Markus Pantsar

Truth, Proof and Gödelian Arguments:
A Defence of Tarskian Truth in Mathematics

Philosophical Studies from the University of Helsinki 23

Filosofisia tutkimuksia Helsingin yliopistosta
Filosofiska studier från Helsingfors universitet
Philosophical Studies from the University of Helsinki

Publishers:
Department of Philosophy
Department of Social and Moral Philosophy
P.O. Box 9 (Siltavuorenpenger 20 A)
00014 University of Helsinki
Finland

Markus Pantsar

Truth, Proof and Gödelian Arguments:
A Defence of Tarskian Truth in Mathematics

ISBN 978-952-10-5373-3 (paperback)
ISBN 978-952-10-5374-0 (PDF)
ISSN 1458-8331
Tampere 2009
Kopio Niini Finland Oy

Contents

Acknowledgements

Although writing a monograph on philosophy of mathematics is largely a solitary pursuit, this project could not have been completed without the help of various people. I am very grateful to my supervisor Professor Gabriel Sandu, whose expertise was invaluable in leading my research down the right paths, and whose comments on various drafts of this work have been essential in giving it the current form.

I also owe gratitude to various members of the Department of Philosophy in the University of Helsinki for their help along the years. Professors Matti Sintonen and Ilkka Niiniluoto in particular deserve special thanks. With great gratitude I want to acknowledge the help of everybody who has at one point or another used his time to answer my questions. These include Professor Jaakko Hintikka, Dr. Philippe de Rouilhan and Professor Jouko Väänänen. I also want to thank my pre-examiners Professors Jan Woleński and Sten Lindström for their valuable comments, which helped improve the book a great deal at the final stage.

This project has been supported financially by the University of Helsinki, which is acknowledged with gratitude. I am thankful to all the personnel at the University of Helsinki who have helped me with the practical matters, in particular Dr. Ilpo Halonen, Dr. Panu Raatikainen, Auli Kaipainen, Kristiina Norlamo and Arto Kilpiö.

Thanks also belong to my family and friends for showing interest and support in the project, but most of all for providing the settings that allowed me to forget the thesis every once in a while. My brother Tuomas and his wife Tytti deserve special thanks for their support. In addition, I want to thank Matt and Jessica Powell for their help in revising the language.

Lastly, but in many ways most importantly, I want to thank my parents Lasse and Merja for their unconditional support in this project, and over my entire life. This thesis would never have been completed without their help. To show even some of the gratitude I owe to my parents, this book is dedicated to them.

Helsinki, March 2009 *Markus Pantsar*

1. Introduction

1.1 General background

Paul Benacerraf and Hilary Putnam (1983, p. 3) have divided philosophers of mathematics into those who "*promulgate* certain mathematical methods as *acceptable*" and those who "want to *describe* the *accepted* ones". This work will without doubt fall into the latter category. In general, I think philosophers should be careful about telling mathematicians how to do their jobs. This is not to say that the accepted results and methods of mathematics should be considered sacrosanct. Nor is it to say that philosophy cannot offer anything of interest to mathematicians. I disagree on both of these counts. There should always be room for healthy interaction between mathematicians and philosophers of mathematics. Nevertheless, the philosophical disposition of this work is definitely that of an anti-revisionist. After all, mathematical truth is the subject matter, and philosophical accounts of it should be careful not to neglect the way mathematics is actually practised. Here I am not interested in creating a new concept of mathematical truth as much as I am in explaining the one most of us already have, whether implicitly or explicitly.

As one consequence, interesting philosophical theories such as intuitionism will not be among the subject matter. Intuitionist mathematics is revisionist: it claims that there is something fundamentally *wrong* with mathematics as it is commonly practised. I cannot agree with that. I think that mathematics contains some of the most important knowledge that human beings have. In addition to examining this knowledge and the conditions for it, philosophers of mathematics should take heed of the way this knowledge is acquired and passed on. This is not to say that we should reject revisionist philosophies like intuitionism – but it *is* to say that we must have very good reasons for introducing them. As of now, we should be safe in using classical mathematics as the starting point.

There is an oft-quoted passage by David Lewis (1993, p. 15) about philosophical revisionism (the rejection of classes, in particular) in mathematics:

> …Mathematics is an established, going concern. Philosophy is as shaky as can be. To reject mathematics on philosophical grounds would be absurd. […] Even if we reject mathematics gently – explaining how it can be a most useful fiction, "good without being true" – we still reject it, and that's still absurd. […] I laugh to think how presumptuous it would be to reject mathematics for philosophical reasