whether *any* type of object (concrete or abstract) is distinct from any other object which is introduced in virtue of an implicit definition or abstraction principle. The C-R problem focuses on the issue whether abstract objects introduced by means of two abstraction principles are distinct or not.

and want to settle cross-sortal identity claims of the form:

$$@_1(P) = @_2(Q)$$

Both abstraction principles, however, are silent with regard to this identity. Yet, it is crucial to resolve the C-R problem since otherwise it isn't clear whether a certain abstraction principle is non-imperialistic or not.

Second, we should also try to harness this general idea of *non-imperialism*, which offers a good motivation to reject nuisance-like principles, into a new version of the conservativeness constraint. So, as a first suggestion, we can say that:

> An implicit definition is conservative if and only if it does not introduce new commitments which are expressible in the language as it was prior to the introduction of its definiendum, and it only makes attributions (such as existence, cardinality) to sui generis objects, i.e. to objects the principle is intended to be about.

I think that this reformulation of Hale and Wright's requirement of conservativeness is very much in their spirit and avoids the problematic appeal to a "previously recognised ontology". Surely, this notion might need further sharpening and a successful resolution of the Caesar and C-R problem will have to be offered.[53] Still, I think that the reformulated conservativeness requirement is well suited for the purposes of the traditional connection and also well motivated. Thus, I will adopt this requirement throughout the sequel. How exactly this reformulated notion of conservativeness fares in the context of the recent "embarrassment of the riches" objection in (Weir, 2003) will have to be investigated on a further occasion.[54]

[53](Hale and Wright, 2001b) put forth such a solution. In (Cook and Ebert, prep) we prepare a formal solution to the C-R problem.

[54]For what is worth, my conjecture is that it will neutralise Weir's worry.

To summarise the discussion about non-creativity and conservativeness, I propose that the constraint of non-creativity should *not* in general be in place for implicit definitions. One rather simple reason is that otherwise there is no prospect for the traditional connection (if applied to mathematics) which I here want to explore and defend. Surely, that won't convince the proponent of this constraint. In chapter 4, however, I will discuss a conception that is broadly in line with a proponent of the non-creativity constraint for implicit definitions. I will here show that this position faces various difficulties and so I will here postpone a more detailed discussion of this particular form of opposition to the traditional connection. Also, I have suggested that if the traditional connection is adopted, the condition of conservativeness (in its revised version) should be in place. I think it is well motivated and it provides a sound criterion for rejecting various unacceptable principles. Finally, as I noted before, rejecting the non-creativity condition incurs the further epistemic obligation, which is in place for any foundationalist theory, to explain on what basis one can assume that the *sui generis* objects the principle commits one to exist. This is a question that will inform much of the next chapters.

3.6 Conclusions about the framework for implicit definitions

To conclude these consideration about a framework for implicit definitions, let me quickly summarise our main proposal before moving on to another dimension of implicit definitions, that of *epistemologically tractable* foundational principles for our knowledge of logic or mathematics, as suggested by the proponents of the *traditional connection*.

In the previous sections, I have tried to provide a framework for implicit definitions. To this end, I distinguished implicit definitions from explicit definitions by pointing to various formal as well as semantic differences between them. I also suggested that the notion of stipulation is a key mechanism for an implicit definition and suggested various models concerning the scope of such stipulations. I then outlined three types of achievments for implicit definitions and suggested that, minimally, in order for a stipulation to be successful, the stipulated vehicle of the implicit definition (i.e. the sentences or rules involved) has to be satisfiable. However, I also noted that a stipulation that fails to effect a true sentence(s) might still be, in some respect, valuable or *effective*, as I tried to show with the example of Basic Law V. As a result I suggested that the standard view – which demands that an implicit definition has to be true to introduce a new meaning – has to be relaxed.[55] Lastly, I discussed two constraints that are often noted in the literature on definitions and axioms, which some philosophers and mathematicians have argued are necessary. First, is the constraint of predicativity, which I argued is a constraint that should be located within the aim of effectiveness and so within a theory of concept-constitution. I suggested it is too restrictive to account for all the principles we intuitively understand. Second, I discussed how the condition of non-creativity undermines any prospects for a foundationalist perspective and suggested instead the adoption of a reformulated version of Hale & Wright conception of conservativeness.

The foundational perspective, which regards implicit definitions as capable of underwriting and explaining our a priori knowledge of mathematics and logic, also known as the *traditional connection*, will be roughly sketched in the following concluding section. A more detailed discussion, involving the particular version of this conception that I aim to defend, is reserved for the chapters that follow.

[55] As also noted I will discuss the notion of stipulation more in the following section.

3.7 The traditional connection and its sceptics

The traditional connection is based on the idea that we can acquire knowledge of logic or mathematics in an epistemically tractable way in virtue of grasping implicit definitions of logical or mathematical principles. It credits implicit definitions with a further epistemic dimension, the ability to underwrite knowledge.

The traditional connection was first defended in the work of Carnap and the Vienna Circle in the early 20th century.[56] It has taken on various facets since, and different variations of the idea have been defended. Here, I won't be concerned with offering an historic account of its developments; rather, I will try to identify the basic ideas underlying the traditional connection.

The basic tenets of the traditional connection are two: First, its proponent is likely to be a foundationalist in the sense that basic mathematical and logical knowledge, i.e. knowledge of the axioms of mathematics and the basic rules of logic, is conceived as *non-inferential*. Second, in order to account for such non-inferential knowledge and explain how a subject can be justified in holding the axioms and basic rules to be true or valid, the traditional connection rejects the need to postulate a new faculty of "intuition" as the *intuitive platonist* (see 1.2.1) did. She also rejects the thought that purely pragmatic considerations can count as the appropriate type of justification in this case – a position held by the *naturalised platonist* (see 1.2.2).[57]

Nonetheless, a proponent of the traditional connection claims that there is a type of justification which can underwrite non-inferential knowledge of mathematics. This, she

[56]See especially (Coffa, 1991) for a good overview, chapter 14ff. See also (Boghossian, 1997) for a helpful, briefer overview.

[57]Historically speaking this is not entirely correct as Carnap adopted pragmatic considerations to justify the adoption of one or the other conceptual scheme, nevertheless no appeal was made to a purely pragmatic justification directed at specific axioms.

claims, stems from the meaning-constituting rules or statements that introduce us to the basic concepts of mathematics and logic. It is when the traditional connection is combined with the context principle that it gives rise to a form of platonism that nicely fits the previous description of *Fregean platonism*: It provides a justification for mathematical knowledge which is not constrained by the *fundamental assumption*,[58] rather by taking the 'linguistic turn', it adopts the idea that the acquisition of mastery of a mathematical discourse can all by itself underwrite knowledge of mathematics.[59]

The major question is: How exactly does this Fregan platonist conceive of the relationship between meaning-constituting sentences/rules and knowledge and so hope to resolve the second *epistemic* requirement of Benacerraf's challenge?

The (very) basic idea is that by grasping an implicit definition the subject will become justified in thinking the statement to be true once she has grasped that the definition in question is *effective*. The thought – which gives rise to some immediate problems which will be discussed below – is that the very effectiveness of the definition has a further epistemic dimension and so incorporates a warrant for the subject that gives rise to knowledge of the statement in question. So, the proponent adopts the following model which applies to a range of interesting 'good' cases of implicit definitions:

Model of the traditional connection

Step 1 Effectiveness of S $\Rightarrow_{traditionalconnection}$ Knowledge of S

Step 2 Knowledge of S $\Rightarrow_{factivity}$ Truth/Satisfiability of S

[58]The fundamental assumption states that 'if there is a priori mathematical knowledge and mathematics is construed at face-value then there has to be some form of acquaintance with the objects involved'. See p. 32 of this thesis.

[59]The Neo-Fregean position therefore adopts exactly this general framework (context principle plus traditional connection) and furthermore adopts the view that abstraction principles are in some sense better than implicit definitions that involve a stipulation of the truth of a (simple) sentence rather than a biconditional. More on this in the next chapter.

Step 3 Truth/Satisfiability of S $\Rightarrow_{def(success)}$ Success of S

The latter two steps are justified quite simply by factivity of knowledge and the definition of success. What follows is that a subject can acquire knowledge of the statement by means of its effectiveness, and then infer that the objects purportedly referred to exist. However, crucially, the step from effectiveness to knowledge is a transition that, in the light of the gap between effectiveness and success (which is entailed by knowledge), gives rise to the further issue to differentiate the 'good' cases in which such transition takes place, from the 'bad' ones in which it fails. Before attending to this issue, I will briefly deal with some initial worries concerning the notion of stipulation and the revised version of the standard view of implicit definitions that underlies the traditional connection.

3.7.1 The notion of stipulation reconsidered

The proponent of the traditional connection should not conceive of the idea of stipulation as a specific act that needs to be performed individually, as it were, in order to have the type of justification that implicit definitions can provide. This idea would lead to the unwelcome consequence that only a few subjects could claim knowledge of mathematics and logic. Rather, the idea of a stipulation is merely needed to initially explain how certain rules or statements can be meaning-constituting.[60] Once that rule or statement is adopted and considered effective within a community of language-users, the traditional connection's account of justification can be applied to any member of a language community.

A second worry has to do with the claim to truth involved in the notion of stipulation and what I called the standard view. In the light of my discussion of effectiveness, I have

[60] A further issue in this context is whether there has to be an explicit act of stipulation at all. It seems to me that a practice can evolve, as if there had been a stipulation, but none actually took place. This type of practice should also be acceptable as a candidate for providing the type of justification the traditional connections draws from explicit acts of stipulation.

so far argued that the standard view cannot be correct in that a sentence *has* to be true, in order for the definiendum to acquire meaning. Rather, I suggested that all a stipulation can do is to effect a certain attitude to regard as true the stipulated statement. As a result, I offered the following reformulation: "an implicit definition endows the new term with a meaning, in virtue of thinkers' regarding the sentence, in which it figures, as true" (see p. 121). The result of this modification of the effect of a stipulation is that the transition from effectiveness to knowledge of the implicit definition can't be assumed by default. This much is what the occurrence of Basic Law V should teach the proponent of the traditional connection.

However, even this modification might not be enough. One particular problem, that will lead to a further reformulation of the idea of stipulation, will be motivated by a very specific shortcoming of Boghossian's, as well as Hale and Wright's, proposal. This problem will inform chapter 5 and its resolution in chapter 6 gives rise to my own view of how to understand the traditional connection. Until then, however, let us assume the above suggested reformulation since various criticisms that we will encounter in this and the fourth chapter neither rely upon the specifics of the idea of stipulation nor on the conditions for a justified transition from effectiveness to knowledge.[61]

For example one often-noted worry is that opponents of the traditional connection have argued that the general notion of "truth" involved in either the standard view or the reformulation of the standard view involves a conventionalist account of truth! I think this is a misguided worry and I will briefly clarify this misconception.

The concern is that by stipulating that certain axioms to be true – or rather to be regarded as true – "these postulates [implicit definitions] create the truths that they, the

[61] Some of the criticisms appeal to the standard view, however the same worry will apply to the reformulated conception of stipulation. I will refrain from constantly rephrasing their criticism but assume that implicitly 'to stipulate the truth' is best understood as 'to stipulate to regard as true'.

postulates, express."[62] Therefore, the method of implicit definition, and especially the idea of stipulation that is part of it, is regarded as implicitly presupposing conventionalism about truth. Yet, this latter alleged component of the view – conventionalism about truth – has been subject to very convincing criticism by Quine and others[63]. Here, I will show – by appeal to Boghossian's insightful paper "Analyticity"[64] – that the method of implicit definition (including the notion of stipulation) need not involve in addition the adoption of a conventionalist theory of truth at all.[65]

Boghossian's insight is to clearly specify what the scope of the stipulation is. We need to distinguish between on the one hand the "claim that the conventional assignment of truth to a sentence determines *what claim that sentence expresses* (if any) and on the other hand the issue about what (if anything) determines the truth of the claim that is thereby expressed."[66] The latter is what might be called the truth-maker. In general, however, when a sentence is stipulated to be regarded as true, it is not part of the stipulation to stipulate the truth-maker *per se* into existence. Rather, the stipulation is merely meant to fix the meaning of the definiendum. *If* the stipulated statement is effective and satisfiable, the sentence expresses a truth. Importantly – and this is the crux – while it is part of the stipulation to decide what would count as a truth-maker of the sentence it is not part of the stipulation to bring about that there has to exist a truth-maker for the sentence. A full-blown conventionalist might think otherwise and regard the existence of the truth-maker as stemming from the stipulation. However, the point here is – and I fully agree with

[62](Creath, 1992), p.147 as quoted by (Boghossian, 1997), p.365. Creath provides a nice overview of Carnap's conventionalism in this article.

[63](Quine, 1935) and (Quine, 1954), I won't be able to rehearse his criticism here but again in (Boghossian, 1997) a good overview can be found of Quine's criticism.

[64]Two rather similar versions of this paper can be found in (Boghossian, 1996) and (Boghossian, 1997)

[65]Boghossian might not be the first to revive the traditional connection; others have defended similar views such as (Dummett, 1991b), (Peacocke, 1992), (Peacocke, 1993), (Peacocke, 2000), (Boghossian, 1996), (Boghossian, 1997), (Boghossian, 2000) and (Hale and Wright, 2000)

[66](Boghossian, 1997), p.351

Boghossian – that conventionalism is an independent component of the method of implicit definitions and the idea of stipulation, and so it is a viable option to conceive of it without adopting the full-blown conventionalist viewpoint.

This result also relieves Dummett's tension that I outlined in 2.3.3. There, I offered an interpretation of an "indisputable" tension according to Dummett. It arises, when the notion of truth involved in stipulating a sentence is not "substantial" but, rather mind-dependent; while the objects the sentence is about are claimed to be mind-independent (the platonist component). Now, if what I have argued for is accepted, the notion of stipulation can and should be regarded as involving a substantial notion of truth and not a conventionalist one. Only if the latter were the case, would there be Dummett's tension. Therefore, I contend, there need be no "head-on collision" nor tension between the methodology of abstraction principles or more generally implicit definitions, the notion of stipulation, acceptance of the context principle and platonism.

So much for some initial clarifications of the role of stipulation and some initial worries about the prospects for upholding the traditional connection. The final section of this chapter will outline what I take to be the most important issues the traditional connection has to address if it is to be a viable contender for resolving Benacerraf's *epistemic* challenge.

3.7.2 Challenges to the traditional connection

The first issue is the problem about the *ontological commitments* of some meaning constituting implicit definitions as already pointed out in my discussion of the requirements of *non-creativity* and *conservativeness* in section 3.5.2. According to the proponent of the traditional connection, meaning-constituting sentences are in some sense harmless since all they do is to fix a new concept and, if all goes well, these statements express truths. However, the statements involved in mathematics have, in case they are in fact true, strong ontological commitments. So for example, Hume's Principle entails, on a face-value reading, the existence of infinitely many objects of whose existence we can know a priori once the traditional connection is adopted. The *rejectionist* denies the justifiability of such a priori existence claims. In the next chapter, I will survey different types of rejectionist positions and discuss and evaluate them.

In chapter 5, I will discuss two attempts to explain in detail how the traditional connection works – how a subject can acquire a priori knowledge by means of implicit definitions. So far, we have yet to review any detailed model of the acquisition of knowledge in virtue of implicit definitions, especially in the light of the gap that I argued for between effective and successful definitions. To this end, I will discuss Boghossian's proposed template and his notion of epistemic analyticity.[67] I will argue that his, as well as Hale and Wright's[68], explanation of how a subject can come to have a warranted belief in the disquoted definiendum, falls foul of a transmission of warrant failure. I will then diagnose this mistake and offer a therapy by looking again at the effects of stipulating a sentence. The last chapter will suggest an alternative account of how basic knowledge by means of an effective definition can be acquired, while avoiding the previously highlighted problem.

[67]First suggested in (Boghossian, 1996) and (Boghossian, 1997).
[68]As proposed in (Hale and Wright, 2000).

Chapter 4

Rejecting Rejectionism

Introduction

In the literature, various types of opposition have been developed to the traditional connection and in particular to the invocation of abstraction principles in connection with it, as purportedly a special kind of implicit definition. One type of opposition, which I won't be discussing in detail here, is *reductionism*. The basic idea is that instead of regarding an abstraction principle as introducing a new expression that enables us to talk and know about abstract entities, such as numbers, directions, etc., we should view abstraction principles as a means of explaining away apparent reference to such objects. So, instead of reading abstraction principles from "right-to-left", and so as introducing a new kind of expression referring to abstract objects, the reductionist reads abstraction principles from "left-to-right". That means that the problematic vocabulary that purportedly refers to abstract objects on the left-hand side of an abstraction principle is eliminated in favour of the accepted vocabulary on the right-hand side. This basic idea branches out into various different types of reductionism which have been discussed extensively in recent literature.[1] Here, I won't rehearse this type of opposition, but focus on a different line of criticism of the traditional connection.

The second type of opposition is now standardly called *rejectionism*.[2] The rejectionist position presents another resort of opposition, even if it is assumed that the above conflict can be resolved in favour of the traditional connection. The rejectionist agrees that *if* the stipulations involved in abstraction principles succeed, then we should accept the "right-to-left" reading. However, he questions whether abstraction principles, conceived as implicit definitions, *do* succeed – whether, that is, they are so much as true (or a priori knowable). And the grounds for the opposition is that these principles (at least in the case of Hume's Principle) imply the existence of *too many* abstract objects.

To be sure, this is a mere crude charicature of the rejectionist view and two forms of rejectionism will be distinguished in this chapter. The first I label *semantic* rejectionism, a position that is taken by Horwich, and the other is an *epistemic* rejectionism, which is defended by Field and Boolos. I will discuss the two positions in turn and evaluate various responses to them that have been offered in the literature. Finally, having examined these replies to rejectionism, I will offer a new argument that aims to destabilise the rejectionist position. Although it might not be sufficient to convincingly *refute* the rejectionist thought, I believe it will put the onus on the rejectionist to explain away various unwelcome consequences and thereby render this position less attractive.

[1]Initially discussed in (Wright, 1983) and (Dummett, 1991a). A good overview of the three types of reductionism can be found in (MacBride, 2003), p.115ff. See also (Heck, 2000) who argues that (syntactic) reductionism is untenable.

[2]First labelled as such in (Wright, 1990), p.157

4.1 What is Rejectionism?

Roughly speaking, the rejectionist conception rejects implicit definitions that involve existential commitments and thus, at least *prima facie* adopts the non-creativity constraint as a necessary condition for the success of an implicit definition.[3] But, this characterisation is still too crude and we can distinguish two types of rejectionism.

First, consider *semantic* rejectionism. This conception rejects *tout court* the idea that implicit definitions can have ontological commitments and offers a tool to rephrase ontologically committing definitions into non-committing ones, while retaining the meaning-constituting component of such definitions. Hence, according to semantic rejectionism, any legitimate type of stipulation has to be construed in such a way that it does not involve ontological commitment.

Second, and distinct from *semantic* rejectionism, is *epistemic* rejectionism. This position accepts that stipulated sentences – stipulated for the purpose of implicit definitions – can have ontological commitment. However, epistemic rejectionism repudiates the thought that notwithstanding such commitments this type of implicit definitions can still be considered analytic, or a priori. Rather, if an implicit definition has ontological commitments then it can neither be regarded as a priori knowable nor as an analytic truth. So, epistemic rejectionism adopts what Tennant once called the *fourth dogma of empiricism*: "Having existential commitments entails syntheticity."[4]

In the following, I will discuss both types of rejectionism in turn, followed by a review of various responses to each as they occur in the literature. In the last section of this

[3] I say merely that rejectionism *prima facie* adopts the non creativity constraint. Some rejectionists reject *any* implicit definitions with ontological commitments, no matter what the background ontology is. So, even non creative definitions are – in some cases – rejected on his view. More detail will follow below.

[4] (Tennant, 1997), p. 303. Assuming that there is no category of the synthetic a priori, the dogma could also be rephrased as "having existential commitments entails that such statements are a posteriori."

chapter, I will focus on the notion of *ontological commitment* that underlies both rejectionist conceptions.

4.2 Semantic Rejectionism

The main recent proponent of semantic rejectionism is Horwich, who has argued against *substantive* (i.e. ontologically committing) implicit definitions in various papers.[5] Horwich contends that the meaning-constituting part of an implicit definition should always be distinguished from the ontologically committing part. In addition to this demand, Horwich proposes – by appeal to so-called Carnap conditionals – a means whereby one can separate what belongs to the meaning-constituting part from what is the substantive, ontologically committing part. In the following, I will briefly outline the conception behind Carnap conditionals and then discuss Horwich's motivation and use of them.

4.2.1 Ramsey sentences and Carnap conditionals

We have already encountered Ramsey sentences[6] in section 3.4. Here, I will review Carnap's own use of them and then show how they are utilised to construct so-called Carnap conditionals.

Carnap[7] recognised the fruitfulness of Ramsey sentences, and it became an integral part of his theorising to establish the linguistic rules within a linguistic framework and to distinguish so-called internal and external questions.[8] His own motivation for using Ramsey

[5](Horwich, 1997), (Horwich, 1998), (Horwich, 2000) – it should be noted that Horwich's views have changed in the course of these three papers. In his first paper he is a kind of epistemic rejectionist; however later and most markedly in (Horwich, 2000), he adopts semantic rejectionism.

[6]First introduced by Ramsey in (Ramsey, 1929).

[7]See various article of his (Carnap, 1936), (Carnap, 1939), (Carnap, 1952), (Carnap, 1963) and especially (Carnap, 1966), chapter 26-8, see also (Lewis, 1970).

[8]See his influential paper (Carnap, 1950), especially §2, where he introduces the idea of different linguistic

sentences is twofold, Carnap writes: "[t]heir function [i.e. of the Ramsey-sentence] is to serve in the explication of experiential import and, more importantly, in the explication of analyticity."((Carnap, 1963), p. 963, my addition)

The first function of the Ramsey sentence, as Carnap conceives matters, is thus to show what ontological commitments a theory carries. This can be explained quite easily. Take a theory M, with a mixture of terms for observables as well as theoretical terms. Hence take M to be the set of sentences such that:

$$\mathrm{M} := \{S(t_1), S(t_2), \ldots S(t_n), S(o_1), S(o2), \ldots S(o_m)\}$$

while $(t_{1 \ldots n})$ are theoretical terms and $(o_{1 \ldots m})$ are observable terms.[9]

By using the Ramsey method we substitute the theoretical terms within the sentences with variables, which – in a second step – are bound by existential quantifiers. The result will then consist of the following sentences:

$$M_R := \{\exists x_1 \ldots x_n S(x_1), S(x_2), \ldots S(x_n), S(o_1), S(o2), \ldots S(o_m)\}$$

M_R is a logically weaker theory that contains no theoretical terms and which, through the presence of the existential quantifiers, wears its existential commitment on its sleeve.[10] For Carnap, one virtue of the ramsified theory is that, arguably, it has the same explanatory value as the initial theory while avoiding potentially problematic vocabulary.

Although, this application was important at the time, a second usage of Ramsey sentences, as figuring in so called *Carnap conditionals*[11] has received widespread acceptance

frameworks and distinguishes internal from external questions.

[9]Whether the distinction between observables and theoretical terms can be sharply drawn need not concern us here. This example merely explains but does not defend how Carnap applied Ramsey sentences to exhibit the existential commitments of a theory.

[10]Assuming the ontological commitment the observable terms have is independently obvious.

[11]It is labelled Carnap conditional by Horwich in (Horwich, 1997), Lewis in his (Lewis, 1970) calls it the Carnap sentence.

as an efficient method for separating empirical/factual content from analytic content – a separation which was fundamental to Carnap's system.

Carnap's ingenious idea was, on the hand, to identify the factual content of a theory M with that of the corresponding ramsified theory M_R, and, on the other, to identify the *meaning postulates* or analytic content of M with the following conditional:

Carnap conditional

$M_R \rightarrow M$

Thus, if the Ramsey sentence turns out to be true, then combined with the meaning postulates we can infer by modus ponens the truth of the theory M (including the new terms). However, if the theory is falsified, then the Ramsey sentence will be false. Consequently, the antecedent of the Carnap conditional is false, which makes the conditional itself true. Therefore, the Carnap conditional can be regarded as necessary, and so still accounting for the meaning of the terms, despite the possible falsity of the underlying theory. By adopting this model, Carnap hoped to resolve the problem of *how theories can fix the meaning of terms despite the fact that they might be falsified.*[12]

To put matters differently: One may quarrel over the truth of the Ramsey sentence, but one will still be in agreement over the truth of the conditional, which provides the meaning for the scientific terms, whether or not the Ramsey sentence is true. Carnap thus concluded that the Carnap conditional "is a purely analytic sentence, because its truth is based on the meanings intended for the theoretical terms."[13]

Horwich, as we will see in the following, uses this conception of Carnap conditionals to try to undermine the idea that there can be implicit definitions that have ontological import.

[12] For example, in this way, the meaning of the term "phlogiston" can be accounted for, despite the falsity of the underlying theory.

[13] (Carnap, 1966), p.270

4.2.2 Horwich's usage of the Carnap conditional

Horwich picks up on the idea by Carnap and claims that no implicit definition *should* be considered as having ontological commitment. Rather, every implicit definition is, or if not overtly then at least 'implicitly' of the form of a Carnap conditional. So what one might call the 'vehicle' of the implicit definition, i.e. the part which fixes the meaning of the term, is always to be reconstructed as a conditional of this kind. He writes:

> "Thus the stipulation that 'f' shall mean whatever is needed for '#f' to be true, is *not* purely meaning constituting; rather, our commitment to the conditonal, '$\exists x(\#x) \rightarrow \#f$', is what provides 'f' with its meaning. ((Horwich, 2000), p. 157)

Shortly after this claim, he applies the same point to Hume's Principle and also to other mathematical and logical implicit definitions. He writes:

> As in the case of a physical theory, Hume's Principle can be divided up into two parts. First, there is
>
> (HS) $\exists\eta\forall f\forall g(\eta(f) = \eta(g) \leftrightarrow f \sim g)$[14]
>
> [...] which asserts that there is some function from properties to objects that satisfies Hume's Principle. Second, there is the conditional
>
> (HM) $\exists\eta\forall f\forall g(\eta(f) = \eta(g) \leftrightarrow f \sim g)$
> $\rightarrow \forall f\forall g$(the number of f's = the number of g's$\leftrightarrow f \sim g$)
>
> And surely it is no less plausible here than it is in the case of a physical theory to suppose that it is our acceptance only of this conditional, (HM), that is required for 'the number of 's" to mean what it does. [...] Our acceptance of (HS) is

[14]Basically the ramsified version of Hume's Principle

> an *additional* requirement, [...], which is not required for 'the number of 's' to mean what it does.((Horwich, 2000), p. 158)

To repeat, according to Horwich, an implicit definition whose aim is to fix the meaning of a term is always to be understood as encapsulating a conditional. That is, we may overtly stipulate a sentence such as '#f' to fix the meaning of a term, but the properly conceived vehicle of the stipulation is '$\exists x \# x \rightarrow \# f$'.

Having clarified how we should, according to Horwich, conceive of implicit definitions, not only in the scientific discourse but also in logic and mathematics, we need to look at *why* this is the view we *ought* to adopt. Unfortunately, Horwich offers little in terms of positive motivation. His main motivation stems from the apparent success of the model in the scientific case. He writes:

> "If such [Carnap] conditionals do give the proper form for an implicit definition of a scientific term, then it would seem **natural** to suppose that *all* implicit definitions have something like this form, including those of arithmetical, geometrical and logical terms."((Horwich, 1997), p. 426 my bolding, italics in original text)
>
> "As in the case of a physical theory, Hume's Principle can be divided into two parts" [...] "I see little reason not to treat logic analogously." ((Horwich, 2000), p.158)

I shall offer two observations about Horwich's reasoning. First, I will exhibit a prima facie weakness in the structure of Horwich's reasoning. Second, I offer what seems to me the best way to repair the weakness, but then point out that it is still insufficient to threaten Horwich's opponent.

As it is evident from the quotations above, Horwich's reasoning is somewhat tentative, in that he concludes that only a Carnap conditional is *required* to fix the meaning of a term and nothing stronger, for example Hume's Principle itself, is needed. Crucially, however, in order to convince the opposition, some additional grounds are needed to show that it is *wrong* to adopt the more substantial view (the non-conditionalised version) or, at least, to point to independent reason as to why the substantial view is problematic. So far, it seems, that Horwich's "punchline" is that the Carnap conditional is *sufficient* to account for the meaning of the term. Yet, it seems that a compelling argument is missing that also shows that it is necessary to use nothing stronger than Carnap conditionals – or at least to provide some reason why it is advantageous to do so.

Despite this initial weakness, one can so interpret the analogy to the empirical case which Horwich makes so as to produce a strengthening of his reasoning. Carnap conditionals are not only sufficient but also necessary in empirical science, since it is *only* by appeal to such conditionals that we can account for the possible falsification of the empirical theory while still having the means to account for the meaning-constituting function of the implicit definition. This is exactly Carnap's, and now Horwich's, motivation for adopting these conditionals. Once Carnap's model for implicit definitions is adopted for scientific theories, Horwich contends that it is *natural* to use the same model in mathematics and logic. On this basis, he arrives at the conclusion that implicit definitions can neither carry ontological commitments nor found any knowledge of objects (a priori or otherwise), since the vehicle of implicit definition is always an ontologically non-committal conditional.

In the following, I will investigate two thoughts underlying Horwich's position: First, I examine what reason there is to suppose that, if Carnap's model is accepted in empirical sciences, it should be applicable in mathematics and logic as well. Second, I reconsider the initial motivation to adopt Carnap conditionals in the first place.

The proponent of the traditional connection can, I think, justifiably reject the transition from the applicability of Carnap conditionals in science to their adoption in mathematics.

Since empirical statements are subject to empirical (dis-)confirmation, there is a genuine role for the Ramsey sentence, if we are to explain how implicit definitions might be falsified through experiments without compromise of their meaning. For empirical statements it seems that conceptual clarity, i.e. understanding of the terms involved, is presupposed in order for (dis-)confirmation of a statement, for which an appeal to Carnap conditionals might well be necessary.

In contrast, for anyone who accepts the traditional connection, there is no such role for Ramsey sentences and the Carnap conditional for mathematical statements. First, whether or not a thinker believes certain mathematical statements to be true is independent of an empirical confirmation of its Ramsey sentence. It is the driving thought of the traditional connection to show how, without empirical, pragmatical or intuitive resources a thinker can still be justified in thinking that mathematics is true. Second, it is generally unclear whether the same kind of conceptual clarity is needed for a disconfirmation of a mathematical statement. Showing that an implicit definition in the mathematical case is false/inconsistent, need not, at least prima facie, presuppose the same kind of conceptual clarity or understanding of the implicit definition that has to be accounted for by the corresponding Carnap conditional.

As a result, for a proponent of the traditional connection, it will be *far* from *natural* to regard empirical and mathematical statements as on a par. So, at least in the context of the current dialectic, I think the proponent of the traditional connection is justified to reject Horwich's uniformity assumption.[15]

[15] Horwich's reason to adopt the Carnap conditional in mathematics and logic is driven by broadly Quinean conception which I briefly discussed and put aside in the first chapter of this thesis. Noteworthy here is that according to Horwich there can't be any substantial a priori knowledge of mathematical statements and so in this respect semantic rejectionism leads to its epistemic counterpart.

Also, I think that in addition, Horwich's position itself is unstable. If it can be assumed that Hume's Principle or the Peano axioms are both conservative on the reading I put forth in the previous chapter,[16] then it is reasonable to suppose that the Peano axioms and Hume's Principle are both *dispensable* in scientific theorising and so these mathematical statements are not subject to confirmation or disconfirmation in any scientific sense. To clarify, neither of the two statements is essential to the scientific theorising since neither of them gives rise to new theorems or predictions about any other objects but numbers (previously recognised or not), that cannot be proved without them. Therefore, the adoption of a Carnap conditional for these mathematical statements is pointless since the corresponding Ramsey sentence can't be subject to empirical confirmation or disconfirmation.[17]

As a result, Horwich's claim that it is *natural* to suppose Carnap conditionals as the vehicle of implicit definition in mathematics is not only unfounded in the present dialectical setting but even problematic. Not only is there no genuine motivation to adopt Carnap conditionals in the case of mathematics that applies to the perspective of a proponent of the traditional connection, but more worryingly adopting this model in mathematics, or at least for those mathematical statements that are conservative, is utterly pointless even on Horwich's conception.

In light of this concern, let us now investigate Horwich's initial considerations in favour of the adoption of the Carnap conditional for empirical statements in the first place. He argues that this model nicely accounts for how a theory can fix the meaning of a term despite the theory later being falsified. The meaning postulates remain true, even if the Ramsey sentence is false.

[16] Alternatively, it would suffice for these principle to be just non creative in the previously outlined sense (see section 3.5.2)

[17] This is very much in line with ideas of Field's error-theoretic nominalism and his rejection of the indispensability argument in section 1.2.3 and 1.2.2 respectively.

However, recall the framework for implicit definitions that I outlined in chapter 3, where the distinction between the effectiveness and the success of an implicit definitions was drawn. I argued that even certain inconsistent implicit definitions can be considered effective in that they still fix the sense of an expression. This, I argued gives rise to a revised version of the standard view of implicit definitions, where terms acquire meaning in virtue of our merely *regarding* a sentence as true which might later be falsified, rather than stipulating that the sentence *has* to be true and *remain* true. Adopting this revised conception of how an implicit definition fixes meaning, therefore, makes redundant the need to account for the meaning-constituting part by appeal to *necessary truths* such as the Carnap conditional and so undermines Horwich's initial motivation.[18]

I contend then, that an alternative explanation of how terms can retain their meaning is available despite the falsity of the underlying theory and without the appeal to Carnap conditionals. Moreover, we observed that there is no relevantly analogous motivation for resorting to such conditionals in mathematics in any case. Finally, the fact that mathematical principles are conservative and dispensable from scientific theorising and so not subject to empirical confirmation or disconfirmation, leads me to conclude that Horwich's version of semantic rejectionism is neither a stable nor a well motivated position and as such no genuine threat to the prospects of the traditional connection.

4.3 Epistemic Rejectionism

In contrast to semantic rejectionism, epistemic rejectionism is not so much concerned about the possibility of ontological commitments' being carried by meaning-constituting implicit definitions. Rather the epistemic rejectionist worries about the additional claim that ontolog-

[18]Interestingly Horwich actually adopted the above revised version of the standard view himself, while at the same time arguing for a need for a Carnap condition, which renders his position 'slightly' puzzling.

ically committing statements can be regarded as a priori knowable or as analytic. There are various proponents of this position. I will focus here mainly on George Boolos' arguments.[19]

Boolos' formulation of the worry can be found in the suitably entitled paper "Is Hume's Principle analytic?" Importantly, his reflections do not only apply to Hume's Principle, but more generally to any principle, axiom or implicit definition that entails existential commitments and so it is a general worry for the traditional connection. Furthermore, I believe that similar considerations would presumably apply to the idea of thinking that Hume's Principle or any other ontologically-committing principle is a priori knowable, rather than analytic (if there is any difference).

Boolos' view is that statements that are analytic involve (at least) two components: "Firstly, they are true; secondly and roughly speaking, they lack content, i.e., they make no significant or substantive claims or commitments about the way the world is; in particular, they do not entail the existence either of particular objects or of more than one object."[20] Clearly, Boolos here adopts what Neil Tennant has dubbed the 'fourth dogma'.[21] More precisely, and concerning Hume's Principle, Boolos continues: "The main significant worry for the defender of the analyticity of HP concerns the quite strong content that it appears to possess."[22] Namely, "if HP holds [...] there must be infinitely many objects"[23], and so he concludes: "[HP] taken as an *axiom*, might then entail that, for example there are many, many objects, *too many for it to be capable of being regarded any longer as analytic.*"[24]

[19]See especially (Boolos, 1997), also there are similar thoughts in his earlier papers, for example (Boolos, 1987a), (Boolos, 1987b) and (Boolos, 1990). Field can also be considered an epistemic rejectionist. See his review article (Field, 1984) of (Wright, 1983). However, there are also passages in Field that suggest a semantic rejectionist reading. So for example he also suggest a conditionalised reading of Hume's Principle (slightly different than Carnap conditionals).

[20](Boolos, 1997), p. 303

[21]That is "having existential commitments entails syntheticity."(Tennant, 1997), p. 303

[22](op.cit, p. 304)

[23](op.cit, p. 306)

[24]op.cit., p.309

These passages provide an initial picture of the position I labelled epistemic rejectionism. In section 4.5 I will return to the issue of the extent of the ontological commitments of Hume's Principle. First, I will discuss various replies to epistemic rejectionism that have been offered by Hale and Wright. One might expect an opponent of epistemic rejectionism to merely point to a fundamental disagreement concerning the 'fourth dogma' of empiricism. This is not the case however, and as it turns out, the ensuing discussion between Hale and Wright and Boolos offers important insights into the respective positions and rather intricate arguments.

4.4 Hale and Wright's responses to the epistemic rejectionist

Hale and Wright's responses turn on the particular form of *abstraction principles* and the way they conceive, in general, of the implicit definition as legitimately working. A first line of response, which occurs in different papers (in different degrees of detail), argues for the need to distinguish implicit definitions that involve stipulation of the *truth* of a statement from abstraction principles that merely involve the stipulation of *truth-conditions*. Grasping this distinction offers a line of response to the epistemic rejectionist. Based on our framework for implicit definitions we can enlarge on this – according to Hale and Wright – crucial distinction.

In section 3.1.2 we discussed various different models of implicit definitions and distinguished the following two cases:

Model for implicit definitions by stipulating truth

"$\ldots\xi\ldots$" is true

and on the other hand:

Model for implicit definitions by stipulating truth-conditions

"$\ldots\xi\ldots \leftrightarrow A$" is true

and also noted that Abstraction Principles fall under the latter model (at least if regarded in their schematic form).

Abstraction Principles

"$(\xi(F) = \xi(G) \leftrightarrow F \approx G)$" is true

The question, then, is why should this difference make a significant difference? After all the second model is just a special case of the first model and in both cases the fundamental stipulation is the truth of a sentence – whether it involves a biconditional or not seems prima facie irrelevant.[25]

However, Hale and Wright argue that there is an important feature in merely stipulating biconditionals which marks a significant difference from other implicit definitions. They express their point like this:

> Whenever the meaning of a functional expression 'ξ' is fixed by means of an abstraction principle "$\forall F \forall G(\xi(F) = \xi(G) \leftrightarrow F \approx G)$", what is stipulated as true is always a (universally quantified) *biconditional*, so that what is done is to fix the truth-*conditions* for identities linking ξ-terms. The truth-value of instances of the abstraction's left-hand side is never itself a matter of direct stipulation – if any identities of the form "$\xi(F) = \xi(G)$" are true, that is always

[25] Also reconsider my discussion of the interrelationship between the different models of implicit definitions on p.118 where I hinted towards the idea that each model of implicit definitions can be turned into the others (although I also noted at least some restriction). The fact that I'm using the standard view of stipulation in this discussion is irrelevant. Similar consideration apply to the revised version. Note that in the following quotations Hale and Wright do adopt the standard view.

the product of two factors: their truth-conditions, as given by the stipulation, together with the independently constituted and, in the best case, independently ascertainable truth of corresponding instances of the abstraction's right-hand side. The existence of the referents for ξ-terms is therefore never part of what is stipulated. (Hale and Wright, 2000), p. 146 (as in (Hale and Wright, 2001a))[26]

The point made here is that abstraction principles in general are not – *per se* – ontologically committing. It is consistent with the stipulation in question that the statements on the right-hand side are regarded as false and so, as long as the statements on the left-hand side are regarded as false too, the stipulation is still true. Surely, in the case of Hume's Principle, the right-hand side is, if instantiated with the predicate "non-self identical", a logical truth – however, and crucially, this is an independent component which is not part of what is stipulated.

Consequently, according to Hale and Wright, the stipulation of Hume's Principle and abstraction principles in general does not, in contrast to the stipulation of other forms of mathematical axioms, such as the Peano axioms, have *direct* ontological commitments. They write:

> In sum: the stipulation that the Peano axioms are true would be a stipulation that there is an infinite population of objects behaving as they require. We are urging, by contrast, that the stipulation of *Hume's Principle should be seen first and foremost as a meaning-conferring stipulation*–[...]–of which it is a relatively *un-immediate*, interesting and welcome consequence that there is a population of objects of which the Peano axioms are true of." (op.cit, p. 148, my italics)

Two strategies to respond to the epistemic rejectionist can be detected in these passages.

[26]This type of response has been put forth in various papers for example (Wright, 1990) (p.162f) and (Hale, 1994) (p.184) (as quoted from (Hale and Wright, 2001a))

Firstly, in the second quotation Hale and Wright make the claim that Hume's Principle should be considered *first and foremost as a meaning-conferring stipulation.* Exactly how this could help to resist the epistemic rejectionist view will be discussed in the next section below. Secondly, and, as we will see, connected to the first claim, is Hale and Wright's distinction between *un-immediate* commitments of Hume's Principle and direct commitments of the Peano axioms. What exactly constitutes this difference and how it might help in the current discussion will be discussed in the next but one section. What both strategies of reply have in common is the aim to draw an epistemically significant distinction between abstraction principles and other mathematical implicit definitions, which do not share the specific features of abstraction principles.

4.4.1 Abstraction principles as meaning-conferring stipulations

In various passages Hale and Wright focus on attributing to Hume's Principle, and for that matter to abstraction principles in general, the role of meaning-constituting principles which serve to explain a new concept. The above quotation points to the role of Hume's Principle as "first and foremost a meaning-conferring stipulation", and there are various earlier passages where it is suggested that abstraction principles should mainly be regarded as meaning-constituting and not as ontologically committing:

> "The fundamental truths of number theory would be revealed as consequences of an explanation: a statement whose role is to fix the character of a certain concept." (Wright, 1983), p. 153
>
> "A legitimate abstraction principle, in short, ought to do no more than introduce a concept by fixing truth-conditions for statements concerning instances of that concept. In a limiting case, it may associate such statements with truth-

conditions of such a kind that they are necessarily fulfilled – that is just what Hume's Principle sometimes does. *But that should not blow away the distinction between concept-formation on the one hand and mere axiomatic stipulation about existence on the other.*" (Wright, 1997), p.296 my italics

"I really don't see why the fashion in which Hume's Principle – if it indeed succeeds in doing so – determines the truth-conditions of statements which configure the cardinality operator with second-order logical concepts, should be epistemologically any more problematic than *any definition or other form of stipulation whose effect is to fix the truth-conditions of statements containing a targeted (type of) term.*" (Wright, 1999), p, my italics

The quotations hint towards the thought that Hume's Principle (and other 'good' abstraction principles in general) should be regarded as a mere explanation of a concept and that as such they are epistemologically unproblematic. This again is reminiscent of my earlier distinction between the effectiveness and the success of a definition that I drew in the previous chapter. In my terminology, Hale and Wright are urging us to regard Hume's Principle first and foremost as an implicit definition whose aim is *effectiveness*. So, the ontological commitments are not directly relevant to the achievement of this aim of effectiveness.

However, as a reply against the epistemic rejectionist this is insufficient. Namely, the proponent of the traditional connection aims to move from the effectiveness of Hume's Principle to knowledge of its truth, which in turn requires the success of that statement. And for Hume's Principle to be successful there have to be infinitely many objects that satisfy the principle. However, the rejectionist will deny that the effectiveness of Hume's Principle suffices for a priori knowledge of it, since to know it a priori requires (assuming closure) the possibility of knowing a priori that there are infinitely many objects; and that this is possible is exactly what the rejectionist denies.

Consequently, the fact that "*Hume's Principle should be seen first and foremost as a meaning-conferring stipulation*" makes no immediate difference for the rejectionist. He will maintain that regarding Hume's Principle as meaning-constituting should not distract from the additional demand that it needs to be satisfiable in order to found our a priori knowledge of arithmetic. Even if the function of an implicit definition is merely to introduce and explain new expressions, the point remains that if it thereby incurs ontological commitments, it either should be re-parsed into a meaning-constituting part and an ontologically committing part [the previously criticised claim of the semantic rejectionist], or else one has to give up on the claim that the definition in question is a priori knowable or analytic. Hence, I don't believe that a mere emphasis on meaning-determination helps, all by itself, to undermine the rejectionist complaint.

But, there are further considerations that might make the difference. As noted in the last quotation of section 4.4, Hale and Wright maintain that there is a genuine epistemic difference between Hume's Principle and the Peano axioms: Hume's Principle's and other abstraction principles' ontological commitments are in some sense less direct, or "un-immediate", whereas the Peano axioms involve a direct commitment. On what basis this distinction can be drawn and whether it can make the sought-after epistemic difference will be the topic of the next section.

4.4.2 Direct vs. indirect ontological commitments

Hale and Wright maintain that there is a difference between direct ontological commitments (or mere axiomatic stipulation) of the Peano axioms on the one hand, and the un-immediate or indirect commitments of Hume's Principle on the other. The fact that abstraction principles are of the second category provides them with an epistemic advantage, which could be used to address the epistemic rejectionist worry.

In the following, I will suggest three ways in which this distinction might be drawn and discuss in turn the fruitfulness of the three different approaches.

The first option to draw the difference between direct and indirect ontological commitments can be extracted from the above quotation: "the distinction between concept-formation on the one hand and mere axiomatic stipulation about existence on the other." ((Wright, 1997), p. 296) The suggestion would be that meaning-constituting definitions only have indirect or un-immediate commitments while axiomatic stipulations about existence have direct commitments.

Accordingly, one might say that the set-theoretic axiom of infinity or Russell's axiom of infinity are mere axiomatic stipulation about existence while Hume's Principle is meaning-constituting and so has mere indirect commitment. Nonetheless, drawing the distinction in this way invites the question why we should not also regard the Peano axioms as meaning-constituting and so not as a mere axiomatic stipulation? Remember, in the broader framework of implicit definitions outlined in the previous chapter, I distinguished various models of implicit definitions, including the model of structural definitions (see p. 117):

Model of structural definition

"$\ldots \xi R \pi \ldots$ & $\ldots \pi S \tau \ldots$ & $\ldots \tau T \phi \ldots$ " are true

where either terms such as ξ, π, τ and ϕ or predicates (R. S or T) are introduced as new expressions.

Consider now the conjunction of the Peano axioms which is the following statement:

Peano axioms as an implicit definition

$$\forall F \forall x \forall y \forall z \Big(N0 \wedge ((Nx \wedge Sxy) \supset Ny) \wedge ((Sxy \wedge Sxz) \supset y = z) \wedge ((Sxz \wedge Syz) \supset x = y) \wedge (Nx \supset \exists! t Sxt) \wedge ([F0 \wedge ((Fx \wedge Sxy) \supset Fy)] \supset \forall t (Nt \supset Ft)) \wedge \neg \exists t St0 \Big)$$

Now, why can't the Peano axioms be reasonably regarded as an example of the structural model, and so as an implicit definition fixing the sense of the expressions "Nx", "Sx" and "0"? Unless it can be argued that either structural definitions in general, or this specific version of the Peano axioms, fails to be meaning-constituting (and so at best fulfils the achievement of *productiveness* for implicit definitions), there is at least a *prima facie* case for regarding the Peano axioms as an *effective* definition.[27] For that reason, drawing the distinction between direct and indirect commitments by appeal to the meaning-constituting character does not provide the right result for Hale and Wright.

A second option for drawing the distinction between direct and indirect commitments would be to base such contrast on the *deductive complexity* required in order to make the existential commitments explicit. Axioms such as Russell's axiom of infinity or the set-theoretic axiom of infinity are statements that have *explicit* commitments. One does not need to make further deductive steps to make explicit the ontological commitments of these statements. By contrast, in the case of Hume's Principle the proof to exhibit the full ontological commitments – i.e. the proof of the existence of infinitely many numbers – is a substantial mathematical accomplishment involving numerous deductive steps.[28]

I believe one could draw the distinction on the basis of "deductive complexity", but it is very doubtful that Hale and Wright intend to found the distinction in this way, nor in any case would this idea draw a line between Hume's Principle as indirectly committing and the Peano axioms as directly committing. For, to exhibit the full ontological commitments

[27] I say 'prima facie' since there is logical space for a position that denies the effectiveness of the Peano axioms, although I don't think that this is an attractive position.

[28] The theorem was first explicitly noted by Parsons, in his (Parsons, 1965) and later independently "rediscovered" in (Wright, 1983), p.158-69. More recent presentation of the proof can be found in (Boolos, 1987a) (discursive), (Boolos, 1990) (rigourous),(Boolos, 1995) and (Boolos, 1996).

of the Peano axioms also involves a complex deduction and so they would likewise involve merely indirect ontological commitments.

Rather, I believe that the intended distinction between direct and indirect committments is founded in more specific features of abstraction principles. It is not merely that an abstraction principle fixes truth-conditions for statements concerning instances of the new concept as opposed to being an outright stipulation of truth. As we have noted, it does involve the stipulation of the truth of a (quantified) biconditional. But since it merely stipulates a (quantified) biconditional, it leaves open the possibility that instances of the biconditional are true, due to the falsity of both the right and the left side. The fact that, in the case of Hume's Principle, an instance of the right-hand side is a logical truth is an *independent* fact and it is only with the addition of this truth that Hume's Principle is ontologically committing. Consequently, one might propose that the sense in which abstraction principles in general have merely *indirect* ontological commitments, consists in the fact that there are collateral assumptions, namely the truth of the right-hand sides, which then coupled with the stipulated truth of the whole (quantified) biconditional involves ontological commitments. It is due to these additional truths that this abstraction principle, in contrast to Peano axioms, does not have direct commitments whereas the latter is not so dependent on other independent truths (or so it seems). Importantly, the claim here is not that certain logical resources are needed to deduce the ontological commitments (which obviously are needed in the case of Peano axioms), but that there are independent assumptions arising from a specific conception of the *metaphysics* of logic that are crucial to establish the ontological commitments of Hume's Principle.

What exactly this specific conception of the metaphysics of logic involves and how in detail a reply to the epistemic rejectionist on this basis can be formulated will be discussed in the next section.

4.5 Whence the infinity?

The aim in this section[29] is to further scrutinise the necessary conditions for Hume's Principle to be ontologically committing and as such to identify the source of the ontological commitments. The conclusions, I hope, can then be generalised to other abstraction principles in mathematics. Once it is clear what the source of the infinity of Hume's Principle is, we should be able to re-evaluate the epistemic rejectionist claims. First, however, I will clarify further what epistemic rejectionism is committed to and provide a theoretical basis for the position by introducing the notion of *presumptuousness* as the underlying criterion. Then I will review certain results as shown by (Shapiro and Weir, 2000) and in formal detail by (Cook, 2003) which prima facie put pressure on epistemic rejectionism. In the sub-section 4.5.4, I will propose a short thought-experiment to highlight the problem for rejectionism posed by the formal results and then suggest various responses on behalf of the epistemic rejectionist. The upshot will be to elicit a new and basic disagreement between epistemic rejectionism and the traditional connection.

4.5.1 Epistemic rejectionism scrutinised and the condition of presumptuousness

Previously, I offered a rough characterisation of the epistemic rejectionist position by appeal to the following claim of Boolos: "The main significant worry for the defender of the analyticity of HP concerns the quite strong content that it appears to possess."[30] Namely, "if HP holds [...] there must be infinitely many objects."[31] And, he holds, such commitments

[29]In various respects, the ensuing discussion is reminiscent of the debate between Boolos and Dummett concerning the source of the inconsistency of Basic Law V. See (Dummett, 1991a), chapter 17 and Boolos's reply in (Boolos, 1993).

[30](op.cit, p. 304)

[31](op.cit, p. 306)

can't be known analytically or a priori.

But, what exactly is it that generates the commitment to infinitely many objects? If we construct the Ramsey sentence of Hume's Principle – as suggested in chapter 3 – in order to discern its ontological commitments, we get following statement:

Ramsified Hume's Principle (RHP)

$\exists\eta\forall f\forall g(\eta f = \eta g \leftrightarrow f \sim g)$[32]

That is, Hume's Principle commits one to the existence of a higher-order function – an entity, whichever it is that maps concepts on to objects, giving distinct values for each pair of non-equinumerous concepts. Hence, it seems in the case of Hume's Principle that it is the existence of this function which is problematic and can't be assumed a priori. This seems to be exactly what Boolos has in mind in the following passage:

> Our present difficulty is this: just how do we know, what kind of guarantee do we have, why should we believe, that there is a function that maps concepts to objects in the way that the denotation of octothorpe [the symbol Boolos used to denote the number operator] does if HP is true? If there is such a function then it is quite reasonable to think that whichever function octothorpe denotes, it maps non-equinumerous concepts to different objects and equinumerous ones to the same object, and this moreover because of the meaning of octothorpe, the number-of-sign or the phrase "the number of". *But do we have any analytic guarantee that there is a function that works in the appropriate manner*? ((Boolos, 1997), p. 306, my italics)

So, Boolos believes that it is – we might say – *presumptuous* to assume that we can have

[32]'η' is a variable ranging over functions from properties to objects and 'f $\sim$ g' means that there is a one-to-one relation between the f's and the g's.

analytic or a priori knowledge of the existence of such a function since it commits one to the existence of *too many objects.* So, as a reconstruction of Boolos' thought we can put forward the following general criterion of *presumptuousness*, challenging the a priori/analytic knowability of an abstraction principle.

Criterion of Presumptuousness

Assuming (a priori) the existence of a function is *presumptuous*, if and only if its application has *further ontological commitment* on the object level.

In order to be more precise we have to clarify the notion of *further ontological commitment.* There are various interpretations that might be used to elicit this intuitive thought and we need to be very careful what the epistemic rejectionist might mean when he complains about *substantial*, i.e. ontologically committing, definitions.

One might render Boolos' worry about ontologically committing definitions as about those statements that commit one to "too many objects", defining the notion of *further ontological commitment* as something like the following:

Further Ontological Commitment I

A function has further ontological commitment if the assumption of its existence commits one to the existence of at least κ many objects in the first order domain, whereas κ is such that it has at least cardinality $\aleph_0$.

This just means that a function is *presumptuous*, if taking it as well-defined, it commits us to infinitely many things.[33] Thus, on this reading the notion of *ontological commitments* is independent of the ontological commitments of our underlying and already accepted theory. And it seems to me that it is this non-relative notion of "further ontological commitments"

[33]Surely one might choose a smaller or bigger cardinality than $\aleph_0$ in order to capture the notion of "too many objects".

which does motivate Boolos when he claims that Hume's Principle commits us to too many objects – i.e. too many to be regarded as an *a priori* principle.

However, there is space for a second understanding of further ontological commitment, also in line with the idea of the epistemic rejectionist, but closer to the idea of the previously encountered non-creativity constraint.

> *Further ontological commitments II*
>
> A function has *further ontological commitments*, if the assumption of its existence commits one to *more* objects – in terms of cardinality – than previously acknowledged.

To clarify, if our previously accepted domain of objects has cardinality κ, then a function has further ontological commitments, if the assumption of its existence commits us to the existence of at least λ many objects in the first order domain, and $\lambda > \kappa$.[34]

So, we have two notions of *presumptuousness* to underwrite the epistemic rejectionist complaint that the Neo-Fregean *presumptuously* assumes the existence of the function referred to by the number-operator. On the first reading of *further ontological commitment* Hume's Principle clearly is presumptuous while on the second reading, it is presumptuous only if we are not already committed to the existence of infinitely many objects.

In the following, I want further to explore the epistemic rejectionist's thought by discussing different background assumptions. As noted by Hale and Wright, Hume's Principle

[34] A third and slightly stronger version of the notion of *further ontological commitments*, which takes into account the kind of objects involved and thus involves a resolution of the Caesar/C-R problem, can be characterised in the following way:

> *Further ontological commitments III*
>
> A function has *further ontological commitments*, if the assumption of its existence commits one to *new kinds* of objects, which we previously have not acknowledged.

Here I will leave aside this characterisation as it merely complicated issues.

itself does not entail the truth of instances of the right-hand side; rather, it is only in conjunction with another presumably necessary truth that Hume's Principle entails the existence of infinitely many objects. In the next section, I want to exhibit how crucial this additional component is, and that it is not without additional non-trivial assumptions that Hume's Principle is ontologically committing. To this end, I will briefly discuss two different *metaphysical* pictures giving rise to two different comprehension schemes which, arguably, affect the presumptuousness of the introduction of the operator involved in Hume's Principle.

4.5.2 Different comprehension schemes

There have been a number of previous discussions of aristotelian logic and Hume's Principle.[35] Hume's Principle is (obviously) a statement in second-order logic which involves higher-order quantification. Higher-order quantification itself raises various questions and here we need to distinguish a number of interrelated issues pertaining to the domain of quantification of the higher-order quantifiers.

One issue is what *kinds* of entities are in the domain of the higher-order quantifier, another is *how many* of these entities are there and lastly, *what does it take* to be in the domain of the higher-order quantifers. Leaving aside the first question and thus issues raised by Quine, Boolos and many others[36], let us grant the existence of properties as the entities that higher-order quantifiers range over and focus on the second and third issue. In the context of this chapter, I won't argue for a specific answer to these questions; rather I would like to note that two traditional positions on *what it takes* for a property to exist can be distinguished, which lead to different conception of *how many* properties there are.

[35]Well, two publications actually. It is first mentioned in (Shapiro and Weir, 2000) and recently worked out in formal details in (Cook, 2003).

[36]See (Quine, 1986) for his notorious "Wolf in sheep's clothing" – phrase and (Boolos, 1985), and (Boolos, 1984) for Boolos' suggestion of plural interpretation of the second-order quantifiers.

We can distinguish an *aristotelian* conception from a *platonist* conception. What it takes for a property to exist according to the aristotelian is that, the property has to be *in rebus*. Roughly speaking, properties are concrete universals that exist provided that they are exemplified by a concrete object. The platonist, in contrast, has no such restriction, rather properties are *ante rem*, and their existence does not depend on whether they are concretely instantiated or not.

As result, *how many* properties there are depends upon the metaphysical framework that is adopted. For the platonist – translated into modern philosophers "slang" – standard comprehension is true, while the aristotelian, on the other hand, insists on a modification of the standard comprehension scheme, since only those properties that are actually instantiated (*in rebus*) exist. So, the modern platonist embraces the following comprehension scheme:

Full comprehension scheme (FCS)

$\exists F \forall x_{1...n}(Fx_{1...n} \leftrightarrow \phi x_{1...n})$

while the modern aristotelian would embrace the following restricted comprehension scheme:

Aristotelian comprehension scheme (ACS)[37]

$\exists x_{1...n} \phi x_{1...n} \rightarrow \exists F \forall x_{1...n}(Fx_{1...n} \leftrightarrow \phi x_{1...n})$[38]

Let us assume that both positions are genuine positions in metaphysics in that each conception can be sufficiently motivated (at least from a neutral position). What I will show in the next section is that these two distinct positions that are reflected in two different comprehension schemes have a major effect on the issue of presumptuousness.

[37]There are various ways of characterising the aristotelian comprehension scheme. An interesting question is whether the aristotelian should be characterised more as an agnostic about the empty property, rather than a believer in the non-existence of this property. Here I won't dwell on the agnostic characterisation.

[38]To be sure, ACS could have the same form as FCS provided the existential quantifier binding the 'F' is suitably restricted to instantiated properties only, or as the aristotelian would say 'to those that are and not more'. This technique is used in (Cook and Ebert, 2005). However, I decided to sketch the aristotelian conception from a platonist perspective to strengthen the contrast.

4.5.3 Hume's Principle within an aristotelian framework

We need to note the interesting result that the standard proof for the existence of infinitely many numbers on the basis of Hume's Principle does not go through without the existence of the empty property. Intuitively, this is because, within the aristotelian framework, it can't be proven that there are n + 1 numbers less than or equal to n.[39] When embedded in a logic with an aristotelian comprehension scheme, Hume's Principle hardly entails any additional ontological commitments. (Cook, 2003) shows that if there is a non-empty universe, then Hume's Principle only entails the existence of the number 1 (but not 0), otherwise Hume's Principle in an aristotelian setting is satisfiable on the empty domain.

Having this result in place, how does it affect our previous discussion of the presumptuousness constraint? prima facie, it might seem that this constraint does not apply to Hume's Principle, if it is embedded in an aristotelian comprehension scheme. However, we need to be careful here, and take into account the different notions of *further ontological commitment* which may underlie the presumptuousness constraint. Let us discuss each in turn.

Adopting the first conception of further ontological commitment and regarding commitments as critical once they are *too many* in the sense of at least $\aleph_0$-many, then whether or not the universe is empty, Hume's Principle embedded in aristotelian comprehension is *not* presumptuous. It only is presumptuous if the principle is not allowed to have *any* commitments and the universe is not-empty.

On the second conception, whereby a function has further ontological commitments if we are, by accepting its existence, committed to *more* objects in terms of cardinality than previously acknowledged, Hume's Principle is not presumptuous in an aristotelian setting. It never *inflates* the previously acknowledged ontology.[40]

[39] See especially (Shapiro and Weir, 2000), p.168 where this result is further explained.

[40] Again at this stage I leave aside the issue about *further ontological commitment III*. See footnote on

Hence, it seems that Hume's Principle is not presumptuous (apart from one special case) if embedded in an aristotelian framework. And, assuming I correctly reflected Boolos' thoughts, I think that at least prima facie, Boolos himself would not regard Hume's Principle, if embedded in an aristotelian comprehension scheme, as problematic, since it does not involve the existence of *too many objects*. This result is already interesting, since it further elaborates and gives credence to the thought that it is not Hume's Principle itself that entails infinitely many objects; rather additional metaphysical assumptions have to be in place.[41] The following section will introduce a puzzle about the ontological commitments of Hume's Principle based on these formal results which puts pressure on the rejectionist position.

4.5.4 A puzzle about Hume's Principle

Consider the following story:

"Imagine a young upcoming logician/philosopher, well-trained and a committed *epistemic rejectionist*, who thinks that it is impossible to know a priori any principles that involve a *presumptuous* operator. However, our logician/philosopher – call him George – is also an aristotelian, thus he thinks that only properties exist that are instantiated. Well, now let George come across Hume's Principle (as a good logician should); having studied it, he does not have any complaints about Hume's Principle at least none on the grounds of the constraint of *presumptuousness*. Thus, assuming the principle fulfils other constraints which

p.176.

[41]Shapiro and Weir (see their (Shapiro and Weir, 2000) especially p.168ff.) (sometimes) makes it seem as if the empty property is what is responsible for the ontological commitments. They write: "The neo-logicist cannot claim, then, that the existence of infinitely many numbers is demonstrable in an epistemically innocent fashion unless he or she can show that it is demonstrable, in such an epistemically innocent fashion, that there is an uninstantiated property. [...] The 'aristotelian neo-logicist' will, it is true, be able to prove $(\exists F \forall y \neg Fy \rightarrow Inf)$
where Inf is a theorem of infinity; that is, if an empty property exists then there are infinitely many numbers." (op.cit) The following discussion will, I hope, clarify the points raised in their discussion.

are necessary for its success (and it's a priori knowability), George is content to claim to know a priori Hume's Principle.

Now, as it happens George visits the Academy in St. Andrews one day, encounters some philosophers there and by talking to them, he becomes convinced that he should allow the empty property within his metaphysics. That is he changes his metaphysics of properties from an aristotelian conception to a platonist one.

But George has been set up by those philosophers. Having changed his metaphysical framework, they challenge him to explain how his change is compatible with his position as a committed epistemic rejectionist and his previous *a priori* endorsement of Hume's Principle.

The problem thus is that George has previously accepted the existence of the number-function (i.e. the truth of the Ramsey-sentence of HP) while accepting aristotelian metaphysics. But by adopting a different metaphysical conception (whose only relevant difference is that it involves the existence of one more property), the number function picked out by Hume's Principle has apparently become *presumptuous*!"

So what is George to do? Let me outline the above puzzle in more detail and develop it as a challenge to, or rather reductio of, epistemic rejectionism. This is followed by a discussion of different lines of reply to the premisses outlined below.

Premise 1

The existence of the number-function in aristotelian logic is not presumptuous, and thus (assuming it fulfils other criteria) is a priori acceptable.

Premise 2

Changing the underlying metaphysics from aristotelian to non-aristotelian is innocuous.

Conclusion

As the existence of the function, denoted by the number-operator is already accepted on a priori grounds, a rejectionist is committed to allowing the a priori justified claim of the existence of infinitely many objects.

and the underlying and "hidden" major premise of this reasoning is:

Major Premise

The nature of the number-function is independent of the existence of the empty property, i.e. making such justified changes in the metaphysical picture has no bearing on the identity of the number-function.

In the following, I will discuss the different claims of the argument in turn and suggest various replies that can be made to each premise.[42]

4.5.5 Rejecting premise 1

One line of response is to reject the first premise. Here, one might doubt this claim on the grounds that appreciating that Hume's Principle is presumptuous in a platonistic meta-

[42] A further premise is assumed: The opponent might agree with all the premises but deny the conclusion on the basis that the function is not total. In the following I will make the assumption that the function is total.

physics shows that it is presumptuous *tout court.* In both definitions (that is Hume's Principle embedded in aristotelian and embedded in platonist comprehension) the number operator refers to the same function, only that its occurrence in the platonist framework shows that the principle is presumptuous.

This line of reply, however, is unattractive since it has the consequence that even if George was to remain an aristotelian, he has to give up Hume's Principle on the ground of its "potential" presumptuousness. So, it seems that rejecting the first premise on this basis to avoid the puzzle presupposes the thought that the platonist conception has a somehow higher priority over the aristotelian framework in evaluating the ontological commitments of a principle. Consequently, what matters concerning the notion of presumptuousness would have to be rephrased in order to properly reflect the epistemic rejectionist's reasoning, if this strategy were pursued, such that:

> *Strong condition of presumptuousness*
>
> Assuming the existence of an (higher-order) operator is presumptuous, if and only if its acceptance *does* imply *further ontological commitments* on the object level *in a platonist comprehension scheme.*

This change of the condition of presumptuousness is clearly unattractive since it demands that even a committed aristotelian should take into account the possibility of further ontological commitments in the platonist framework – which, one might well have supposed should not matter to him. Therefore, rejecting the first premise is not an attractive strategy.

4.5.6 Rejecting premise 2

The second premise comprises the claim that a change from an aristotelian to a platonistic metaphysics of properties is innocuous, in that it should not be considered as a very substantial move. This claim can be challenged and let us scrutinise what a change of metaphysical framework involves.

What might well be called an 'inconsequential' change in our metaphysical conception would involve merely the quantity of properties that one assumes to exist. On the platonistic conception for example, there exists one more property – the empty property – that doesn't feature in the aristotelian conception. One view then, is that the aristotelian properties are of the same kind as the platonistic ones; the two frameworks just differ in terms of quantity of properties that are out there.[43]

In contrast, a more substantial understanding of the change takes into account the kind of motivation for adopting one conception over the other. As mentioned before, an aristotelian and a platonist differ in *what it takes* for a property to exist. While the former considers properties as concrete universals, existing *in rebus*, the platonist conceives of properties as pure universals existing as abstract entities. This difference might then not only give rise to a different answer to the question of *how many* properties are there, but also might lead one to suppose that the *kinds* of properties are substantially different in the different metaphysical settings. Therefore, since there are different *kinds* of properties involved in the different metaphysical views merely changing the comprehension scheme is not enough to accurately reflect the change: Different *kinds* of properties bring with them a different *content* for principles such as Hume's Principle – and so the sense of the number-operator is substantially different in both cases. On this basis, one can argue that a transition from an aristotelian to a platonistic framework is not innocuous since substantially different ontologies of properties are involved. In turn, this result could be used as a motivation to reject the *Major Premise*, and so argue that different functions are being picked out by the expression "the number

[43]With quantity I not merely mean cardinality as both are committed to (at least) $\aleph_0$ many objects. Rather the platonist has one more specific property that does not exist in an aristotelian setting.

of" within the different metaphysical frameworks – different, since they have different *kinds* of entities as arguments.

So, an important lesson – a lesson that is missing from the discussion of (Shapiro and Weir, 2000) and (Cook, 2003) – is that unless one looks more closely at the views underlying the respective comprehension schemes, many crucial metaphysical issues remain untouched. However, is the possibility of the 'substantial reading' of change enough to defuse the puzzle?

I think the following formal trick can be used to re-instantiate the puzzle, at least in the case of Hume's Principle: Instead of conceiving of the number operator as denoting a function from *properties* to numbers, one could regard it as a function from the Fregean *extensions* of properties to numbers (thereby relying, to be sure on a suitably restricted version of Basic Law V[44]). Presumably, even if one should regard platonist properties as distinct in kind from aristotelian properties, the *extensions* of such properties *do* overlap. To use an example, we may grant that the platonistic property which holds just of Hale and Heck is distinct from the aristotelian property which holds just of Hale and Heck, as the former is *ante rem* while the latter is *in rebus*. Nonetheless, the extension of the former property is identical to the extension of the latter and so the worry about the substantiality of the transition from aristotelian conception of properties to a platonistic conception should not get a grip here. The function involved in this revised version of Hume's Principle is a function from objects to objects and so, the substantiality of a change in one's metaphysics of properties, does not have a bearing on whether this function from objects to objects is the same.

So, let us therefore directly tackle the issue raised in the Major Premise, namely whether the operator (be it the standard version or the above revised version) may be viewed as

[44] As, for example, suggested by (Boolos, 1989) or (Cook, 2006a)

denoting the same function in both aristotelian and platonist comprehension schemes.[45]

4.5.7 Rejecting the Major Premise

The standard conception for the identity-conditions of functions spells matters out extensionally. So, the identity-conditions of a function depend upon the entities that do figure as arguments and values of the function. On the standard view, if two functions have the same inputs but provide different outputs, then the functions are distinct. Thus, the epistemic rejectionist could maintain that the number-function from properties to objects – or, the number-function from extensions of properties to objects – in an aristotelian setting, is a different function from the one embedded in a Platonic setting. This, for the simple reason that Hume's Principle$_{platonist}$ entails the existence of infinitely many objects, while the latter Hume's Principle$_{aristotelian}$ does not. This observation can then motivate the rejection of the *Major Premise* and gives credence to the thought that different entities are denoted by the number-operator occurring in Hume's Principle$_{platonist}$ and Hume's Principle$_{aristotelian}$.

However, one motivation for thinking that the Major Premise holds, is to think of identity-conditions for functions non-extensionally. Think of the operator involved in Hume's Principle in analogy to a "vending machine". Given a certain input (properties) the vending machine (Hume's Principle) gives you a certain output (objects). But the peculiarity about this particular "machine" – the number-function involved in Hume's Principle – is that by giving it the "magic first coin" (the empty property) it not only returns you the "product" (number), but also guarantees that anytime this is repeated with a different "coin", a further new "coin" is issued, which enables you to do the same again and again, so to get infinitely many "products".

[45] If the above formal trick is used the Major Premise has to be suitably revised: The issue becomes whether the identity of the number-function is affected by the (non-) existence of the extension of the empty property.

However, and here the analogy becomes relevant to the issue at hand, no matter what "coins" are given to the machine, whether you start with the magic one or not, the nature and constitution of the machine is the same. So – spelling out this metaphor – if we accept Hume's Principle embedded within aristotelian logic we encounter the same function as in the non-aristotelian case. The coins (properties) themselves and the amount of coins (properties) may change from one metaphysical setting to another but the nature of the machine (Hume's Principle) stays the same. Hume's Principle$_{platonist}$ involves the same function and indeed is the same principle as Hume's Principle$_{aristotelian}$, although different consequences can be deduced in the different metaphysical settings.

Do these considerations help us to see how we should conceive of the dialectic between the epistemic rejectionist on the one hand and the traditional connection (and proponent of abstraction principles) on the other? I will draw together the major points and try to diagnose in what position we now find ourselves.

4.5.8 Diagnosis: Rejectionism rejected?

The "puzzle" about the ontological commitments of Hume's Principle was introduced make the point vivid that Hume's Principle, and other abstraction principles, do not commit to objects all by themselves; rather other components have to be in place. The puzzle was used to put pressure on the the epistemic rejectionist view and the notion of presumptuousness. Hume's Principle embedded in aristotelian logic is not presumptuous so, if it is assumed that the identity of the function involved in Hume's Principle is invariant with respect to a change of the underlying comprehension scheme, the principle should also be non-presumptuous in a platonist frameworks. This conclusion could then be regarded as a reductio of the notion of presumptuousness presupposed by the epistemic rejectionism.

We have seen that there is at least one way to avoid this problematic conclusion for the epistemic rejectionist, namely by denying the Major Premise (i.e. the invariance claim) on the basis of an extensional conception of properties. Here, I won't be able to foreclose this way out for the rejectionist. Thus, a main issue for further discussion remains the question whether the identity of functions should be characterised extensionally or not. If there are independent reasons for a non-extensional treatment, then the force of the puzzle is reinstated and the threat for the viability of epistemic rejectionism will return.[46] Furthermore, we should also appreciate the more general lesson noted above, that focusing exclusively on Hume's Principle per se (and so abstraction principles in general) in the discussion of ontological commitments is at least potentially confusing, at worst wrong! We should always keep in mind the necessary presuppositions (which go beyond mere logical reasoning, but are located within the metaphysics of properties and logic) in order to establish the ontological commitments of an abstraction principle.

Lastly, also worth bringing back into the discussion is a further element that triggered the last main section: the distinction between direct versus indirect ontological commitments, which was meant to draw a difference between the Peano axioms and Hume's Principle. It seems that, if Hume's Principle embedded in an aristotelian metaphysics is – in some non-extensional sense – the same as Hume's Principle in a platonist setting, the distinction between direct and indirect commitments could be made in the following way: Hume's Principle has only indirect ontological commitments, because the same principle has in one setting certain ontological commitments which, in another, it lacks. Consequently, putting forth Hume's Principle as an implicit definition has merely indirect commitments.

[46]If that is the right analysis, it seems that we can put further, at least prima facie, pressure on an extensional treatment of functions. Consider for example the Direction Abstraction: Should it really matter to the identity of the direction-function what lines there are? Should it matter whether the lines are concrete or abstract? I think not, still I will here have to leave a fuller investigation of this next round of discussion to another occasion.

Nonetheless, to what extent this draws a distinction with the Peano axioms is still an open question. It remains to be seen whether a similar trick could be pulled with the Peano axioms. A closer look at the details and assumptions for a proof of infinity from these axioms is needed to evaluate whether the Peano axioms have direct or indirect ontological commitments.

4.6 Conclusions and Outlook

In this chapter we have discussed rejectionism and examined various arguments against this position. I believe *semantic rejectionism* has emerged as both unstable and unmotivated.

Epistemic rejectionism, characterised in terms of the adoption of the additional constraint of *presumptuousness*, proved more resilient. I think that a fair evaluation of the *status belli* is that it has reached a stalemate. While the rejectionist will happily embrace an extensional conception of operators and functions, his opponent will be tempted to accept a non extensional conception. And while the former will accept the fourth dogma of empiricism, the latter will see the case of mathematics as a reason to reject it – another case of "one man's *modus ponens* is another's *modus tollens*"[47]. Despite this, I hope, however, to have clarified, structured and advanced the discussion.

With this, I leave this particular criticism of the traditional connection behind and now turn to explore the question how implicit definitions can help to explain a subject's knowledge of logic or mathematics and thereby address the second epistemic requirement of Benacerraf's dilemma. To this end, I will discuss a proposal put forth by Paul Boghossian, which is in various respects similar to the view of Hale and Wright. I will argue that this conception fails to account for our knowledge of mathematics or logic.

[47](Boolos, 1997), p.308

Chapter 5

Explaining knowledge by means of implicit definitions

Introduction

This chapter[1] is concerned with explaining how a proponent of the traditional connection may aim to account for our a priori knowledge of logic and arithmetic by appeal to implicit definitions. Previously, in chapter 3, we have discussed various constraints that have to be fulfilled by an implicit definition to be either effective, productive or successful. I also noted in 3.7 an important epistemic dimension for implicit definitions: The effectiveness of an implicit definition does, in the good cases, underwrite or give rise to knowledge of it.

This epistemic dimension has yet to be explored and explicated and it seems that two issues are pertinent for a satisfying explanation: First, the fact that some effective definitions are unsuccessful and do not give rise to knowledge raises a type of Bad Company Objection;

[1]The following discussion of Boghossian and parts of the discussion of Hale and Wright's position is based on (Ebert, 2005).

there are cases in which the transition from effectiveness to knowledge fails and so an account is needed to distinguish the "good" cases from the "bad" cases. I will come back to this in chapter 6. The second issue, is the demand for an explanation of how exactly a subject arrives purely in virtue of grasping an implicit definition at a *warranted belief* in the principle in question.

This chapter will focus on one such explanation as originally proposed by Paul Boghossian. Boghossian, another proponent of the traditional connection, offers a template in virtue of which a subject can acquire a warranted belief, and if it is true, knowledge of the implicit definition. I will argue that his model ultimately fails, since his explanation commits a *transmission of warrant failure.*[2] In addition, I argue that the same worry comes up in other models and templates, such as Hale and Wright's. The structure of this chapter will be the following:

In the first part I will focus on Boghossian[3] and provide a brief outline of his account. This is followed by an explanation of what a transmission of warrant-failure consists in and show that this charge is independent of the worry about *rule-circularity* which has been raised concerning the justification of logical principles and of which Boghossian is fully aware.[4] In the third section I will argue for the insufficiency of Boghossian's template as an explanation of how we might acquire a warrant for logical principles. I will show however that the insufficiency of his template can be remedied by adopting a Disquotational Step. In the fourth section, I will then argue that incorporating this necessary further step makes his template subject to a transmission of warrant failure, assuming that certain rather

[2]Here I focus on his discussion of logical knowledge; however I contend that the same worries would apply if his model were applied to mathematical knowledge.

[3]See his (Boghossian, 1996), (Boghossian, 1997) and more recently (Boghossian, 2000), (Boghossian, 2001) and (Boghossian, 2003).

[4]Boghossian actually uses rule-circularity as a *genuine and legitimate type* of justification. However, this claim won't be discussed here.

basic and individually well-motivated principles hold. Thus, Boghossian's account faces a dilemma: Either he adopts the Disquotational Step and subjects his account to the charge of a transmission of warrant-failure, or he drops this additional step leaving the account vulnerable to the gap that has previously been highlighted. The fifth section discusses various rejoinders that Boghossian might adopt but none of which – I will argue – can resolve the dilemma. Lastly, I will raise and discuss the question how this worry generalises to other accounts. Here, I will focus on other proponents of the traditional connection – specifially Hale and Wright[5] – to explain our knowledge of logic and/or mathematics in virtue of implicit definitions. I will argue that their account, naturally interpreted, is also affected by essentially the same transmission of warrant-failure.

5.1 Boghossian's template for a priori knowledge of logical principles

Boghossian's explanation of our a priori knowledge of logical principles has two steps. The first is based on assuming what might be called *inferentialism*, the thesis that genuine meaning-constituting rules provide an *entitlement* to reason according to such rules. Inferentialism is thus a version of the traditional connection which focuses on the rule formulation of implicit definitions rather than whole statements. However, Boghossian also notes that by itself inferentialism does not yet *explain* our belief that certain rules are valid. On his view the subject does not need to *know* that a certain rule is meaning-constituting in order to be entitled to infer according to it. Therefore a second step is required in order to explain our *knowledge* of certain presumably a priori principles. We need to explain how we can be

[5] I won't have the space however to discuss other similar positions as for example in (Peacocke, 1999) and (Peacocke, 2000)

warranted not just in using principles in inferences, but in believing that those principles are valid or true; this is explained in virtue of Boghossian's *template.*

In his earlier paper "Analyticity"[6], Boghossian suggests that we can explain how to gain a "warrant for our belief in the elementary truths of logic" (op.cit., p. 358) by the following (slightly adjusted) template:[7]

Boghossian's template

1. If C is to mean what it does, then A has to be valid
2. C means what it does
3. A is valid

This template is a metalinguistic construction. The first premise is what Boghossian considers to be an implicit definition of C, where C should be regarded as a name for a term, such as "and" and A as a name for the meaning-constituting inferences for "and", such as its Introduction and Elimination rules (I-E rules). The second premise is the antecedent of the conditional of the first premise. It is phrased slightly confusingly and I suggest that it be understood such that "C means what it does" is roughly equivalent to "C refers as intended by the implicit definition" or "C refers as intended" for short.[8] I will discuss this ambiguity further in due course.

[6]See (Boghossian, 1997) p.357 and p.359

[7]Boghossian's original formulation is more cumbersome, nevertheless the adjusted version conveys the same idea in a simplified formulation.

[8]In conversation Boghossian has agreed that the notion of meaning as used in his template is closer to reference than to Fregean sense. Also note that I will neglect the minor difference between premise (1) and (2) in that (1) says "[...] *is to* mean what it does" and (2) reads "[...] means what it does". I will assume that both occurrences are intended to be used interchangeably to warrant the modus ponens inference.

Various problems have been highlighted concerning Boghossian's account.[9] The following criticism is independent of points previously raised in the literature. Before outlining my criticism let me quickly explain the basic notions I will use in my criticism such as 'warrants', 'transmission of warrants' and 'transmission of warrant-failure', and the contrasting idea of a 'rule-circular justification'.

5.2 Warrants, Transmission of warrant(-failure) and Rule-circular justification.

I will be brief in my explication of the notion of a "warrant" and only later introduce further complication in the context of possible rejoinders for Boghossian. For the moment it will be enough to regard "warrant" – a notion hardly explained in Boghossian's article – as a placeholder for that which turns a true belief into knowledge. The notion of *transmission of warrant* also needs clarification and can be characterised by contrasting it to the notion of *closure of warrant*.[10] The latter notion merely claims that if there are warrants for the premises of a valid argument, then there is a warrant for the conclusion too. The notion of *transmission of warrant* is stronger in that it provides a stricter connection between the warrants for the premises and the resulting warrant for the conclusion. It imposes requirements on the thinker acquiring a warrant for the conclusion *in virtue* of the premises being warranted. So, the idea is that a warrant is transmitted if a thinker who acquires a warrant for the premises of a valid argument, and knows that it is valid and *thereby* – i.e. in virtue of both components – acquires a warrant for the truth of the conclusion.

[9]See for example (Margolis and Laurence, 2001) and (Wright, 2001) and (Williamson, 2003) for raising various specific worries with Boghossian's proposal. Also (Horwich, 1997) who we discussed in the previous chapter, discusses critically Boghossian's model.

[10]This distinction was previously drawn in (Wright, 2002). I follow closely this discussion. The notion of closure of warrant goes back to (Dretske, 1970).

We also need to clarify the notion of a *transmission of warrant-failure.* There are different ways in which such failure might occur. One type, which will be important to my argument, can be characterised as a form of *begging the question.*[11] The basic idea is that having a warrant for the truth of the premises, already presupposes having acquired a warrant for the truth of the conclusion. As a result, there can't be transmission of warrant for the truth of the conclusion from those premises, since such warrant already has to be acquired in order to acquire a warrant for the premises. So, an argument can be sound yet it might fail to transmit a warrant from the premisses to the conclusion. Thereby, it is rendered unsuitable to underwrite a *warranted belief* in the conclusion on the basis of its premises.

It is important to note that this charge is independent of the problem already highlighted by Quine and others[12] that to justify some form of logical inference results in a justificatory circle – another form of begging the question – as we already need to presuppose the logical *inference* in question in order to appreciate the validity of the argument which aims to provide a justification for its validity. Boghossian is aware of the issue of *rule-circular justification*, but he aims to provide an account in which we can regard this rule-circularity as providing a genuine justification.[13] Let me briefly clarify the notion of a rule-circular justification, highlight some differences and then move on to my argument against Boghossian's account.

A canonical statement of the rule-circularity worry is that an argument is unpersuasive if the conclusion of an argument establishes the validity of an inference rule, whose validity is already presupposed in reasoning from the premises to this very conclusion. To wit, both worries (transmission of warrant-failure and rule circularity) have to do with the persuasiveness or conclusiveness of an argument, but I think there are various considerations that show

[11]Note that the case of *begging the question* is not a counterexample for the closure of warrant-thesis.

[12]See for example (Quine, 1935) and (Quine, 1954)

[13]As noted above, his positive account won't be subject of this paper, nevertheless I will later discuss and reject a similar reply Boghossian provides against the rule-circular justification charge to my transmission of warrant-failure charge. More on this in section 5.5.2.

that the worry of rule circularity is distinct from the charge of a transmission of warrant failure.

Rule circularity applies to the justification of inferences, where an inference is used in the argument and is mentioned (or schematised) in the conclusion. Reconsider the above example with the connective "and" and its meaning-constituting rules in the template. The conclusion is about the validity of the "and"-rules and the reasoning merely uses modus ponens, so there is no relevant charge of rule-circularity. The transmission of warrant-failure worry is in this respect more general, as, if it is correct, will show that the template is wrong for any newly introduced expression. Furthermore, the charge also applies to statements that might be justified in virtue of the above reasoning which are not inferences. Consider, for example an extension of Boghossian's account to mathematical principles such as Hume's Principle, which also fails, if I am right, due to a transmission of warrant failure.[14]

There are further differences, but I won't dwell on them here and instead present the main thrust of this chapter. In the next section I will further investigate Boghossian's template as I think it suffers from serious insufficiencies. Important amendments have to be made in order to regard it as being able to at least *prima facie* provide a "warrant for our belief in the elementary truths of logic." (op.cit.) But having these amendments in place will be to commit it to a transmission-failure, as I will show in section 5.4.

5.3 Step 1: The insufficiency of Boghossian's template

A key issue concerns the elusive and inherently ambiguous "is to mean what it does" of the first premise or "means what it does" of the second premise, which presumably are synonymous. Still, there are various ways to interpret these phrases. One option is to deflate

[14]Disregarding for a moment that a biconditional such as Hume's Principle might be rephrased as Introduction and Elimination rules, as noted on p. 118f.

this notion and accept that even an empty term means what it does, namely we can give some descriptive content to it. So consider again the example of the expression "the extension of" as introduced by the inconsistent axiom Basic Law V. Despite the axiom's inconsistency one can contend that on the deflated version of the second premise the term 'extension' means what it does, in that we can explain the notion with appeal to co-extensiveness. Thus, on this reading, the second premise of Boghossian's template is close to trivial, and most of the weight is put on the first premise, which becomes rather substantial: Why should the mere having of descriptive content be sufficient to figure in valid patterns of inference? Just consider Basic Law V again: even if it is effective – as argued for in chapter 3 – it is surely neither consistent nor valid (if regarded as a pair of I-E rules).[15]

Alternatively, we can – as I will do for the rest of this chapter – interpret the above phrases as equivalent to "refers as intended". This now renders the *first* premise close to trivially true, for what it claims is that, if a term has the intended reference in order to make a pattern of meaning-constituting rules valid, then this pattern will be valid. On a material conditional interpretation this is a necessary truth. But, the second premise now becomes rather substantial, as it claims that a term refers as intended. On what grounds can we rule out that it is not an empty term? It is this issue that will become prominent in my later discussion. The main point will be that in order to appreciate or to have a warrant that the term involved refers as intended, we already have to appreciate, or have to have a warrant for believing, the conclusion of the reasoning of the template, i.e. that the meaning-constituting patterns are valid. The core of my charge consists of two key thoughts that highlight the shortcomings of Boghossian's account. Firstly, the template as outlined by Boghossian has to guarantee the knowledge of the validity of a pattern of inference as an *object-language*

[15]This important insight gave rise to the distinction between effectiveness, success and productiveness which informs this discussion.

statement and not merely as a metatheoretical statement – a claim that I will substantiate in this section. Secondly, to grasp the (object-language) content of the second premise – that is, the intended reference of the term – an understanding of the term's role in whole sentences such as the conclusion is a prerequisite. Thus, in order to be in a position to know that the term has its intended reference, we need to know that the conclusion has the meaning to make the pattern of meaning-constituting inferences valid. This will eventually lead to a transmission of warrant failure of Boghossian's template – the details of this claim will be expounded in section (5.4).

To develop my worries culminating in the claim that Boghossian's account needs further amendments let us look closer at the conclusion of the template, as cited it says:

Boghossian's template

1. If C is to mean what it does, then A has to be valid
2. C means what it does
3. A is valid

As already mentioned C and A are names for connectives and meaning-constituting rules, and thus the conclusion is merely a meta-linguistic statement. This however raises an immediate worry, namely, how does the meta-linguistic statement concerning the validity of the rule guarantee that the subject has grasped the *content* of the rule in question? That is granting that the template provides a warrant for the validity of A for the subject, it is not secured that the subject has also *understood* the rule whose validity he has a warrant for. The meaning-constituting rules for "and" or perhaps some more complex connectives might be written in Chinese and, as it happens, my Chinese friend tells me that premise one and two are true without telling me which connective he is talking about. In that case, I would be able to assent to the conclusion and agree that such-and-such signs represent valid rules of inference, without *grasping the content* of the rules in question. So it is questionable – to say the least – that the template as stated is sufficient to provide the *right type* of warranted belief in the rules/principle.

Therefore, I believe Boghossian's template needs to be adjusted in order to account for such a case (the Chinese friend example). What seems necessary is to add a further step which I will label the Disquotational Step.[16] The idea is that we can account for the kind of *identifying knowledge* that is needed in explaining our knowledge of logic or the rules in question, provided there is a warrant to disquote on what Boghossian took to be the conclusion – that is (3) above. Thus, I will adjust the above template by adding a further step from the initial conclusion (3) to the new conclusion (4), i.e. the object-language counterpart of (3). For purposes of presentation and clarity I will take "and" for A and "P and Q $\Rightarrow$ P" for C, while also omitting the second elimination rule for simplicity.

The revised template

1. If "and" is to mean what it does, then "P and Q $\Rightarrow$ P" has to be valid

2. "and" means what it does

3. "P and Q $\Rightarrow$ P" is valid

[16]This step is not strictly speaking an instance of disquotation. Disquotation normally concerns the truth of a statement and not the validity of an inference rule as in the above example. One option to bring my use of disquotation closer to the standard usage is to reinterpret the Boghossian's template in terms of statements of a rule and consequently to use truth instead of validity and so make the additional step a genuine instance of disquotation. However, I choose to remain as close as possible to Boghossian;s initial characterisation of the template and merely label the additional step, which I argues is necessary, a 'Disquotation Step' for lack of a better name. See also the following footnote for further discussion.

4. P and Q $\Rightarrow$ P[17]

Adding the last step forecloses on the worry about knowledge by testimony, but it raises the question how we can justify the move from (3) to (4). As we have seen above, the premises as so far understood are insufficient to provide a guarantee for the Disquotational Step (remember the Chinese friend counterexample). Thus we have to examine how we can warrant this move and I will argue that, based on other assumptions I will outline below, this will lead to a transmission of warrant failure for the revised template.

5.4 Step 2: Transmission of warrant failure of the revised template

If the stronger conclusion (4) is needed to establish a warranted belief in the validity of the principle in question, how can we justifiably move to this conclusion? What is it that warrants the application of the Disquotation Step – moving from premise (3) to the conclusion – and thus disquote? The worry I will raise in the following is an instance of the transmission of warrant-failure charge. Namely, that under a suitable reading of (2) and (3) we need, in order to justify the Disquotational Step, a prior warrant for (4). This claim is

[17]As noted above, in order to turn the step from (3) to (4) into a genuine disquotation move we could further adjust the template such that we are not talking about the validity of I-E rules ($\Rightarrow$ is meant to represent an inference), but rather the truth of the statement of the rule. Accordingly, (3) would then be represented as 'P and Q $\models$ P' is *true*. I contend however that nothing hinges on these variations since my argument concerns a *structural difficulty* with Boghossian's template: What is needed is merely a way to represent the step from being warranted in believing that conjunction elimination is valid (or that the statement of conjunction elimination is true) to being warranted in believing in believing conjunction elimination, and (DS) is meant to represent just this move. If, as I argued, this move is necessary for a subject being warranted believing (and grasping the content of) such inference, the following section will outline a transmission of warrant-failure for Boghossian's position. If the step is rejected then I find it hard to see how the template could suffice to explain a subject's warranted belief in, and thereby grasp of the truths of logic. Another problem worth noting here is that the new conclusion still involves schematic letters and is not strictly speaking an object-language statement – however, I won't pursue this problem any further (this would just be another problem for Boghossian's account).

based on four assumptions, which individually are either implicitly held by Boghossian or can be considered as uncontroversial claims:

A1: Being warranted to disquote a sentence requires an understanding of that sentence

The idea is that in order to be *warranted* in making the Disquotational Step it is necessary to *understand* the sentence being disquoted – this much seems to me a *platitude.* And, in order to have such understanding one has to possess, or rather grasp, the concepts that are involved in the statement. This assumption is introduced so to avoid knowledge by testimony and thereby to account for the justification to use the Disquotation Step. The following assumption is rather distinct and more general.

A2: Sentences as basic units of understanding

Basically the thought here is that the basic semantic unit for our understanding are sentences and hence in order to appreciate what reference a term has one has to understand the role this term plays in various sentences. This assumption is close to Evan's *Generality constraint* and we can assume that *prima facie* it is accepted by Boghossian.[18]

A3: Understanding involves knowledge of meaning

Again I consider this principle to be benign and uncontroversial. Thus, in the following I won't assume a specific model of what understanding involves (knowledge of truth-conditions perhaps), rather the minimal conditional that – in general – to understand a sentence is to have knowledge of the meaning of the sentence.

A4: Epistemic notion of analyticity

[18]We might also regard this as part of the context principle thesis, but as the context principle as I outlined it concerns various other theses as well I refrain from labelling it the context principle.

Lastly (A4) is Boghossian's very own conception of *epistemic analyticity.*[19] More explicitly he claims that a statement S is *epistemically analytic* for a person T, if "the mere grasp of S's meaning by T sufficed for T's being justified in holding S true".[20] Thus Boghossian's conception of analyticity, slightly rephrased, involves that knowing the meaning of an analytic statement S, implies having a warrant for the truth of S. This, it seems to me, is the component of Boghossian's view that makes him a proponent of the traditional connection. It is grasping certain effective principles that suffices for a subject to acquire a warrant to hold true and so *know* the principles in question.[21]

With these assumptions in place let me outline my charge of a *transmission of warrant failure* in the revised template. Let us ask what it is to have a *warrant* for the Disquotational Step. The following argument will show that a necessary condition for such a warrant is that one has to have a prerequisite warrant for the conclusion.

The argument is as follows: Based on (A1), to have a warrant for making the Disquotational Step one has to have an understanding of (3). This means that – in order to avoid the Chinese friend example – the subject has to have a grasp of the content of (3). In order for a subject to have such grasp, she has to have a suitable understanding of premise (2). Namely, the subject has to grasp what the intended referent of A (in our example "and") is. Yet, due to an application of (A2), sentences are the basic unit of understanding, and so in order to grasp the intended reference of A, the subject has to understand the role such term plays in appropriate sentences; that is, it is a *prerequisite* to disquote to have a grasp of the statement such as the conclusion in Boghossian's argument. But, in order to have an understanding of the conclusion involves by (A3) knowledge of the meaning of this statement. However,

[19]This notion has been attacked in (Margolis and Laurence, 2001) however I won't quarrel with it here.
[20](Boghossian, 1997), p. 334
[21]In addition, Boghossian clearly distances himself from superficially similar view, such as (BonJour, 1998) who appeals to some form of rational insight.

and this is the crux of the argument, the logical principles in question are, for Boghossian, to be regarded as epistemically analytic. And under Boghossian's conception of epistemic analyticity grasping the meaning of a sentence provides for a warrant for the truth of such sentences. Since, grasping the meaning is a prerequisite to be in a position to disquote, the warrant for the truth of the conclusion *precedes* the warrant to disquote. Being in a position to disquote *already* involves having a warrant for the truth of the conclusion.[22]

Therefore, we can confront a proponent of Boghossian's account with the following dilemma:[23] Either his template is inherently insufficient to account for our knowledge of the intended principles since it cannot underwrite disquotation of the initial conclusion (3), or he faces, based on the tenability of the above assumption, a transmission of warrant-failure. The revised template and his thesis of epistemic analyticity are incompatible. The incompatibility arises because his template requires, that understanding a sentence is distinct from knowing or being warranted in believing that the sentence/rule is true/valid. Otherwise the template can't transmit a warrant for the conclusion. However, Boghossian's own conception of epistemic analyticity is such that a mere understanding of the sentence suffices for knowledge that the sentence/rule to be true/valid. This is the basic tension in Boghossian's account.

[22]The argument can be represented using the following abbreviations:
DS stands for Disquotational Step; K(α) for knowledge of α; U(α) for understanding α; M(α) for meaning of α; W(α) for warrant/being justified for α and T(α) for truth of α.

Step 1
W(*DS on* 3^{rd} *premise*)$\Rightarrow$ U(2^{nd} *premise*) - due to **A1**
Step 2
U(2^{nd} *premise*) $\Rightarrow$ U(*conclusion*) - due to **A2**)
Step 3
U(*conclusion*) $\Rightarrow$ K(M(*conclusion*) - due to **A3**
Step 4
K(M(*conclusion*)) $\Rightarrow$ W(T(*conclusion*) - due to **A4**

[23]Note that various replies to this problem and further clarification of the argument will be offered in the following section.

5.5 Various rejoinders to my criticism

In this section I will discuss three possible replies on behalf of Boghossian to my criticism. The first discusses the application of the assumption that the basic units of understanding are sentences. The second takes up a distinction between internal and external warrants and develops a reply along externalist lines that Boghossian seems likely to adopt. The third draws a distinction within the notion of *acquiring a warrant* on the basis of which a further response can be developed. I will argue that none of these lines of response will undermine my argument.

5.5.1 Assumption (2) revisited

A possible reply against the argument is that the application of the assumption that the basic units of understanding are sentences, does not guarantee that in order to understand the second premise one has to have an understanding of the conclusion. Rather, granting that this assumption correctly applies here, it only demands that one has to understand *some* sentences in which the newly introduced term occurs; this does not necessarily imply that it has to be the conclusion of the template.

Although I think this is a correct observation I don't believe it has much force. The conclusion of the argument is a disquoted version of what is supposed to be a meaning-constituting rule for the new term in question. Hence, to grasp the reference of the term in question, which implies (at least) knowing what role the term plays in whole sentences (A2), one must grasp what role the term plays in its own meaning constituting rules, i.e. those sentences that are meant to fix the reference of the term. The fact that the conclusion is concerned with meaning-constituting inferences for the term in question surely implies that grasping the reference of the term already involves a grasp of the conclusion.

But to reinforce the argument note the following two points: Firstly, if the role of the assumption that the basic unit of understanding are sentences is accepted and one has to appeal to *some* sentences to grasp the reference of the new term, while excluding the conclusion to avoid my argument, then this implies the following insufficiencies: The template, by itself, is not sufficient to account for a subject's knowledge of the logical principle in question. Rather appeal has to be made to other sentences not mentioned in the template in order to explain the acquisition of knowledge of that principle. Furthermore, remember that the target of putting forth the revised template is to have knowledge of the conclusion. To argue that the conclusion isn't part of what needs to be understood threatens the very idea that the conclusion can be known, since knowledge of it surely presupposes understanding!

Secondly, to focus the discussion we might imagine a scenario in which a character, who barely has a grasp of language and who has no grasp of any logical connectives, is meant to be introduced to these new connectives in virtue of the template. This is a case in which Boghossian's template should be able to generate a warrant for the logical principle for the character in question (Hero). The crucial point in this scenario however, is that Hero only knows of the meaning-constituting rules for the connectives (on the basis of premise (1)) and thus, if (A2) is in place, *has* to appeal to the conclusion of the argument, as *per impossibile* there are no other occurrences of the term he can refer to. Hence, I think that exploiting the admitted 'structural' weakness of my argument won't establish a serious line of reply to the charge of transmission of warrant-failure.

5.5.2 Internal vs External warrants

In his papers on rule-circular justification of logical inference, Boghossian could be interpreted as adopting a type of externalist position. He holds that rule-circularity can underwrite a warrant for knowledge of logical principles![24] Similarly, it might be suggested that an appeal to external warrants might help to avoid a transmission of warrant-failure of the template.

Let us therefore distinguish two different statuses a warrant might have. A warrant is *internal* if the warrant that underlies a certain belief (that a certain inference is valid for example) can be reflectively accessed by the subject in virtue of a priori reasoning, self-knowledge or reflection. A warrant is *external* if the warrant that underlies a certain belief is not so accessible and thus cannot be claimed through a priori reasoning, self-knowledge or reflection.

If that is the correct distinction to draw, then it should be clear that the internalism-externalism debate is here of no relevance. Firstly, in my discussion of a transmission of warrant-failure of the template, I have nowhere relied upon a *specific* notion of warrant, rather I contend that my worry is applicable to any type of warrant, internal or external.

Secondly, I think there is an awkwardness in interpreting Boghossian as invoking only external warrants in the context of logic. This is because, he provides a template to explain how a subject could *on the basis of reasoning* through the template acquire a warrant for a logical principle. Surely, a warrant acquired in this way has to be internal in the sense that it can be claimed by a priori reasoning or reflection itself; otherwise it is unclear what the point of his template is. Furthermore, consider that Boghossian appeals to a notion of *epistemic analyticity* where grasping the meaning of terms suffices to be warranted in holding true certain statements. It seems rather intuitive also in this case that the *type* of warrant can't just be purely external – that is not claimable by reflection or self-knowledge. To invoke external warrants here would seem to imply that grasping the meaning of the term or understanding a certain expression is to be spelled out externally, which (at least intuitively) goes against the initial motivation for the notion of epistemic analyticity. Hence,

[24] See for example his (Boghossian, 2000), p.250ff.

I have strong reservations in interpreting Boghossian's position in a strict externalist fashion. So, I think there is no prospect in avoiding the transmission of warrant-failure charge by appeal to the distinction between internal and external warrants.

5.5.3 Two notions of "acquiring a warrant"

Here I want to suggest a last rejoinder by drawing a different distinction within our notion of warrant. It is not concerned with internal and external warrants, rather it has to do with the relation a warrant (internal or external) can have to a subject. Thus, we might distinguish two different statuses for a warrant to count as 'being acquired'.

On the one hand a thinker might be in *possession* of a specific warrant for a belief. That is – metaphorically speaking – the subject has the warrant within his epistemic field. In contrast a thinker might have a specific warrant merely *available* without actually possessing it. That is, to use the above metaphor, the subject has the warrant close to his epistemic field ready to be possessed without actually possessing it. Although this distinction might not be as clear-cut as one would like, the intuitive idea, I hope, is clear enough and, as we will see, it will prove useful for further discussion.[25]

There are various ways in which Boghossian could apply this distinction. Here I will discuss the most obvious option. He might argue that the notion of epistemic analyticity should be understood in such a way that instead of maintaining that a statement S is epistemically analytic for a person T, if "the mere grasp of S's meaning by T suffices for T's being justified in holding S true"[26], we should say that "the mere grasp of S's meaning by T suffices for T's having *available* a warrant in holding S true", without necessarily *possessing* it. This alteration might be used to forestall the charge of a transmission of warrant-failure

[25]Williamson in his (Williamson, 2000) draws this distinction with respect to knowledge – furthermore a similar distinction is often made in the context of self-knowledge.

[26](Boghossian, 1997), p. 334

in the following way. In order to make the last step in the above argument I need to assume that what the notion of epistemic analyticity provides is the *possession* of a warrant, which would then lead to a transmission of (possessed) warrant-failure, as to possess a warrant for the second premise involves already possessing a warrant for the conclusion. But – so the line of response might go – having acknowledged the above distinction, the notion of epistemic analyticity involved is weaker in that it only makes *available* a warrant for the conclusion. And, this does not suffice for a genuine transmission of warrant-failure, since what we end up with is that the possession of a warrant for the premises will involve the availability of a warrant for the conclusion. Indeed this result, by itself, can't be sufficient to reject the template, because for *every* valid argument it will be the case that somehow the possession of a warrant for the premises will involve the availability of a warrant for the conclusion – otherwise how could there be any transmission of warrant in the first place?

However, I don't think that this response will yield the right result for Boghossian. The suggested revision of the notion of epistemic analyticity would weaken it to such an extent that the conception of epistemic analyticity would virtually do no work in Boghossian's intended account of how a subject can come to know or be warranted in believing a logical principle. Recall that the main motivation for the notion of epistemic analyticity is that a grasp of meaning plays a substantial role in being justified in holding certain sentences to be true. Now, on the weakened version of epistemic analyticity, understanding can't be appealed to in order to provide actual *possession* of a warrant for holding a statement true. Understanding merely makes a warrant available to a subject, which however, is not sufficient as a *genuine* justification for holding true certain statements. Thus, on this weakened version of epistemic analyticity a further story, over and above the template and the idea of epistemic analyticity, has to be told in order to explain how a subject comes to *possess* a warrant for a logical principle. Yet, it was the aim to account for a subject's warranted belief in a logical principle solely by means of the template and the notion of epistemic analyticity.

So to conclude, I believe to have shown that Boghossian's conception faces genuine difficulties which can't be remedied in any obvious way, as my discussion of the various possible rejoinders showed. The question that remains to be discussed here is whether these problems are specific to Boghossian's account or whether they can be extended to other similar conceptions that aim to found our knowledge of logic and arithmetic on implicit definitions. If that were the case, the worry outlined would constitute a general obstacle to the very idea underlying the traditional connection.

In the following section, therefore, I will discuss whether other proponents of the traditional connection – Hale and Wright – are also affected by the transmission of warrant-failure charge. On the basis of a recent paper by Hale and Wright, I will outline another template to explain the acquisition of a priori knowledge and argue that the means to generate the same problem for their view are available.

5.6 Extending the argument to Hale and Wright's conception

In this section I will suggest an understanding of Hale and Wright's conception of how implicit definitions can found a priori knowledge, that is similar in spirit to Boghossian's view. I will argue that my reading of their proposal also falls foul of a transmission of warrant-failure.

There are two important features in Boghossian's conception that generate the general worry of an epistemic circularity. First, his attempt to recover basic a priori knowledge starts out with meta-linguistic premises. In order to arrive at the object-language conclusion, reasoning along a template is required. So, the first feature of his view is that acquiring a priori knowledge involves a transition from the meta-language statement of the implicit definition to its object-language version.

The second feature is Boghossian's conception of epistemic analyticity: The idea that understanding a sentence provides a warrant to hold the statement true. The transmission of warrant-failure of the template arises because in order to be justified to make the transition from the meta-language statement to its object-language version – i.e. to warrantedly disquote – already involves an understanding of the statement in question. However, the type of understanding that is needed to disquote will *not precede* the kind of knowledge that is aimed for in that conclusion, since the notion of epistemic analyticity draws together an understanding of a sentence with knowledge of it. As a result, reasoning in accordance with the template can't transmit a warrant from the premisses to its conclusion.

Let us now investigate in what respect Hale and Wright's proposal shares these two general features. In their important paper on this topic, entitled "Implicit definitions and the a priori"[27], Hale and Wright offer an explanation of a priori knowledge which, like Boghossian's, attempts to recover a priori knowledge by starting from meta-linguistic premises involving the vehicle of the implicit definition. In contrast to Boghossian, they do not provide a specific template to mark a transition; however they offer the following explanation of how a subject can acquire a priori knowledge from meta-linguistic premises involving an implicit definition. They write:

> "How, just by stipulating that a certain sentence '#f' is true - where '#_' is already understood, and 'f' is a hitherto contentless expression [...] – is it supposed to be possible to arrive at an a priori justified belief that #f? [...] Well, the route seems relatively clear *provided* that two points are granted: first that a stipulation of the truth of the particular '#f' is so much as properly possible *[1.]*

[27](Hale and Wright, 2000)

> [...] and second that the stipulation somehow determines a meaning for 'f'. *[2.]* [...] Moreover if the stipulation has the effect that 'f' and hence '#f' are fully *understood* [...] then nothing will stand in the way in the way of an intelligent disquotation *[3.]*: the knowledge that '#f' is true will extend to knowledge that #f. In other words: to know both that a meaning is indeed determined by an implicit definition and what that meaning it is, ought to to suffice for a priori knowledge of the proposition thereby expressed. ((Hale and Wright, 2000), p. 296)

If the intention of this passage is to explain how a subject can acquire, by following the reasoning alluded to in that quotation, a priori knowledge by means of an implicit definition, then we should also be able to extract a general pattern of inference on their behalf. Let me suggest the following template, and also assume that for the statement in question, it is "so much as properly possible to stipulate" it.

1. The stipulation that '#f' is true is successful

2. If '#f' is true then it determines a meaning for 'f'

3. If the stipulation that '#f' is true has the effect that 'f' and '#f' are fully understood then the subject is warranted to disquote on '#f'

4. #f

What is clear from this template, is that Hale and Wright are aware of the understanding problem (the Chinese-friend example) and so evade the first horn of the dilemma, which I discussed in 5.3. Moreover, there is a difference in the details between this reconstructed version of Hale and Wright's template and Boghossian's template. Reconsider Boghossian's first premise which states – rephrased in the Hale and Wright terminology: 'If 'f' has a determined meaning ['means what it does' or 'refers as intended'], then '#f' is true', while premise 2 in the suggested Hale and Wright template states that: 'If '#f' is true then it determines a meaning for 'f''. So, Hale and Wright use the converse of the conditional adopted by Boghossian. Here, however, I won't be further concerned with this difference since I don't think that it constitutes a difference that will matter with respect to the looming transmission of warrant-failure.

Also, note that there is no direct link between premise 2 and premise 3 in this reconstructed template. Premise 2 somehow seems to be idle since its consequent is nowhere used in the later stages of the template. Therefore, it seems that maybe a more suitable interpretation should use a slightly rephrased third premise – call it 3*:

> 3*. If the stipulation that '#f' is true and 'f' has a determined meaning, then 'f' and '#f' are fully understood.

Since premise 1 states that the stipulation that '#f' is true is successful, the first part of the antecedent of premise 3* is redundant and thus we can focus on the following simplified conditional:

> 3**. If 'f' has a determined meaning [refers as intended], then 'f' and '#f' are fully understood.

As a result the suggested final template will be:

1. The stipulation that '#f' is true is successful.
2. If '#f' is true then it determines a meaning for 'f'.
3. If 'f' has a determined meaning [refers as intended], then 'f' and '#f' are fully understood.

4. #f

If, the intention of Hale and Wright is similar to Boghossian and the quoted passage is meant to offer a pattern of reasoning a subject has to go through in order to arrive a priori knowledge of the implicit definition, then I think the reconstructed template seems adequate to represent this idea.[28] Moreover, I believe that on this reading, Hale and Wright clearly accept the first feature of Boghossian's conception and represent the acquisition of a priori knowledge by means of a transition from the meta-linguistic version of the vehicle of the implicit definition to its object-language version.

Let us now consider whether Hale and Wright also share the second feature of Boghossian's account and adopt a similar conception of epistemic analyticity. Although Hale and Wright nowhere *explicitly* use the notion of epistemic analyticity, they have appealed to the notion of 'recarving contents'. The underlying idea here is in important respects similar to that of the conception of epistemic analyticity.[29]

Recarving of content concerns a feature of abstraction principles, and the idea can be traced back to Frege, who, in *Grundlagen*, §64 writes:

> The judgment 'line a is parallel to line b', or, using symbols,
>
> a // b
>
> can be taken as an identity. If we do this, we obtain the concept of direction, and say: 'the direction of line a is identical with the direction of line b'. Thus we replace the symbol // by the more generic symbol =, through removing what is specific in the content of the former and dividing it between a and b. We carve

[28]I highlight this assumption since I suspect that Hale and Wright disagree that the quoted passages should be understood in this way.

[29]There has been extensive discussion of the notion of recarving of content. See (Hale, 1997) and (Hale, 1999) for a defence, (Potter and Smiley, 2001) for extensive criticism and (Hale, 2001) for a reply. Also, see (Stirton, 2000).

> up the content in a way different from the original way, and this yields us a new concept.[30]

In this passage, Frege uses the idea of recarving of contents in order to explain how an abstraction principle can generate a new content and thereby introduce a new concept. In my terminology, Frege explains why certain abstraction principles can be legitimately regarded as *effective.* Yet, Hale especially, takes the idea of recarving of content to do more. Without going into too much details under what conditions a recarving of content takes place, Hale seems to use this idea to explain how a subject can be justified in believing that Hume's Principle is correct and so to move legitimately from the effectiveness of an abstraction principle to knowledge of it.

His proposal is that the idea of recarving of contents can be used to discern that two sentences have the same content or rather truth-conditions. He writes, as a final proposal for his notion of recarving of contents:

> Two sentences have the same truth-conditions iff anyone who understands both of them can tell, without determining their truth-values individually, and by reasoning involving only compact entailments, that they have the same truth-value. ((Hale, 1997), p. 102)

Assuming that reasoning by "compact entailment" suffices to understand Hume's Principle – whatever the details of this additional requirement are – this principle will coincide in truth-conditions just in this sense: The right-hand side and the left-hand side of Hume's Principle will be such that anyone *who understands both of them can tell* – with certain provisos in place – that they have the same truth-value. So, a subject, having *understood* both sentences, thereby arrives at *knowledge* of the biconditional.

[30] In a similar fashion Hume's Principle carves up content in a different way and so yields the concept of number.

In comparison, reconsider the underlying idea of Boghossian's notion of epistemic analyticity: a statement S is *epistemically analytic* for a person T, if "the mere grasp of S's meaning by T sufficed for T's being justified in holding S true".[31] In analogy, we can now suggest an interpretation of Hale's notion of recarving of content following the same schema: A biconditional S is an instance of *recarving of content* for a person T, if an understanding (which includes certain provisos) of both sides of S by T suffices for T's being justified that they have the same truth-value and that consequently T is justified in holding S true.

Here, I won't claim that Hale and Wright are committed to exactly this formulation; rather I suggest that even if they were to reject the notion of epistemic analyticity in full generality, their adoption of recarving of content has strong affinities to the conception of epistemic analyticity. So, an analogue to the second feature of Boghossian's account can be found in the notion of recarving of content, which applies provided that a suitable biconditional is the vehicle of the implicit definition.

As as result, there is an interpretation of Hale and Wright's overall position which is – at least with respect to the two features that give rise to the worry of a transmission of warrant-failure – very similar to Boghossian's proposal. I've noted various differences in the details between the two conception; nonetheless, I think that on this reading of Hale and Wright's proposal, we can discern the same kind of epistemic circularity.

Presupposing that Hale and Wright are also adopting the other three assumptions I noted before[32], the following thought leads to a transmission of warrant-failure of the reconstructed Hale and Wright template:

Assume that the intended usage of the suggested template is to provide an explanation how, by reasoning in accordance with it, a subject can acquire a priori knowledge of an ab-

[31] (Boghossian, 1997), p. 334

[32] A1: Being warranted to disquote a sentence requires an understanding of that sentence; A2: Sentences as basic units of understanding and A3: Understanding involves knowledge of meaning. See 5.4.

straction principle. The subject starts out with a meta-linguistic version of the so-stipulated abstraction principle. Having grasped the premises, how can the subject be warranted to disquote on the meta-linguistic version of the abstraction principle? In order to be in a position to disquote, she must have grasped the content of the statement in question. Here, it is the crucial step from premise 3 of the reconstructed Hale and Wright template to its conclusion which *is* reminiscent of Boghossian's template.

Namely, how can a subject be warranted to believe that 'f' has the determined meaning [refers as intended] to fully understand 'f' and '#f', which is needed to be in a position to disquote the target statement? What is needed, according to assumptions 1-3, is that the subject understands the role the term plays in the object-language version of the relevant sentence. But, the notion of recarving of content ties the understanding of the sentence directly to knowledge of that statement and so the threat of a transmission of warrant-failure appears also on this interpretation of Hale and Wright's proposal. Their account, if understood in this way, fails for exactly the same reasons as Boghossian's: To explain a priori knowledge of implicit definitions by means of reasoning from the meta-language version of the implicit definition to its object-language version is wanting since the understanding which is needed to make this crucial transition will not precede the kind of knowledge which is aimed for of the conclusion. A warrant for the conclusion is already presupposed, for the subject to be in a position to warrantedly disquote. Therefore, no warrant for the conclusion can be transmitted on this template.

To be sure, various open questions remain in this discussion of my understanding of Hale and Wright's proposal. It can be argued that the underlying idea of the notion of recarving of content is in important respects different from Boghossian's conception of epistemic analyticity. Furthermore, it can be contested whether Hale and Wright intend the quoted passage to be understood as providing a template to *explain* how a subject can acquire a priori knowledge. And if so, it can be questioned whether my reconstruction, culminating in the above template, is appropriate. Despite these reservations, however, I believe that the general idea behind my argument against Boghossian's account is applicable to my interpretation of Hale and Wright's proposal.

Finally, to generalise, I believe what this discussion has shown is that the transmission of warrant-failure comes with the *general territory* of the traditional connection once the two features – a transition from meta- to object-language and the idea of epistemic analyticity – that seem inherent to this conception, are adopted. If my argument is correct, it shows that this general approach of the traditional connection of how to explain the acquisition of a priori knowledge is flawed: It is impossible to provide an explanation of how a subject can arrive at a warrant for logical or mathematical principles by reasoning from meta-linguistic premises once the notion of epistemic analyticity is adopted.

Despite this very general problem for the traditional connection, I think there is a treatment for this 'illness' which lies at the heart of the traditional connection as conceived by Boghossian and Hale and Wright. In order to be in a position to offer the right kind of therapy which will give rise to an alternative approach that is still in the spirit of the traditional connection, we first need to clearly diagnose what the source of this 'illness' is.

5.7 Diagnosis of the transmission of warrant-failure of the template

In this section, I will take a step back and diagnose the source of the transmission of warrant-failure for Boghossian's and Hale and Wright's conception. To this end, let us ask why, in the first place, is there a need to adopt a template to explain a priori knowledge?

The main motivation to use a template arises from the standard and even the revised view of implicit definitions. The thought here is that a statement is *stipulated* to be (regarded as) true and so, not surprisingly, the stipulation takes place in the meta-language. In order to properly explain the knowledge acquisition from the meta-language version of the implicit definition, a transition to the object-language version is then needed and an appeal to a general template, which accounts for the transition, is made. So, it is this specific conception of stipulation, that it is located on the meta-linguistic level, which generates the requirement for a disquotation step.[33]

In addition to this main motivation to adopt a general template, it is often claimed – at least by Boghossian – that the type of warrant needed to accept the meta-linguistic premises is somehow easy to come by and also a priori: The first and second premise involve meta-linguistic statements that can be grasped to be correct in virtue of *mere* linguistic knowledge. Allegedly, this type of (tacit) linguistic knowledge, should suffice to underwrite a priori knowledge of the mathematical/logical principle that figures in the conclusion of the template.

Lastly, a further feature of the template might be that it can be regarded as providing an explanation of a priori knowledge which also respects the phenomenology of knowledge acquisition in logic and mathematics. The type of reasoning that is needed in order to be justified in believing mathematical or logical principles, is very simple and intuitive and so it is very close to the idea that logical and mathematical knowledge comes "in a flash" and the principle is readily accepted once it is properly understood.

These three considerations in favour of a template have to be defused and indoing so I will offer a new way to conceive of the traditional connection. Let us start by a treatment of the last two 'features' of the general template, before tackling the main motivation and

[33]Note here that Boghossian's initial template fails to appreciate the need for this disquotation step!

with it the notion of stipulation.

A full discussion of Boghossian's claim concerning the a priori status and 'easy' knowledge of the two premises would lead us too far afield. However, in general, it surely can't be the case that having the kind of linguistic knowledge needed to accept the premises is *always* easily accessible and a priori. So, what is needed is a proposal why in the case of mathematics and logic – in contrast to say scientific sentences – the warrant needed to accept the premises is a priori and easily accessible.[34]

More importantly, however, reflect upon the last consideration. The claim is that the template nicely explains the *directness* of knowledge-acquisition of logical and mathematical principles and thus sits squarely with the phenomenology of how we acquire mathematical and logical knowledge.

However, I think that – quite to the contrary – the result of adopting a template leads to a highly *rationalised* account of knowledge acquisition: A subject acquires, by means of *reasoning* through the template, a priori knowledge of logical and mathematical knowledge. And, not only doesn't the phenomenology of knowledge-acquisition seem to involve this highly rationalised process, but also the resulting *inferential knowledge*, is at odds with the aim of the traditional connection!

To be sure, the *intended type* of a priori knowledge of logic and mathematics is meant to be *basic* in the sense that a subject acquires *non-inferential* knowledge of logic and mathematics by means of implicit definitions. And, since *reasoning through a template* can hardly be regarded as providing *non-inferential* knowledge, the very idea of providing such a template is incompatible with the very idea of the traditional connection.

As a result using a template – involving a transition from meta- to object-language – to

[34]Boghossian is aware of this problem concerning the a priori status of his second premises and addresses it in the appendix to (Boghossian, 1997). My previous discussion has shown why I don't think that knowledge of the premises is *easily accessibly* since it involves a warrant for the truth of the conclusion.

explain knowledge acquisition by means of implicit definitions *has* to be the wrong approach. Not only does the intended explanation fall foul of a transmission of warrant-failure but even if it could avoid this problem it would generate the wrong type of knowledge.

So, we are only left with the main motivation for the adoption of a template. The thought that the vehicle of the implicit definitions is stipulated to be (regarded as) true and that therefore a transition from meta to object-language is needed. It is this specific conception of stipulation which is at the root of the problem for the traditional connection; it is this conception that needs treatment.

5.8 Therapy: Stipulation as primitive acceptance

The idea of stipulation was first introduced in the context of the theory of implicit definitions. The intended view here was to stipulate a sentence to express a truth and thereby fix the meaning of the newly introduced term.

This simplistic conception of stipulation was rejected in the light of my arguments for the effectiveness of certain implicit definitions that, despite achieving this aim, fail to express a truth. As a result, I suggested a revised version of the notion of stipulation, which incorporates the possible falsity of the vehicle of the implicit definitions. So, the idea became that an implicit definition endows the new term with meaning, in virtue of thinkers *regarding* the sentence in which it figures as true. This type of stipulation was meant to allow that the vehicle can turn out to be false; however, since the vehicle was introduced with the intention of expressing a truth, it could still be regarded as having provided the new expression with meaning.

The revised version also faces its difficulties. For example, it cannot be understood in the sense that the vehicle of the implicit definition has *constantly* to be regarded as true. It is hard to imagine how a thinker could regard a statement to be true even if he knows that it is false. Thus it can't be the effect of the stipulation that the subject should behave as she would were the statement true, because if it fails to be true and the subject knows that, then such behaviour can hardly be triggered.[35] So, further refinements are needed to clarify what the attitude of "regarding as true" really involves.

Furthermore, the revised conception of stipulation still has the feel that it talks *about* the statement it is concerned with; so it should also be located on the metalinguistic level, since the idea is that a certain statement, say '#f', is to be regarded as true. Consequently, even on this conception of stipulation, there is, in order to arrive at knowledge of the so-stipulated sentence, the need for a template to account for the transition from the meta- to the object-language.

So, again let us take a step back and ask what *is* the *main* function or the *intended* effect of stipulating a sentence, disregarding how it works in detail? It seems to me that the main effect of the stipulation is that it should result in the *direct acceptance* of the statement that is stipulated. What the previous discussion shows is that such direct acceptance should not be understood as resulting from stipulating that the target sentence is to be regarded as true, or simply that it is stipulated to be true. It is this additional link – the appeal to the truth (regarding as true) of the statement – to induce the acceptance of the statement that leads to the need for an additional disquotation step on the so-stipulated sentence.

But then, how should we conceive of the idea of stipulating a sentence, if not by appeal to the truth of the statement? The key thought to how, I believe, the notion of stipulation is to be understood, should comprise the idea that we have to conceive of the type of *direct* acceptance by means of a stipulation, as *primitive*. So, the main effect of the idea of

[35]Imagine the case of Basic Law V again. It can't be now demanded that a subject should regard the statement as true and behave in such a way as if it were true.

stipulating a statement is for a thinker to directly accept the statement, that is to accept the statement and continue to accept it so long as there is no reason to believe it is incorrect. As such, the primitive acceptance involves no additional *prerequisite* cognitive work on behalf of the subject to investigate whether the statement is correct; rather the subject should be directly compelled to accept it.

Thus, this type of primitive acceptance does not go *via* the meta-linguistic claim; rather the primitive acceptance of a statement immediately gives rise to a grasp of the content of the so-accepted statement, assuming it is effective. In addition, in the best cases, it is this primitive acceptance that gives rise to a priori *non-inferential* knowledge of the statement.

To conclude, this understanding of the notion of stipulation – as primitive acceptance – does not need to appeal to *any* type of reasoning on behalf of the subject in order to be credited with knowledge by means of implicit definitions. Also, so conceived, it at least holds the prospects for being part of an explanation of how a subject can acquire genuine *non-inferential* and *basic* knowledge of logic and mathematics by means of implicit definitions. In addition, this initial proposal also nicely captures the phenomenology of knowledge acquisition in mathematics. Coming to know a basic logical or mathematical principle has a type of immediacy and obviousness that can be explained in virtue of the primitive and so direct acceptance of an effective implicit definition.

To be sure, however, this is a mere outline of a proposal to account for non-inferential logical and mathematical knowledge which is in the spirit of the traditional connection. The last chapter will be examining further, and rather generally, the prospect of this idea by investigation the presuppositions that have to be in place – i.e. the conditions for singling out the good cases – for the primitive acceptance of a statement to be knowledge-conferring.

Chapter 6

The traditional connection revived

Introduction

In this final chapter I will outline a general account how non-inferential a priori knowledge resulting from a primitive acceptance is best conceived. To this end, I offer a general model for non-inferential a priori knowledge and discuss whether it is best placed within either externalist or internalist conceptions of knowledge, or whether Crispin Wright's recent proposal of *entitlement* should be invoked in order to accommodate this model. All three conceptions face various difficulties that will be highlighted in the course of this chapter. These considerations will give rise to a proposal – a broadly externalist conception that appeals to entitlements at the level of *claims* to knowledge – which, I argue, is best suited for an account of non-inferential a priori knowledge by means of a primitive acceptance of implicit definitions. To be sure, various issues remain open in this discussion, and, at the end of the chapter, I will highlight what I think are the most important questions for the tenability of my account and I will suggest various strategies to tackle them.

6.1 Primitive acceptance and knowledge

In the previous chapter, I suggested a very basic understanding of the functioning of stipulating a statement that does not involve stipulating the truth of the statement, nor the attitude of regarding as true the statement; rather it effects a primitive acceptance of the so-stipulated statement. I suggested it is this conception of regarding stipulation as primitive acceptance that will help to revive the traditional connection. It will avoid the appeal to reasoning in accordance with the kind of template illustrated by Boghossian's discussion, because the so accepted statement need not be disquoted, and so holds the prospect of providing genuine non-inferential knowledge.

In this chapter, I will be concerned with the question how, and under what circumstances, a subject's primitive acceptance of a statement can give rise to knowledge of that statement. I think we can identify two important issues that need further discussion. First, when is it that a subject, having primitively accepted a statement, arrives at an understanding of that statement? Second, under what circumstance does a subject have knowledge of the primitively accepted statement?

To resolve the first issue, I suggest that the idea of a primitive or direct acceptance of a statement comprises the thought that the subject is directly confronted with the *content* of what is stipulated. This is why, there is no need for an additional disquotation step. Instead, provided that the so-stipulated vehicle of the implicit definition is *effective*, the subject acquires by means of his primitive acceptance of the statement a grasp of the content of that statement.

This proposal, I think, fits nicely with the phenomenolgy of grasping the content of a basic mathematical principle, such as Hume's Principle, or of various logical principles "in a flash". Certainly this suggestion should be elaborated, but, I believe, it should be done with respect to a specific theory of meaning and understanding, which surely I won't be able to settle on here. Therefore, let me now turn to the second issue.

The aim is to explain under what circumstance a primitive acceptance of a statement can give rise to knowledge. Since, the intended type of knowledge is non-inferential, no additional cognitive work involving inferences should be required in order to gain knowledge by means of a primitive acceptance. This, however, should not mean that there are no presuppositions that have to be in place, if a primitively accepted statement is to be known. To fix ideas, let me suggest the following model of how a subject acquires non-inferential knowledge:

> *Model for non-inferential knowledge*
>
> A subject, primitively accepting an implicit definition p, acquires thereby an understanding of p, provided that p is *effective.* If, in addition all relevant (epistemic) presuppositions are met, the subject thereby acquires knowledge that p.

What should be clear is that the primitive acceptance of an implicit definition won't – all by itself – *guarantee* the truth of the target statement (or the validity of the target rule).[1] What remains, therefore, for a satisfying explanation of this type of non-inferential knowledge is to scrutinise further the presuppositions that are involved for knowledge by means of implicit definitions. This may then provide a general explanation how those effective definitions that can lead to knowledge (the good cases) are distinguished from those which are effective, yet fail to found knowledge (the bad cases).

What I mean by *presuppositions* here can be highlighted by an analogy to the case of perceptual knowledge, which is often regarded as a type of non-inferential knowledge: Roughly speaking, the thought – obviously not uncontentious – is that acquiring knowledge through perception involves having a certain belief p triggered by a certain perceptual experience.

[1] Otherwise this idea of non-inferential knowledge would be very close to traditional conventionalism.

On the basis of the *correct functioning* of the faculty of perception, this belief that p constitutes non-inferential knowledge that p. In order to have knowledge the *presupposition* of the correct functioning of the faculty of perception has to be in place.

Similarly, if basic mathematical knowledge is non-inferential, we need to investigate what the presuppositions are in order to have non-inferential knowledge on the basis of a primitive acceptance. But it is important to note that by drawing this analogy, I exactly do *not* wish to imply that non-inferential knowledge of mathematics is based on a perception-like faculty. Quite the opposite: On the suggested model of non-inferential knowledge, there are *no* true beliefs formed in virtue of a *faculty*. Instead, it is the stipulative character of the implicit definition that effects a primitive acceptance – which, at least for the moment, we can regard as a belief-like state.[2] What would defeat the primitive acceptance of p as counting as genuine non-inferential knowledge, it seems to me, is if the primitively accepted statement turned out to be unsatisfiable, inconsistent or non-conservative.[3] Therefore, the presupposition for a primitive acceptance of a statement to count as knowledge is that there are no such defeaters (inconstistency, non-conservativeness) for the target statement.

Yet, is it acceptable to just assume – as externalism does in the case of perceptual knowledge – that this presupposition is met and that there are no defeaters? In the light of the case of Basic Law V, it seems that a further investigation is needed to discuss how, on the model for non-inferential knowledge I am suggesting, we should best conceive of the underlying presuppositions. Are they in any relevant respect *epistemic*, or can can they be conditions which merely have to be satisfied, whether or not anyone knows it? In the following sections, I will discuss very broadly, what general framework, if any, is most suitable for this type of non-inferential knowledge by means of primitive acceptance.

[2]I will say more about this in section 6.5.1.

[3]I assume here and in the sequel that the target statement is effective. Also I will leave open whether there are other defeaters.

6.2 Externalism and the traditional connection

Broadly externalist conceptions of knowledge generally consider the presuppositions for knowledge – such as the correct functioning of the perceptual apparatus – as *non-epistemic.* Provided that, as a matter of fact, the presuppositions for a perceptual belief that p are met, the subject can be credited with knowledge that p. No additional cognitive work on behalf of the subject is needed to be credited with perceptual knowledge. In particular, the satisfaction of the presuppositions for perceptual knowledge is, in general, not *reflectively* accessible to the subject.[4]

Prima facie, this general externalist framework fits very well with the idea of non-inferential knowledge, even in the case of mathematics. The model for non-inferential mathematical knowledge comprises the idea that by means of a primitive acceptance of p, which does not involve any additional cognitive work, the subject can in the best case come to knowledge of the so-accepted statement p. The fact that there are no defeaters for p – i.e. that the presuppositions for knowledge of p are met – is not something that has to be considered and ascertained by the thinker in order to be credited with knowledge. More specifically, the fact that the presuppositions are met is not accessible in virtue of reasoning or reflection; that is it can't be a cognitive insight that the presuppositions for knowledge are met.

This last claim also nicely suits as an interpretation of Gödel's results.[5] He has shown that there is no proof available – no cognitive insight – of the consistency of a mathematical theory, assuming it is consistent and the theory is of appropriate strength. Hence, it seems that once we are concerned with non-inferential knowledge of the basic mathematical principles

[4]This is a standard characterisation of externalism, namely denying access internalism. See for example (Alston, 1998) or (Pappas, 2005).

[5]First published in (Gödel, 1931).

that found such a theory, externalism provides a perfectly appropriate framework.

However, we need to be careful not to overstretch the analogy to the perceptual case and remember that there are differences between perceptual and mathematical knowledge. Noteworthy here is that in the perceptual case an appeal is usually made to the *reliability* of the faculty of perception in order to underwrite the general success of knowledge acquisition by its means. Yet, it hard to make such a case for the reliability of the "faculty" of *primitive acceptance* of a statement! More has to be said about why, in normal circumstances, such acceptance produces knowledge and does not involve the acceptance of an effective, yet unsuccessful implicit definition. The case of Basic Law V can, once again, be regarded as prototype for this epistemic version of a *bad company* objection.

Moreover, there is a general intuition – prima facie incompatible with externalism – that since mathematical knowledge is a priori, it should have a kind of transparency. If a subject knows a priori that p, it seems – in normal circumstance – that the subject should also be in a position to reflect on this knowledge and higher-up, as it were, know that she has this kind of knowledge that p. To sharpen these considerations consider the following short argument:

'W_{cog}' stands for cognitive warrant and 'K' stands non-inferential knowledge:

(1)	1	$K(p)$	Assumption
(2)	2	$\neg W_{cog}(\text{no defeaters for p})$	Externalism
(3)	3	$K(p) \rightarrow \text{no defeaters for p}$	Knowledge
(4)	3	$W_{cog}\ (K(p) \rightarrow \text{no defeaters for p})$	Phil. reflection
(5)	3	$W_{cog}\ K(p) \rightarrow W_{cog}\ (\text{no defeaters for p})$	Closure of W_{cog}
(6)	2,3	$\neg W_{cog}K(p)$	mtt
(7)	1,2,3	$K(p)\ \&\ \neg W_{cog}K(p)$	&- Intro

The idea here is the following: Assume that the primitive acceptance of p gives rise to non-inferential knowledge of p – i.e. the presuppositions for non-inferential knowledge are met. The second premise comprises the externalist assumption that there is no cognitive access to the presuppositions for this knowledge. Moreover, it follows that since knowledge of p entails the truth of p, p is not defeated. Furthermore, it seems very intuitive that the subject can reflect on this condition of knowledge and so reflectively know (and thereby have a cognitive warrant) that if there is knowledge of p, then there are no defeaters for p. Provided that cognitive warrants are closed, a subject – despite having non-inferential knowledge of p – has no cognitive accessibly warrant that she has knowledge of p.

The problem the reasoning highlights ties in with the two issues we raised before. It is questionable at best whether a pure externalist framework can show more than how it is *possible* for a subject to have non-inferential knowledge. Since there is no cognitive access to the presuppositions, there is no means accessible to the subject to know whether, in a specific case of primitive acceptance of p, p is known. This raises the general issue *how*, on an externalist proposal, a subject can *legitimately claim knowledge* which is grounded in a primitive acceptance. This question is particularly pressing in the light of doubts – based on Basic Law V – concerning the reliability of primitive acceptance as generating knowledge. Moreover, lacking cognitive access to the presuppositions, it is hard to see how there can be higher order knowledge – a subject's knowledge that she knows that p.[6]

These considerations should not be regarded as decisive against externalism; rather they are put forward to set the agenda. We need to determine what resources are needed and available to explain the following sets of issues that are crucial to a satisfying account of non-inferential knowledge by means of a primitive acceptance of effective implicit definitions.

[6]What I mean with reflective higher-order knowledge is that the subject can – on the basis of reasoning, self-knowledge etc, acquire knowledge that it knows. The argument above should still hold once W_{cog} is substituted by K_{ref}.

1. *Knowledge ascription of p*

   What preconditions have to be met for a subject to non-inferentially know that p on the basis of a primitive acceptance of p?

2. *A subject's legitimate claim to non-inferentially know that p*

   On what basis can a subject legitimately claim non-inferential knowledge that p in such cases?

3. *Higher-order knowledge of p*

   How, if at all, can a subject know that she non-inferentially knows that p in such cases?

4. *Higher-order claims of knowledge of p*

   On what basis can a subject legitimately claim to know that she knows that p in such cases?

The externalist framework can quite simply explain how a subject can know a primitively accepted statement p. Yet it remains questionable what resources there are, on the externalist outlook, for a subject to legitimately claim non-inferential knowledge of p, or even have higher-order knowledge of p. We will come back to the externalist option, since I will defend a version of it in a later section. Let us now look whether a broadly internalist position is so much as compatible with the idea of non-inferential mathematical knowledge that we are presently reviewing.

6.3 Internalism and the traditional connection

Internalism, or rather access internalism – the most common form of internalism – holds that knowledge requires access, by means of reasoning or self-knowledge, to a knowledge-conferring justification.

In this setting a subject having primitively accepted an effective implicit definition p, can acquire non-inferential knowledge of p, provided, not merely that the presuppositions are met, but that their being so is – at least in principle – accessible to the subject (by means of reasoning, etc.). If this is a tenable position, it immediately resolves the problem how a subject can claim knowledge on the basis of a primitive acceptance: The knowledge-conferring justification is to be at least in principle accessible to the subject, and so the subject will be in a position to cite the knowledge-conferring justification. Once having done so, she will then be in a position to claim knowledge of the primitively accepted statement.

The problem for this conception is that, in the case of mathematical knowledge, one would generally consider a *proof* to be what provides the knowledge-conferring justification. Here however, we are concerned with the *basic principles* themselves that are meant to be foundational of a mathematical system, and with knowledge that arises by primitive acceptance, subject to certain presuppositions. Now, if a proof is required – or at least has to be in principle accessible – that the presuppositions, for instance consistency, are met, then its seems that the threshold for non-inferential knowledge is set too high, at least for knowledge of arithmetic and beyond. Gödel's result forecloses the possibility of access to this type of knowledge-conferring justification.

The problem we are encountering here is that once it is claimed that there is *cognitive access* to the presuppositions – the internalist position – the issue immediately arises what this type of cognitive warrant might consist in? In the case of consistency, at least, it can't be a matter of having a proof. But then, how else should we conceive of it? In the next section, I will review one suggestion that is broadly in line with the general spirit of the internalist proposal.[7]

[7]In the following I will continue to focus on the presupposition of consistency of a basic principle.

6.3.1 Internalism and Tarski's thesis

Instead of thinking of the knowledge-conferring justification as a proof that the principle in question is consistent, the suggestion here is to provide a statement equivalent to its consistency to whose truth, allegedly, access is available. Tarski's thesis is what provides this equivalence:[8]

Tarski's Thesis

1. The implicit definition p is consistent $\leftrightarrow$
2. $\neg p$ is not a logical truth $\leftrightarrow$
3. It is not the case that $\forall D(\neg p)^D$ (Tarski's characterization of logical validity) $\leftrightarrow$
4. not: $\neg\exists D(p)^D$ (Double Negation) $\leftrightarrow$
5. $\exists D(p)^D \leftrightarrow$
6. There are some objects (the D's) under which p^D is true.

The internalist – adopting Tarski's thesis – could argue that there can be non-inferential knowledge by means of a primitive acceptance of p, and that this knowledge is internal, when the subject in question has, in principle, access to the fact that there are enough objects to enable p^D to be true.[9] This access could then underwrite the subject's *legitimate claim* to know p.

[8]What I call *Tarski's thesis* is implicit in recent writings of (McGee, 2005). However, I should note that he actually does not use Tarski's thesis in the context I invoke it here.

[9]Note that Tarski's thesis has the following restriction: If the notion of consistency is understood proof-theoretically (in premise (1)), then in order hold the biconditional true, the logic has to be complete (left to right direction) and sound (right to left direction). So, Tarski's thesis would not be of any help to someone who adopts second-order logic.

However, this approach puts the cart before the horse, at least once the outlook of the traditional connection is adopted. The knowledge-conferring justification that is accessible to the subject can't be that there are enough objects around to make the statement true. Rather, according to the very idea of the traditional connection, it should be on the basis of non-inferential knowledge of the basic principle that a subject can *infer* from *it*, that the objects in question exist. Consequently this piece of knowledge can't figure as the knowledge-conferring justification of the basic principle in the first place.[10]

From these considerations, the following lesson emerges for the idea of non-inferential knowledge based on primitive acceptance: The type of warrant or justification that gives rise to non-inferential knowledge once an implicit definition is primitively accepted need not, and cannot, itself depend upon further cognitive work. The whole idea is to account for *immediate* knowledge, in that the *knowledge-conferring* justification does not itself involve further reasoning, or need to be of the type that is accessible by reasoning to the subject in order to be credited with this type of knowledge. So, it is extremely doubtful whether a broadly internalist conception is at all compatible with the idea of non-inferential knowledge as defended here.

The externalist conception, in contrast, claims that no such cognitive warrant is required of the subject and so this proposal sits much better with the idea of non-inferential knowledge. However, the problem for externalism, as we saw, is that it leaves the subject with no warrant on basis of which she can *claim* knowledge by means of a primitive acceptance.

The important insight of the entitlement theory, to which I now turn, is that for certain basic beliefs, no warrants for their truth can be acquired by means of cognitive achievements

[10]Note also that the prospects of a priori knowledge is under threat on this conception, since it is not clear how this type of knowledge-conferring justification – existence of infinitely many objects in the case of Hume's Principle – could be known a priori yet independent of the acceptance of the mathematical principle itself.

– a broadly externalist idea. Yet, there is a type of warrant, or rather an *entitlement* that does not involve cognitive work, and in this respect it is an *unearned* warrant. On its basis it is not *irrational* for a subject to have such a belief. In the following section, I will outline this position as advocated by Crispin Wright[11] and examine how it could contribute to our discussion of non-inferential mathematical knowledge.

6.4 Entitlements and the traditional connection

The key thought for Wright's notion of entitlement is that in pursuing any type of cognitive project (perceiving the world, acquiring knowledge of the past, mathematical knowledge etc.)[12], a subject needs to make specific *unavoidable presuppositions*. So, when she undertakes a project, say finding a phone number in the directory, the thinker needs to take the reliability of her senses for granted. This presupposition can, in certain cases be challenged – one could go and get ones eyes tested, for example. However, this investigation (to get ones eyes checked) would in turn also raise its own presuppositions which are of a similar type. The overall point is that "whenever any cognitive achievement takes place, it does so in a context of *specific* presuppositions, which are not themselves an expression of any cognitive achievement to date".[13]

As a result, there is an element of *adventure* in any cognitive achievement, because at the bottom there will be certain presuppositions that haven't been subject to a cognitive investigation and so they have to be assumed on *trust*, i.e. without evidential justification. Wright introduces the notion of *entitlement* to capture the idea that it is not *irrational* or even *unwarranted* to invest such trust in a presupposition. According to Wright however,

[11]See his (Wright, 2004)

[12]Wright also considers other types of entitlements for rational deliberation (concerned with rational action) and entitlements to substance, these need not concern us here.

[13](Wright, 2004), p.189

entitlements are not underwriting a belief, but rather, due to this lack of evidential support, only give rise to the attitude of *trust* in the presupposition in question. Nonetheless, it can be regarded as a *rational trust* provided that the presupposition counts as a genuine entitlement. Wright proposes the following conditions for a presupposition to qualify as an entitlement. He writes:

> Let me try to harness these ideas to a definite proposal about entitlement. First (to tidy up a bit) a definition: let us say that
>
> P is a presupposition of a particular cognitive project if to doubt P (in advance) would rationally commit one to doubting the significance or competence of the project.
>
> Then the relevant kind of entitlement – an entitlement of cognitive project – may be proposed to be any presupposition of a cognitive project meeting the following additional two conditions:
>
> i We have no sufficient reason to believe that P is untrue and
>
> ii The attempt to justify P would involve further presuppositions in turn of no more secure a prior standing, ..., and so on without limit; so that someone pursuing the relevant enquiry who accepted that there is nevertheless an onus to justify P would implicitly undertake a commitment to an infinite regress of justificatory projects, each concerned to vindicate the presupposition of its predecessor. ((Wright, 2004), p. 190)

So, the idea is that whenever we need to carry out a specific cognitive project whose presuppositions fulfil the first and second requirement stated, then "we should – are rationally entitled to – just go ahead and trust that the [presuppositions] are met."[14]

[14](Wright, 2004), p. 190-1

Let us now consider if and how this proposal could apply to the cognitive project of accumulating *arithmetical knowledge* by means of Hume's Principle. Clearly, it is a presupposition of the cognitive project to acquire arithmetical knowledge that Hume's Principle is consistent, simply because to doubt the consistency of Hume's Principle would rationally commit one to doubting the significance of this cognitive project. Moreover, one can argue that this presupposition – the consistency of Hume's Principle – does fulfil the two conditions to qualify as an entitlement. Consider the second condition first. The issue is whether an attempt to justify the claim that Hume's Principle is consistent, would involve further presuppositions which are of no more secure a prior standing. I think this is the case here, since in order to justify that Hume's Principle is consistent, the only foreseeable means is to provide a relative consistency proof – relative to another mathematical system. However, this system again incurs the *same* presupposition, namely that it is consistent. As a result, it is reasonable to suppose that the enquiry to justify this presupposition would itself "involve further presuppositions in turn of no more secure a prior standing, ..., and so on without limit." And thereby one would "implicitly undertake a commitment to an infinite regress of justificatory projects, each concerned to vindicate the presupposition of its predecessor." So, I believe the second condition is fulfilled.

The first condition is more tricky and also involves some ambiguity. One might think that the occurrence of the paradox in the case of Basic Law V establishes some reason to at least entertain a doubt as to whether Hume's Principle is consistent, since for example, it shares the same form. And so it is unclear whether the condition that "having no sufficient reason to believe that [Hume's Principle is consistent] is untrue" is fulfilled.

To be sure 'having no (sufficient) reason' shouldn't – at least on my understanding – be understood in the sense that there 'being no (sufficient) reason' suffices for a subject is 'not having (sufficient) reason'. To explain, if Hume's Principle is consistent, there would *be* no (sufficient) reason to doubt it – in the "realm of reason" as it were. However, this should not automatically mean that a subject *has* no (sufficient) reason to doubt it. Rather, I think, we should look for an understanding of 'having no sufficient reason' that is information-state relative and so incurs a genuine epistemic responsibility for the thinker.

Thus, let me suggest a proposal under what conditions a subject qualifies as "having no sufficient reason to belief that P is untrue", at least in the case of basic mathematical knowledge. The idea is that, once all *available evidence* ('evidence' in a very broad sense) which might show that a presupposition P is untrue has been surveyed, and it is established that it does not undermine P, then one has no sufficient reason to believe P is untrue.

It is important here is that the notion of *available evidence* is meant to confine the epistemic work that has to be done to what can be reasonably expected from the epistemic subject. To clarify matters, let us consider the consistency of Hume's Principle as a presupposition and see how it complies to condition (i) for entitlements. The thought would go like this: Do we have sufficient reason to believe this presupposition to be untrue? Well, there are inconsistent principles, looking rather similar to Hume's Principle. So, in order to nullify any sufficient reason to believe that Hume's Principle is inconsistent, we need to look at all the *available* evidence which would defeat the presupposition. This means we need to look at whether *known* types of paradoxes can be developed from Hume's Principle. Once we have surveyed all the 'usual suspects' and all other evidence we know of, which might show that the presupposition of consistency is not met, and the upshot is that the principle is not subject to them, then we can consider the first condition to be fulfilled. In these circumstances then there is "no *sufficient* reason to believe Hume's Principle is consistent to be untrue".

Surely, this is a mere sketch, but I think that reasoning along these lines does promise to vindicate the consistency of Hume's Principle as an entitlement, and so a subject can rationally take this presupposition *on trust*! But, how exactly might the entitlement proposal subserve the idea of non-inferential mathematical knowledge by primitive acceptance?

I think we can discern (at least) two ways to implement the entitlement proposal. One option would be to locate entitlements on the level of what I called the presuppositions for knowledge, i.e. the consistency and conservativeness of the target statement – just as I have done in my exposition up till now of the entitlement proposal. So the idea would be, in the special case of Hume's Principle, that a primitive acceptance of Hume's Principle gives rise to knowledge of it, provided that the presuppositions for such knowledge qualify as entitlements.

The second strategy is to regard Hume's Principle *itself* as a mere entitlement. This idea might be tempting since a primitive acceptance, one may argue, does not really give rise to a genuine belief like state, rather it is closer to the idea of Wright's notion of trust. This strategy is intriguing, yet it faces the worry that since there is only an entitlement for Hume's Principle, the principle itself isn't knowledgeable! Since my intention here is to capture non-inferential *knowledge* of basic principles, I will, in what follows, pursue an investigation of the first strategy.[15]

So then, how exactly would an appeal to the entitled presupposition help with the issue of knowledge acquisition on the basis of a primitive acceptance – and with the issue of how a subject can be in a position to legitimately claim knowledge of the primitively accepted statement? In the following, I will raise some general worries about the relationship between knowledge of a statement, and entitlements for its presuppositions. This will put some strain on the first strategy to deploy entitlements within the conception of non-inferential knowledge based on primitive acceptance.[16]

[15]In 6.5.1 I will furthermore argue, or rather suggest that primitive acceptance is a belief like state, which would rule out Hume's Principle, if arrived at my means of a primitive acceptance, as a belief-like state.

[16]I conjecture that similar worries would apply to the second strategy to regard Hume's Principle itself

6.4.1 Some problems for the entitlement theorist

In this section, I will offer various considerations against Wright's view of entitlement. First, I consider a problem that Wright himself considers and which he terms the *leaching* problem. What is interesting here is not so much the problem itself, but rather how Wright aims to resolve it. Second, I want to suggest a similar problem that I label the *inverse-leaching* problem. Lastly, I will put further pressure on Wright's conception of first-order knowledge claims and the role (or lack thereof) entitlements play therein. My overall aim is to motivate an alternative proposal that is very similar in spirit to Wright's view, but differs in certain crucial details.

The leaching problem is this. Once entitlements are introduced and one has entitled trust in the presuppositions of a cognitive project, then as Wright expresses it:

> "we are bound to risk 'leaching', as it were – an upward seepage of mere entitlements into areas of belief which we prize as genuinely knowledgeable or justified."[17]

To take the example of Hume's Principle again: Once it is acknowledged that our presumed knowledge of Hume's Principle rests on a mere entitlement to trust that it is consistent, then it seems as if we should be in no better position than having a mere entitlement (and no more) to Hume's Principle itself and any other mathematical statement that follows from it.

Wright's answer to the leaching problem is that it is not as problematic as it seems and merely calls for a small concession. He claims that "leaching" does not affect first-order knowledge and a subject's claim to knowledge. That is, so long as we are merely

as a mere entitlement. Note that some of the problems that I will discuss below rely upon assuming that there is non-inferential mathematical knowledge along the lines I have been suggesting. Hence, they are not directly attacking the entitlement account but rather the idea of using entitlements within the previously suggested setting of non-inferential knowledge based on a primitive acceptance.

[17](Wright, 2004), p.178

concerned with knowledge of a mathematical statement, then the mere entitlement to trust a presupposition, such as its consistency, does not undermine the resulting knowledge of the mathematical statement. Rather, Wright continues: "In general, the effect of conceding that we have mere entitlements for cornerstones [presuppositions] is not uniformly to supplant evidential cognitive achievements – knowledge and justified belief – with mere entitlements right across the board but to qualify our *higher order* cognitive achievements." As a result Wright concludes that "scepticism demands a surrender of higher order knowledge – the claim to know that we know." Hence, the problem for the entitlement theorist is that once we go higher-order and ask ourselves to what extent one can know that all the preconditions are met – that is whether we know that we know a certain statement – then it seems the entitlement theorist has to give way. He has to acknowledge that from the perspective of the second-order question the most that can be claimed for the presuppositions are mere entitlements and so these entitlements "leach" into and degrade the higher-order knowledge and its justification.

If I understand Wright correctly, then the following proposal about the relationship between entitlements and first-order knowledge/knowledge claims is emerging. Entitlements do play a role in first order knowledge and first order knowledge claims. Yet, they are not explicit in the type of justification that is needed for such knowledge (claims). Rather they are suppressed in the sense that their verification is not among cognitive achievements that are needed for having first order knowledge – whatever they may be in a specific case. There are entitlements that underwrite knowledge or even knowledge claims but they are – for first order knowledge (claims) – neither required to be accessible to the subject nor is there a *need* to acknowledge the entitlements for the presuppositions, in order to have first order knowledge or claim such knowledge. So for example, the subject doesn't have to check that the presuppositions for knowledge do qualify as entitlements; rather *the fact that they do so* *qualify* is sufficient, so long as we are on the first order level.

It is only on the level of second order knowledge and second order knowledge claims that entitlements enter the sphere of explicit justifications which are required. Since entitlements are merely taken on trust without evidence, the having of *mere* entitlements for the presuppositions of first order knowledge *does* undermine the possibility of genuine higher order knowledge and higher order knowledge claims.

If this is the correct understanding of Wright's proposal, then I think the following two observations give rise to some doubts about its tenability. Consider first the following *inverse-leaching* problem.

Let it be granted that there is a reply to the leaching problem and that entitlements do not affect the whole of one's body of knowledge. Instead entitlements remain at the *foundations*, unless higher-order knowledge or higher order knowledge claims are at issue. So, since there are only entitlements for presuppositions; these presupposition are not known, even when first-order knowledge is considered or first order knowledge is claimed. If that much is granted, then I wonder why it isn't the case that knowledge doesn't – to use a metaphor – sink down into the foundation? Instead of worrying that mere entitlements may supplant everyday knowledge – the original leaching problem – the issue should be raised that if, by hypothesis, we do have knowledge of everyday propositions or knowledge of mathematical propositions that are not at the foundation, why doesn't knowledge sink down to knowledge of the foundations? The point can be made by appealing again to Hume's Principle, the presupposition that Hume's Principle is consistent and the following reasoning.

(1)	1	K (HP)	Assumption
(2)	2	K (HP $\rightarrow$ HP is consistent)	Mathematical knowledge
(3)	2	K (HP) $\rightarrow$ K (HP is consistent)	Closure of K
(4)	1,2	K (HP is consistent)	mpp (1),(3)

The conclusion of the argument is unacceptable to the entitlement theorist since the presupposition of consistency is not known; rather one can merely have entitlements for it. This simple argument merely relies on knowledge of Hume's Principle, the additional mathematical knowledge that if Hume's Principle is true, it is consistent and closure. The first two assumptions can hardly be challenged. Yet, I think a possible rejoinder could be located in a justified rejection of closure by the entitlement theorist.

A possible reply on behalf of Wright relies upon the unique feature of his proposal that warrants are *disjunctive*: they either comprise evidence (broadly speaking) or mere entitlements. Consequently, what looks like a simple closure step is much more involving! So for example, if we were to consider the following analogous argument the failure of the closure step seems well motivated on the entitlement account:

(1)	1	K(I have a hand)	Assumption
(2)	2	K (I have a hand $\rightarrow$ I'm not deceived by a Demon)	Reflection
(3)	2	K(I have a hand) $\rightarrow$ K (I'm not deceived ...)	Closure of K
(4)	1,2	K (I'm not deceived by a Demon)	mpp (1),(3)

In this case, the first premise appeals to a warrant which comprises perceptual evidence *and* an underlying entitlement to assume that there is no evil Demon who deceives the thinker. It is these two elements that underlie knowledge that I have a hand. The second premise involves a warrant based on philosophical reflection and certain entitlements about the faculty of reasoning of that thinker. For this reason, it seems to me the entitlement theorist is well justified in rejecting the closure step. Namely, since warrants are disjunctive, it may well be argued that the perceptual evidence is not closed (since the second premise involves knowledge by reflection) while the entitlement part of the warrant is. Thus, the conclusion does not follow, rather there is only an entitlement for the conclusion.[18]

Yet, does this diagnosis apply in the same way to mathematical knowledge or to a priori knowledge in general? If knowledge of Hume's Principle is based on a primitive acceptance and certain entitlements, then it is not so obvious why the closure step shouldn't hold. Certainly, it is the case that primitive acceptance is not closed under this type of reasoning. However rejecting (rightly) the closure of primitive acceptance can't be sufficient to undermine knowledge of the conclusion. For example, consider the following, analogous argument:

(1)	1	K (HP)	Assumption
(2)	2	K (HP $\rightarrow$ Peano axioms)	Frege's theorem
(3)	2	K (HP) $\rightarrow$ K (Peano axioms)	Closure of K
(4)	1,2	K (Peano axioms)	mpp (1),(3)

Assume that a thinker primitively accepted Hume's Principle. Still, it should be possible to extend his knowledge by proving Frege's theorem and thereby arriving at knowledge of the Peano axioms. Yet, according to my suggested reading of the entitlement approach, an analogous piece of reasoning can't extend knowledge once we are concerned with the consistency of Hume's Principle! What is needed, it seems to me, is a *general explanation*

[18] Wright wouldn't even agree that there is, on the *basis* of the reasoning a *warrant* for the conclusion, since he regards these arguments as falling foul of a transmission of warrant failure. This additional worry need not concern us here.

why in the case of the Peano axioms, closure holds and one can have *knowledge* of them in virtue of knowledge of Hume's Principle; yet, in the case of the consistency of Hume's Principle, there is *no inverse leaching* and one cannot have *knowledge* of it in virtue of knowledge of Hume's Principle since here closure fails. What though justifies the closure failure in the latter case and not in the former when both pieces of reasoning are relevantly similar?

As a result, although I don't think that this problem is yet fully conclusive, it raises an important issue about the *stability* of the notion of entitlement at least in the context of mathematical knowledge. This worry, as we will see in the next section, is avoided on my proposal.

Let me now turn to the last consideration against the entitlement theory. Here, I want further to investigate Wright's claim that the leaching problem affects neither first order knowledge nor first order knowledge claims. I want to register a doubt concerning the idea that first order knowledge *claims* do not involve a subject's acknowledgement of the entitlements for the presuppositions. I will suggest that entitlements *leach* even into first order knowledge claims. Again, let us focus here on the special case of mathematical knowledge.

Assume that a thinker S claims to know Hume's Principle. If the general outlook of non-inferential knowledge is correct, then it will be on the basis of a primitive acceptance of Hume's Principle that he has such knowledge. However, since no cognitive achievements have taken place, let us ask again, on what basis can he claim to know the principle?

It seems to me that claiming knowledge of Hume's Principle involves being in a position to provide at least some reasons for thinking that one knows this principle. And in the case of a foundational mathematical principle this should also involve some reasons for supposing that Hume's Principle doesn't share the same fate as Basic Law V – being effective yet inconsistent. However what type of reasons could the thinker cite to make a justified claim to know? It seems to me that the subject should be in a position to cite reasons that involve the idea that "there are no sufficient reason to believe that the presupposition are untrue", or that "I could give you a relative consistency proof, which however will be of no more secure a prior standing". Hence, the idea I'm driving at, is that for a subject to make a *legitimate claim* to know, she should be in a position to cite the entitlements for the presuppositions. As a result, I think that, even in the case of first order knowledge *claims* (in the case of a mathematical principle), entitlements have the potential to "leach".

To be sure, this consideration is suggestive at most and I won't be able to provide strong *arguments* for my conception. Yet, I think that the above example nicely motivates the view that something akin to entitlements are needed to make a legitimate claim to know a basic mathematical statement.

So, I will use these considerations – about leaching/inverse leaching and first order knowledge claims – to open up space for an alternative position that is very close to Wright's proposal. The proposal that I am suggesting *drops* the idea of entitlements altogether for first order knowledge and adopts a purely externalist conception of this type of knowledge, thereby avoiding both the leaching and the inverse leaching problem. Nonetheless, it acknowledges the need for entitlements on the level of *claims to knowledge*. As a result, I think this *hybrid* view will fit nicely with the idea of non-inferential knowledge on the basis of a primitive acceptance, while avoiding the externalist pitfalls by offering an account on the basis of which a subject can legitimately, yet defeasibly, claim non-inferential a priori knowledge.

6.5 Externalism *cum* entitlement: Towards an epistemology for the traditional connection

This proposal has two components. It adopts externalism for non-inferential a priori knowledge and a version of the underlying idea of entitlements in order to ground legitimate *claims* to know.

The externalist component comprises the idea – as outlined in the previous section – that the primitive acceptance of an effective implicit definition gives rise to knowledge, provided that as a matter of fact the relevant presuppositions for knowledge are met. The subject can therefore be credited with knowledge without having to engage in further cognitive work but merely on the basis of primitive acceptance and the satisfaction of the 'external' conditions. On this conception, the resulting knowledge can justifiably be regarded as non-inferential.

However, as highlighted earlier, a purely externalist approach doesn't seem to offer anything that would underwrite a thinker's *claim* to know the accepted statement. And this is exactly where Wright's important insight comes in that, whenever a thinker engages in a cognitive project there is an element of adventure involved. Yet, this element of adventure, I think, is best located in the thinker's *claim* to knowledge and not in knowledge-acquisition simpliciter.

I believe it is a fatal misconception to think that once a subject has mathematical, or even more general a priori knowledge, she has to be *certain* that she has this knowledge. That is, when the thinker has or even claims a priori knowledge, that she has to have absolute certainty that she knows the statement in question. I think this misconception has lead astray both previous attempts to found mathematical knowledge on the basis of implicit definitions and critiques of it.[19]

[19]For example Quine's argument in (Quine, 1948) relies on the idea that a priori knowledge (claims) have

This idea that a subject need not be certain in order to claim a priori knowledge can be further motivated by appeal to an epistemic version of the "ought implies can" principle: Namely, if a thinker cannot discharge an alleged epistemic obligation, then it is not the case that there is any obligation in the first place. So, in order to *legitimately* claim knowledge of a mathematical statement, the epistemic obligation to do so can't outstrip what is in general possible for the subject to discharge. To be sure, a crucial question here is under what conditions is it genuinely impossible to discharge the epistemic obligation. Obviously, it should not depend on irrelevant circumstances – such as the thinker's mood or motivation (he can't discharge it because he doesn't want to). Rather, what I'm hinting towards is the idea that since it is impossible to provide non-relative consistency proofs for statements of a specific strength (such as Hume's Principle), it is not part of the epistemic obligation to provide such a proof in order for a subject to claim knowledge of it.[20]

To be sure, however, this doesn't let the subject off the hook, as it were, leaving her free to claim knowledge of any mathematical statement. Rather, there are legitimate and illegitimate claims to know, i.e. there are certain epistemic responsibilities that have to be assumed in order for a thinker to be in a position to claim knowledge of a statement by means of a primitive acceptance. And it is exactly at this stage that Wright's proposal for entitlements is adopted to provide an account under what conditions it is *legitimate* and *not irrational* to *claim knowledge* on the basis of a primitive acceptance. The thought would be this: for a subject to *claim* knowledge of a primitively accepted effective definition, say A, the subject has to have access to entitlements for the presuppositions of A. To clarify consider the following slightly revised version of Wright's general schema for entitlement:

to involve the highest degree of certainty.

[20]This idea might need further qualification but at least for the moment I think it seems very intuitive and should be applied to knowledge claims. Moreover, I thereby knowingly adopt a kind of deontological – and so according to some, 'internalist' concept of justification for a priori knowledge (claims). This however, does not undermine the general externalist outlook that I also adopt.

P is a presupposition of a particular primitively accepted implicit definition A if to doubt P would rationally commit one to doubting the truth of A.

Then the relevant kind of entitlement – an entitlement to legitimately claim knowledge of A – applies to any presupposition, P, of A meeting the following additional two conditions:

i We have no sufficient reason to believe that P is untrue and

ii The attempt to justify P would involve further presuppositions in turn of no more secure a prior standing, ..., and so on without limit; so that someone pursuing the relevant enquiry who accepted that there is nevertheless an onus to justify P would implicitly undertake a commitment to an infinite regress of justificatory projects, each concerned to vindicate the presupposition of its predecessor.

So to summarise: For a subject to claim to know p on the basis of a primitive acceptance, he has to make sure that he is entitled to do so, since primitive acceptance can't all by itself provide a guarantee that the statement is true. This is the lesson learned from cases like Basic Law V. Yet, we can't demand that in order to claim knowledge, the subject has to be fully certain that the implicit definition is "in good standing", since such a demand would violate the epistemic version of the "ought implies can" principle. Still, in order to legitimately claim knowledge, what is needed is an entitlement to do so. What this involves therefore, in keeping with my previous discussion of Wright, is that the subject has to make sure that any presupposition for A qualifies as an entitlement.

Consider the presupposition of the consistency of A again. For it to qualify as an entitlement, the subject has to make sure that there are no sufficient reasons to doubt it and that any attempt to justify it, would involve further presuppositions in turn of no more secure a prior standing.

The first condition should be understood in the sense that in order to have no *sufficient reason*, it has to be clear that the available defeating evidence does not apply in the case of this presupposition. I have taken that to involve that it need to be shown that none of the *known* paradoxes apply to the implicit definition in question. In addition, as is the case for the consistency of A, it has to be shown that any attempt to justify this presupposition would involve its own presuppositions, and so on without limit of no more secure standing. Whenever these conditions have been verified, the subject is entitled to claim knowledge of the primitively accepted statement.

One feature of this proposal is that the problematic transition from effectiveness to knowledge of the primitively accepted statement only comes to bear once we are concerned with a *claim* to know the statement. As an externalist about knowledge, the transition on the level of knowledge *acquisition* is automatically made, provided that as a matter of fact the presuppositions are met and the definition is successful. On the level of claiming knowledge of the primitively accepted statement, the transition can be assumed, provided that the thinker has an entitlement to make the claim. However, to repeat, the upshot is merely a defeasible claim to know, which fails if it later turns out that the definition is effective yet unsuccessful.

This general outlook provides a very suitable framework to incorporate the traditional connection and the idea that it is by means of mere primitive acceptance, without appeal to a faculty of intuition, that a theorist can claim basic mathematical knowledge. Yet, this proposal too faces further challenges if it is to provide a satisfying account to meet Benacerraf's epistemic challenge. In the following section I will outline and briefly discuss various issues that will need further attention in future work.

6.5.1 Remaining issues for the hybrid view

Here I will discuss four residual issues that need highlighting in the context of my proposal, they are:

1. Why does the resulting non-inferential knowledge qualify as a priori knowledge?
2. Is there any higher order knowledge, or are there any higher order knowledge claims of basic mathematical principles?
3. How does primitive acceptance relate to belief?
4. Is there an embarrassment of the riches objection, i.e can we claim too much mathematical knowledge?

Certainly, I won't be able to conclusively settle all four concerns here, nevertheless let me note various promising strategies to resolve these questions.

Why does the resulting knowledge qualify as being a priori?

On this proposal, a subject acquires knowledge *merely* on the basis of primitive acceptance and the 'conduciveness' of the presuppositions for the target principle. The type of knowledge acquired is a priori, since it does not rely upon any specific empirical evidence. Obviously, the subject has to grasp the content of the stipulated implicit definition, but that, I think, should not yet undermine the prospects of a priori knowledge in general.[21] Rather, once the statement is grasped, the resulting mathematical knowledge does not rely upon further empirical work or any empirical investigation and so this type of knowledge of the basic mathematical principle should be regarded as genuinely a priori. Also, the fact that *claiming*

[21] This would give rise to general scepticism about the a priori, which I won't discuss here. The question I am raising is more the following: if there is a priori knowledge, would my explanation of non-inferential knowledge of mathematics be a suitable candidate? I think yes.

this kind of knowledge is itself *defeasible* should not undermine its a priori status. As, I argued before, it is a misconception to think that for a subject to claim a piece of a priori knowledge, she has to be certain.

As a result, I think that we should regard basic mathematical knowledge, knowledge arrived at on the basis of primitively accepted basic (effective) principles, as not only non-inferential but also a priori. Since I also believe (without here offering any further arguments) that reasoning within second-order logic preserves the a priori status of a basic principle, any theorems that can be deduced from it will also be knowable a priori.

Is there any higher order knowledge, or are there any higher order knowledge claims of basic mathematical principles?

This question ties in with the third and fourth issue I put forth on p. 229. The question we need to focus on is how we should conceive of higher-order knowledge. It seems to me that often the outer – higher order – 'K' involves a different *kind* of justification than the inner – first order – 'K'. So, for example, assuming that a subject has non-inferential knowledge of p, the question concerning higher order knowledge is not whether it therefore also has non-inferential knowledge that it has non-inferential knowledge of p. Rather, second-order knowledge arises due to a type of transparency. Namely, no matter what type of first order knowledge the subject has, does she have reflective – by means of self-knowledge and reasoning – knowledge that she has the first-order knowledge?

If that is what second-order knowledge involves, then it seems that on the account put forth here there is no genuine second-order knowledge. Quite simply, there is no reflective access to *all* knowledge-conferring justification of the first-order level and so there can't be this type of reflective higher-order knowledge. As a result, this type of knowledge and its corresponding higher order knowledge claim have to be rejected. Nonetheless, this rejection should not be regarded as highly problematic or ad hoc since it is usually associated with the type of externalism I defend here.

How does primitive acceptance relate to belief?

One might object to the ability of acquiring knowledge by means of a primitive acceptance, since – so the opponent may argue – the attitude of acceptance does not yet count as a genuine belief. However, a necessary condition for knowledge is that the subject has – at a minimum – a true *belief* and thus a primitive acceptance of a true statement can't suffice for knowledge of it.

There are two ways of responding to this issue: Either one may drop the demand, motivated by the 'justified true belief' model of knowledge that knowledge *has* to involve a belief; or one may argue that the attitude of primitive acceptance is sufficiently close to a belief state to underwrite knowledge.

Here, I am tempted to argue that the attitude of primitive acceptance is in important respects similar to a belief state. There are various features a belief state has: For example, believing p involves the commitment to think p to be true; believing p connects, and can give rise to, an assertion of p; beliefs are often characterised by contrasting them to desires; a belief is a mental state; a belief interacts with other beliefs, and due to its commitments a subject (being rational and *pace* Graham Priest and other seemingly *rational* Dialethists) would not at the same time believe p and believe not p.

This list of *platitudes* of belief states could surely be further extended, but I think for the purposes here, it suffices to just consider these examples. It seems to me that the attitude of "primitive acceptance" also fulfils these platitudes. To primitively accept p incurs certain commitments, so for example, not to primitively accept not p. A primitive acceptance can give rise to assertions. Moreover, it can be contrasted to a desire and like a belief it is best understood as a mental state.

I think these considerations, although suggestive, should allow one to regard the attitude of primitive acceptance as a belief-like state and thus I think that a primitive acceptance can underwrite knowledge provided certain presuppositions are satisfied.

Is there an embarrassment of the riches objection, i.e. can we claim too much mathematical knowledge?

This issue is one of the hardest issue to resolve. The problem we are facing is to demarcate what type of principles are eligible for primitive acceptance. Here I have been focusing on Hume's Principle and the resulting knowledge of arithmetic. However the issue is whether the same considerations apply to other mathematical principles. For example, can we claim to know the set-theoretic hierachy in this way?

The Neo-Fregean approach holds that it is only abstraction principles – and here only the good cases – that are apt to be known non-inferentially. In chapter 4, I suggested two strategies to argue more generally for a demarcation that might be of use here. The first option is to argue that only abstraction principles are genuinely *effective*. Other mathematical basic principles (or sets of basic principles) fail to clearly specify the meaning of the term and thus, since they fail to achieve the aim of *effectiveness*, can't be subject to a primitive acceptance. Although, I think that abstraction principles fare extremely well in fixing the meaning of the new expression – pace certain Caesar worries – I think it will be hard to maintain that other implicit definitions that take a different form are in general not effective.

The second strategy, pursued at length in chapter 4, was to show that abstraction principles merely incur *indirect commitments*, and so one may argue, they are particularly apt to be primitively accepted. While implicit definitions that do not share the form of an abstraction principle involve *direct commitments* which – it may be argued – renders them unable to be primitively accepted.

I do believe that abstraction principles enjoy a kind of simplicity and obviousness that is missing from other definitions and, previously I also argued that Hume's Principle involves mere *indirect commitments* and so I think this implicit definition in particular is a suitable candidate for primitive acceptance. However, I also suggested at the end of chapter 4 that whether, for example, the Peano axioms in the end also enjoy the status of having mere indirect commitments is a further question that has to be subject to further investigation.

As a result, providing a more general account for demarcating which principles are non-inferentially knowable and other that are not, and investigating in particular which specific principles are so knowable, is high on the agenda for future research.[22]

6.6 Conclusions: Benacerraf's dilemma resolved?

At the beginning of this thesis I outlined Benacerraf's dilemma and posed it as a challenge to any philosophy of mathematics. There, I identified the strategy of the *Fregean* platonist as a conception that rejects the *fundamental assumption* which underlies all alternative positions that I discussed in the philosophy of mathematics. This rejection gave rise to a *Copernican revolution*: Instead of justifying knowledge of mathematics by means of a prerequisite access to the objects mathematics is purportedly about, it is the mastery of that mathematical discourse itself by means of which such justification could be achieved.

I suggested that this perspective will yield a new way to resolve Benacerraf's dilemma

[22]Interestingly, the issue here is not whether certain mathematical principle are a priori knowable or not, since it is already granted that both Hume's Principle and the Peano axioms are a priori knowable (afterall, the Peano axioms are deducible from Hume's Principle using second-order logic). Rather, the issue now becomes which features does a mathematical principle need to have in order to be non-inferentially knowable by means of a primitive acceptance?

by adopting the standard view concerning the semantic constraint and by appealing to the context principle and a theory of implicit definitions to account for the following two epistemic challenges:

1. *Reference and object-directed thought*

   The demand to explain how the objects posited by the semantic theory can, in principle, be in the range of directed thought and talk of the subjects.

2. *Knowledge*

   The 'integration challenge': The demand to reconcile the truths of the subject matter with what can be known by ordinary human thinkers. Crucial here is to provide an explanation of how a subject can have mathematical knowledge and on what basis the subject can claim such knowledge.

I proposed, in chapter one and two, that the first of these requirements could be met by adopting the context principle which guarantees reference to numbers as objects.

I then focused on the second epistemic requirement: the integration challenge. Here, I suggested that the Copernican turn leads directly to a view that is often described as the traditional connection; namelytruth-conditions, that successful implicit definitions can found a priori knowledge of logic and mathematics.

I provided a general framework for implicit definitions and argued for a distinction between effective and successful implicit definitions that, in turn, would lead to a revised version of the notion of stipulation involved in the theory of implicit definitions.

In chapter 4, I then argued against two forms of opposition to the very idea of the traditional connection. I rejected what I called *semantic rejectionism* as defended by Horwich. Epistemic rejectionism, however, turned out more resilient. Although, I don't think I have *conclusively* shown this conception to be wrong, I hope to have furthered the discussion and shown how the traditional connection can avoid the epistemic rejectionist complaint.

Chapter 5 was then aimed to explain, in more detail, how exactly various proponents think implicit definitions can found a priori knowledge of logic and/or mathematics. I argued against two standard conceptions – Boghossian's and Hale and Wright's – and in the light of their problems proposed that the notion of stipulation needs further revision in order to avoid the problems I have highlighted.

In the last chapter, I argued how the traditional connection should be conceived, if it is to be able to explain a subject's non-inferential a priori knowledge of mathematics and logic and resolve Benacerraf's integration challenge. I argued what *broader* framework this conception needs to be set in and how the basic notion of stipulation has to be understood in order to adequately explain how mathematical knowledge can arise by means of effective implicit definitions and so by means of a mastery of the mathematical discourse. Also, I have offered an explanation of how a subject can defeasibly, yet legitimately, *claim* basic mathematical knowledge and so I hope to have offered a general account that reconciles the truths of the subject matter (arithmetic and possibly mathematics) with what can be known by ordinary human thinkers, and so resolves the integration challenge.

To be sure, I have highlighted various open questions that need further attention in order to provide a fully satisfactory answer. Nevertheless, I do believe I have shown the prospects and attraction of pursuing the traditional connection for an explanation of our a priori knowledge of arithmetic and possibly beyond.

On a final note, let us revisit Benacerraf's influential paper "Mathematical Truth" that informed much of my discussion. Even there Benacerraf considered a view that is in some respects similar to the traditional connection, but in his last paragraph he ridicules even the very prospects of this conception. He writes:

> "To clarify the point, consider Russell's oft-cited dictum: "The method of 'postulating' what we want has many advantages: they are the same as the advantages of theft over honest toil" (Russell 1919: 71). On the view I am advancing, that's false. For with theft at least you come away with the loot, whereas implicit definitions, conventional postulation and their cousins are incapable of bringing the truth. They are not only morally but practically deficient as well."(Benacerraf, 1973), p.420

By having provided a general framework for implicit definitions and by having explained the notion of stipulation involved, I hope to have shown that implicit definitions are at least capable of 'bringing' the truth, provided that the conditions for success are met. In addition, by offering an account involving an externalistic conception for mathematical knowledge and the theory of entitlements that underwrite a subject's *legitimate* claim to knowledge on the basis of primitive acceptance of an effective implicit definition, I hope to have demonstrated that nothing is *morally* wrong either with founding non-inferential and a priori knowledge of basic mathematical principles in this way.

Bibliography

Alston, W. P. (1998). Internalism and externalism in epistemology. *Routledge Encyclopedia of Philosophy*, page from http://www.rep.routledge.com/article/P028.

Antonelli, G. A. (1998). Definition. *Routledge Encyclopedia of Philosophy*, page from http://www.rep.routledge.com/article/Y057SECT3.

Azzouni, J. (1994). *Metaphysical Myths, Mathematical Practice.* Cambridge University Press, Cambridge.

Balaguer, M. (1998). *Platonism and Anti-Platonism in Mathematics.* Oxford University Press, Oxford.

Belnap, N. (1962). Tonk, plonk and plink. *Analysis*, 22:30–34.

Benacerraf, P. (1965). What numbers could not be. *The Philosophical Review*, 74:47–73. Reprinted in Paul Benacerraf and Hilary Putnam, *Philosophy of Mathematics.*

Benacerraf, P. (1973). Mathematical truth. *Journal of Philosophy*, 70:661–80. Reprinted in Paul Benacerraf and Hilary Putnam, *Philosophy of Mathematics.*

Benacerraf, P. and Putnam, H., editors (1983). *Philosophy of Mathematics.* Cambridge University Press, Cambridge, second edition.

Boghossian, P. (1996). Analyticity reconsidered. *Noûs*, 30(3):360–391.

Boghossian, P. (1997). Analyticity. In Hale and Wright(ed), *Companion to the Philosophy of Language.*

Boghossian, P. (2000). Knowledge of logic. In Boghossian and Peacocke (ed), *New Essays on the A priori.*

Boghossian, P. (2001). How are objective reasons possible? *Philosophical Studies*, 106:1–40.

Boghossian, P. (2003). Blind reasoning. *Proceeding of the Aristotelian Society Supplementary Volume*, 77.

Boghossian, P. and Peacocke, C., editors (2000). *New Essays on the A Priori.* Clarendon Press.

BonJour, L. (1998). *In Defense of Pure Reason.* Cambridge University Press, Cambridge.

Boolos, G. (1984). To be is to be a value of a variable (or to be some values of some variables). *The Journal of Philosophy*, 81:430–49. Reprinted in George Boolos, *Logic, Logic and Logic.*

Boolos, G. (1985). Nominalist platonism. *Philosophical Review*, 94:327–44. Reprinted in George Boolos, *Logic, Logic and Logic.*

Boolos, G. (1987a). The consistency of Frege's foundation of arithmetic. in J.J. Thomson, ed., *On Being and Saying: Essays for Richard Cartwright*, p.3-20, Cambridge, Mass.:The MIT Press. Reprinted in George Boolos *Logic, Logic and Logic.*

Boolos, G. (1987b). Saving Frege from contraction. *Proceedings of the Aristotelian Society*, 87:137–51. Reprinted in George Boolos, *Logic, Logic and Logic.*

Boolos, G. (1989). Iteration again. *Philosophical Topics*, 42:5–21. Reprinted in George Boolos, *Logic, Logic and Logic.*

Boolos, G. (1990). The standard of equality of numbers. In George Boolos, *Meaning and Method: essays im honor of Hilary Putnam*, Cambridge: Cambridge University Press. Reprinted in George Boolos, *Logic, Logic and Logic.*

Boolos, G. (1993). Whence the contradiction? *Aristotelian Society Supplementary Volume*, 67:213–233. Reprinted in George Boolos, *Logic, Logic and Logic.*

Boolos, G. (1995). Frege's theorem and the Peano postulates. *The Bulletin of Symbolic Logic*, 1:317–26. Reprinted in George Boolos, *Logic, Logic and Logic.*

Boolos, G. (1996). On the proof of Frege's theorem. In A. Morton and S. Stich (eds) *Benacerraf and his Critics*, Oxford: Blackwell. Reprinted in George Boolos, *Logic, Logic and Logic.*

Boolos, G. (1997). Is Hume's Principle analytic? In Richard Heck (ed), *Language, Thought and Logic.* Reprinted in George Boolos, *Logic, Logic and Logic.*

Boolos, G. (1998). *Logic, Logic and Logic.* Harvard, Cambridge, Massachusetts.

Boolos, G. and Jeffrey, R., editors (1989). *Computability and Logic.* Cambridge University Press, Cambridge. third edition.

Brandom, R. (1996). The significance of complex numbers for Frege's philosophy of mathematics. *Proceedings of the Aristotelian Society*, 96:293–315.

Burgess, J. and Rosen, G. (1997). *A Subject With No Object.* Oxford University Press, New York.

Carnap, R. (1936). Testability and meaning. *Philosophy of Science.* 3(1936): 419-71 and 4(1937): 2-40.

Carnap, R. (1939). *Foundations of Logic and Mathematics.* in Readings on the Philosophy of Science, ed. B. Brochy.

Carnap, R. (1950). Empiricism, semantics and ontology. *Revue International de Philosophie,* 4:20–40. Benacerraf and Putnam (ed) *Philosophy of Mathematics* (1983), 2nd edition, p. 241-57.

Carnap, R. (1952). Meaning postulates. *Phil. Studies,*, 3:65–73.

Carnap, R. (1963). Reply to Carl G. Hempel on scientific theories. pages 958–966. in Schlipp(ed) *The Philosophy of Rudolf Carnap.*

Carnap, R. (1966). *Philosophical Foundations of Physics.* Basic Books, New York.

Carnap, R. (1967). *Logical Foundations of Probability.* 2nd ed. 3rd imprint.

Coffa, J. A. (1991). *The Semantic Tradition from Kant to Coffa.* Cambridge University Press, Cambridge.

Colyvan, M. (2001). *The Indispensability of Mathematics.* Oxford University Press, Oxford.

Cook, R. T. (2003). Aristotelian logic, axioms and abstraction. *Philosophia Mathematica,* 11.

Cook, R. T. (2006a). Iteration one more time. *Notre Dame Journal of Formal Logic.* forthcoming.

Cook, R. T. (2006b). What is wrong with tonk (?). *Journal of philosophical logic.* forthcoming.

Cook, R. T. and Ebert, P. A. (2004). Discussion note: Kit Fine, limits of abstraction. *British Journal for the Philosophy of Science*, 55:791–800.

Cook, R. T. and Ebert, P. A. (2005). Abstraction and identity. *Dialectica: Special Edition on the* Julius Caesar Problem, 59(2):1–19.

Cook, R. T. and Ebert, P. A. (prep). Abstraction and identity: Reloaded. Preparation.

Creath, R. (1992). Carnap's conventionalism. *Synthese*, 93:141–65.

Currie, G. (1982). *Frege: An Introduction to His Philosophy.* Harvester Press, Sussex.

Dretske, F. (1970). Epistemic operators. *Journal of Philosophy*, 67:1007–23.

Dummett, M. (1956). Nominalism. *Philosophical Review*, 65. reprinted in Michael Dummett *Truth and other Enigmas.*

Dummett, M. (1978). *Truth and other Enigmas.* Duckworth.

Dummett, M. (1981a). *Frege: Philosophy of Language.* Harvard, Cambridge, MA, second edition.

Dummett, M. (1981b). *The Interpretation of Frege's Philosophy.* Harvard, Cambrdige, MA.

Dummett, M. (1991a). *Frege: Philosophy of Mathematics.* Harvard, Cambridge, MA.

Dummett, M. (1991b). *The Logical Basis of Metaphysics.* Harvard, Cambridge, MA.

Dummett, M. (1998). Neo-fregeans: In bad company. pages 369–87. in Matthias Schirn (ed) *Philosophy of Mathematics Today.*

Ebert, P. (2005). Transmission of warrant-failure and the notion of epistemic analyticity. *Australsian Journal of Philosophy*, 83(4).

Evans, G. (1982). *Varieties of Reference.* Clarendon Press, Oxford.

Field, H. (1980). *Realism, Mathematics and Modality.* Princeton University Press, Princeton.

Field, H. (1982). Realism and anti-realism about mathematics. *Philosophical Topics*, 13:45–69. reprinted in Hartry Field *Realism, Mathematics and Modality.*

Field, H. (1984). Platonism for cheap? Crispin Wright on Frege's context principle. *Canadian Journal of Philosophy*, 14:637–662. reprinted in Hartry Field *Realism, Mathematics and Modality.*

Field, H. (1985). On consistency and incompleteness. *Journal of Philosophy*, 82:239–60. reprinted in Hartry Field *Realism, Mathematics and Modality.*

Field, H. (1989). *Science without Numbers.* Blackwell Publisher, Oxford.

Fine, K. (1975). Vagueness, truth and logic. *Synthese*, 30:265–300. reprinted in Rosanna Keefe and Peter Smith (ed) *Vagueness: A Reader* (1997) MIT Press.

Fine, K. (2002). *Limits of Abstraction.* Blackwell Publisher, Oxford.

Frege, G. (1884). *Die Grundlagen der Arithmetik.* English Translation by J.L. Austin, *The Foundations of Arithmetic*, Northwestern University Press, Evanston, IL, 1980.

Frege, G. (1893/1903). *Grundgesetze der Arithmetik.* Vol 1 (1893), Vol 2 (1903). English Translation by Montgomery Furth, *The Basic Laws of Arithmetic*, University of California Press, Berkeley and Los Angeles, 1964.

Frege, G. (1976). *Wissenschaftlicher Briefwechsel.* Felix Meiner Verlag, Hamburg. edited by G. Gabriel, H. Hermes, F. Kambartel and C. Thiel.

Gentzen, G. (1969). *Untersuchungen ueber das logische Schliessen.* English Translation by M. Szabo, in *The Collected Papers of Gerhard Gentzen*, North-Hollan, Amsterdam.

Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter systeme i. *Monatshefte für Mathematik und Physik*, 38:173–98.

Gödel, K. (1947). What is Cantor's continuum problem? Reprinted in Benacerraf and Putnam *Philosophy of Mathematics* 2nd edition.

Goldman, A. (1967). A causal theory of knowledge. *Journal of Philosophy*, 64:357–72.

Goodman, N. and Quine, W. (1947). Steps towards a constructive nominalism. *Journal of Symbolic Logic*, 12:105–122.

Hahn, E. & Schlipp, P. (1986). *The Philosophy of W.V. Quine.* La Salle, Ill.

Hale, B. (1979). Strawson, Geach and Dummett on singular terms. *Synthese*, 42:275–95.

Hale, B. (1987). *Abstract Objects.* Basil Blackwell, Oxford.

Hale, B. (1994). Dummett's critique of Wright's attempt to resusciate Frege. *Philosophia Mathematica*, 3:122–47. Reprinted in Bob Hale and Crispin Wright, *The Reason's Proper Study.*

Hale, B. (1995). Singular terms (2). in B.McGuiness adn G. Olivieri (ed) *The Philosophy of Michael Dummett*, Dordrecht: Kluwer, pp. 17-44; reprinted in Hale and Wright (2001) *The Reason's Proper Study.*

Hale, B. (1996a). Singular terms (1). in M. Schirn (ed) *Frege: Importance and Legacy*, Walter de Gruyter, Berlin, p. 438-57, reprinted in in Hale and Wright (2001) *The Reason's Proper Study.*

Hale, B. (1996b). Structuralism's unpaid epistemological debts. *Philosophia Mathematica*, 4:122–47.

Hale, B. (1997). Grundlagen, par. 64. *Proceedings of the Aristotelian Society*, 97:243–61. Reprinted in Bob Hale and Crispin Wright, *The Reason's Proper Study*.

Hale, B. (1999). Intuition and reflection in arithmetic: Arithmetic reflection without intuition. *Aristotelian Society (Suppl. vol.)*, 63:75–98.

Hale, B. (2001). A response to Potter and Smiley, abstraction by recarving. *Proceedings of the Aristotelian Society*, 101:339–58.

Hale, B. and Wright, C. (1997). *Companion to the Philosophy of Language*. Blackwell Publishers.

Hale, B. and Wright, C. (2000). Implicit definition and the a priori. In Paul Boghossian and Christopher Peacocke (eds.), *New Essays on the A Priori*; reprinted in Bob Hale and Crispin Wright, *The Reason's Proper Study*.

Hale, B. and Wright, C. (2001a). *The Reason's Proper Study: Essays towards a Neo-Fregean Philosophy of Mathematics*. Clarendon Press.

Hale, B. and Wright, C. (2001b). To bury Caesar.... In Bob Hale and Crispin Wright, *The Reason's Proper Study*.

Hale, B. and Wright, C. (2002). Benacerraf's dilemma revisited. *European Journal of Philosophy*, 10(1):101–129.

Harman, G. (1973). *Thought*. Princeton University Press, Princeton.

Heck, R. (1997). Finitude and Hume's Principle. *Journal of Philosophical Logic*, 26:589–617.

Heck, R. (2000). Syntactic reductionism. *Philosophia Mathematica*, 8(3):124–149.

Hellman, G., editor (1989). *Mathematics without Numbers.* Clarendon Press, Oxford.

Hempel, C. G., editor (1952). *Fundamentals in Concept Formation in Empirical Science.* University of Chicago Press, Chicago. in International Encyclopedia of Unified Science Vol.II,7.

Hilbert, D. (1899). *Grundlagen der Geometrie.* English translation (by P. Benrays): Foundations of geometry, Open Court Publishing Comp, Illinois, 1971.

Hodes, H. T. (1984). Logicism and the ontological commitments of arithmetic. *Journal of Philosophy*, 81(3):123–149.

Hodes, H. T. (1990a). Ontological commitment: thick and thin. pages 235–60. in G. Boolos (ed) *Meaning and Method: Essays in Honour of Hilary Putnam.*

Hodes, H. T. (1990b). Where do natural numbers come from? *Synthese*, 84:347–407.

Horwich, P. (1997). Implicit definition, analytic truth, and apriori knowledge. *Nous*, 31(4):423–440. reprinted in Horwich *Meaning.*

Horwich, P. (1998). *Meaning.* Oxford University Press, Oxford.

Horwich, P. (2000). Stipulation, meaning and apriority. pages 150–169. in Boghossian and Peacocke (ed) *New Essays on the A priori.*

Irvine, A. (1990). *Physicalism in Mathematics.* Kluwer Academic Publishers, Dordrecht.

Lewis, D. (1970). How to define theoretical terms. *Journal of Philosophy*, 67:427–466. Reprinted in Lewis, D. *Papers in Metaphyiscs and Epistemology.*

MacBride, F. (1999). Listening to fictions: a study of Fieldian nominalism. *British Journal for the Philosophy of Science*, 50:431–455.

MacBride, F. (2003). Speaking with shadows: A study of Neo-Fregeanism. *British Journal for the Philosophy of Science*, 54:103–63.

MacFarlane, J. (2004). Double vision: Two questions about the Neo-Fregean programme. Online draft of March 24, 2004, http://philosophy.berkeley.edu/macfarlane/work.html.

Malamet, D. (1982). Review of Field. *Journal of Philosophy*, 79:523–34.

Margolis, E. and Laurence, S. (2001). Boghossian on analyticity. *Analysis*, 61(4):293–302.

McGee, V. (2005). Tarskis' staggering existential assumptions. *Synthese*, 142(3):371–387.

Miller, D. and Thorp, I. (1966). *Croquet and How to Play it.* Faber & Faber, London.

Pappas, G. (2005). Internalist vs. externalist conceptions of epistemic justification. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy.*

Parsons, C. (1965). Frege's theory of numbers. In Max Black (ed), *Philosophy in America*, (Ithaca:Cornell University Press, p. 180-203).

Parsons, C. (1979). Mathematical intuition. *Proceedings of the Aristotelian Society*, 80:145–68.

Peacocke, C. (1992). *A Study of Concepts.* MIT Press, Cambridge, Mass.

Peacocke, C. (1993). How are a priori truths possible? *European Journal of Philosophy*, 1:175–1999.

Peacocke, C. (1999). *Being Known.* Oxford University Press, Oxford.

Peacocke, C. (2000). Explaining the a priori: The programme of moderate rationalism. pages 255–285. In Paul Boghossian and Christopher Peacocke (eds.), *New Essays on the A Priori.*

Potter, M. and Smiley, T. (2001). Abstraction by recarving. *Proceedings of the Aristotelian Society*, 101:327–38.

Potter, M. and Sullivan, P. (1997). Hale on Caesar. *Philosophia Mathematica*, 5:135–152.

Prior, A. N. (1961). The runabout inference-ticket. *Analysis*, 21:38–39.

Putnam, H. (1971). *Philosophy of Logic.* University Press, New York.

Quine, W. V. (1935). Truth by convention. In Willard V. Quine, *The Ways of Paradox.*

Quine, W. V. (1940). *Mathematical Logic (1st ed.).* Harvard University Press, Cambridge, Massachusetts.

Quine, W. V. (1948). Two dogmas of empiricism. In Willard V. Quine, *From a Logical Point of View.*

Quine, W. V. (1953). *From a Logical Point of View.* Harvard, Cambridge, Massachusetts.

Quine, W. V. (1954). Carnap on logical truth. In Willard V. Quine, *The Ways of Paradox.*

Quine, W. V. (1966). *The Ways of Paradox.* Random House, New York.

Quine, W. V. (1986). *Philosophy of Logic, 2nd edition.* Harvard, Cambridge, Massachusetts.

Ramsey, F. (1929). Theories. reprinted in *Foundations* ed. D.H. Mellor, London, Kegan & Paul 1978.

Rayo, A. and Uzquiano, G., editors (2006). *New Essays on Unrestricted Quantification.* Oxford University Press, Oxford.

Read, S. (2000). Harmony and autonomy in classical logic. *Journal of Philosophical Logic*, 29:123–54.

Resnik, M. (1976). Frege's context principle revisited. published in Schirn(ed) *Studien zu Frege (vol III).*

Resnik, M. (1997). *Mathematics as a science of patterns.* Oxford University Press, Oxford.

Resnik, M. D. (1980). *Frege and the Philosophy of Mathematics.* Cornell University Press, Ithaca.

Ricketts, T. (1985). Frege, the Tractatus, and the logocentric predicament. *Nous*, 19:3–15.

Ricketts, T. (1986a). Generality, meaning and sense in Frege. *Pacific Philosophical Quarterly*, 67:172–95.

Ricketts, T. (1986b). Objectivity and objecthood: Frege's metaphysics of judgement. pages 65–95. in L. Haaparanta and J. Hintikka (eds) *Frege synthesized* Dordrecht: Kluwer.

Ricketts, T. (1996). Logic and truth in Frege. *Aristotelian Society Supplementary Volume*, 70.

Rosen, G. (1993). The refutation of nominalism. *Philosophical Topics*, 21:149–86.

Rossberg, M. (2006). *In Defence of Second-order Logic.* PhD dissertation, University of St Andrews.

Schirn, M. (1976). *Studien zu Frege (vol.I-III).* Friedrich Frommann Verlag, Stuttgart - Bad Cannstatt.

Shapiro, S. (1983). Conservativeness and incompleteness. *Journal of Philosophy*, 80:521–30.

Shapiro, S. (1984). Review of Field. *Philosophia*, 14:437–44.

Shapiro, S. (1991). *Foundations Without Foundationalism: A Case for Second-Order Logic.* Clarendon Press, Oxford.

Shapiro, S. (1997). *Philosophy of Mathematics: Structure and Ontology.* Oxford University Press, New York.

Shapiro, S. (2005). Where in the world wide web is the law of non-contradiction. unpublished.

Shapiro, S. and Weir, A. (2000). 'Neo-Logicist' logic is not epistemically innocent. *Philosophia Mathematica*, 8(3):160–189.

Sluga, H. (1975). Frege and the rise of analytic philosophy. *Inquiry*, 18:471–484.

Sluga, H. (1976). Frege as a rationalist. published in Schirn(ed) *Studien zu Frege (vol III).*

Sluga, H. (1977). Frege's alleged realism. *Inquiry*, 20:227–242.

Sluga, H. D. (1980). *Gottlob Frege.* Routledge and Kegan Paul, London.

Sosa, E. (2005). In defence of intuition. published in M. Bishop and D. Murphy, (eds) "Stich and His Critics", Blackwell Pubishers.

Stalnaker, R. (1996). What possible worlds could not be. pages 40–54. in Robert C. Stalnacker *Ways a world might be.*

Stirton, W. R. (2000). Hale's 'weak sense' is just too weak. *Proceedings of the Aristotelian Society*, 100:209–213.

Sullivan, P. M. (2004). Frege's logic. pages 671–762. in Gabbay and Woods (eds) *Handbook of the History of Philosophy and Logic volume 3.*

Suppes, P. (1957). *Introduction to Logic.* Van Nostrand, Princeton, NJ.

Tennant, N. (1997). *The Taming of the True.* Oxford University Press, Oxford.

Tugendhat, E. (1970). The meaning of 'Bedeutung' in Frege. *Analysis*, 30:177–189. Reprinted in Schirn(ed) *Studien zu Frege (vol III).*

Urquhart, A. (1990). The logic of physical theory. in Irvine(ed) *Physicalism in Mathematics.*

Weiner, J. (1990). *Frege in Perspective.* Cornell University Press, Ithaca.

Weir, A. (2003). Neo-fregeanism: An embarrassment of the riches. *The Notre Dame Journal of Formal Logic*, pages 13–48.

Wetzel, L. (1990). Dummett's criteria for singular terms. *Mind*, 99:239–64.

Williamson, T. (1988). Review of B. Hale, 1987, *Abstract Objects. Mind*, 98:487–90.

Williamson, T. (2000). *Knowledge and its Limits.* Oxford University Press, Oxford.

Williamson, T. (2003). Understanding and inference. *Proceeding of the Aristotelian Society Supplementary Volume*, 77.

Wright, C. (1983). *Frege's Conception of Numbers as Objects.* Aberdeen University Press, Aberdeen.

Wright, C. (1986). Inventing logical necessity. In J. Butterfield (ed) *Language, Mind and Logic* (Cambridge).

Wright, C. (1990). Field and Fregean platonism. In Andrew Irvine, *Physicalism in Mathematics*, Kluwer, Dordrecht/Boston/London, p. 7393 reprinted in Hale/Wright *The Reason's Proper Study.*

Wright, C. (1992). *Truth and Objectivity.* Harvard University Press, Harvard.

Wright, C. (1995). Critical notice of Michael Dummett's *Frege: Philosophy of Mathematics. Philosophical Books.* In Matthias Schirn, *Philosophy of Mathematics Today* reprinted in Hale/Wright *The Reason's Proper Study.*

Wright, C. (1997). On the philosophical significance of Frege's theorem. In Richard Heck, *Language Thought and Logic.*

Wright, C. (1998a). On the (harmless) impredicativity of Hume's principle. In Matthias Schirn, *Philosophy of Mathematics Today* reprinted in Hale/Wright *The Reason's Proper Study.*

Wright, C. (1998b). Why Frege did not deserve his *granum salis*: a note on the paradox of the 'concept horse'. *Grazer Philosophische Studien*, pages 239–63. reprinted in Hale/Wright *The Reason's Proper Study.*

Wright, C. (1999). Is hume's principle analytic? *Notre Dame Journal of Formal Logic*, 40. reprinted in Hale/Wright *The Reason's Proper Study.*

Wright, C. (2001). On basic logical knowledge. *Philosophical Studies*, 106:41–85.

Wright, C. (2002). Anti-sceptics simple and subtle: Moore and McDowell. *Philosophy and Phenomenological Research*, 65:330–48.

Wright, C. (2004). On epistemic entitlement. *Proceeding of the Aristotelian Society*, 78:167–212.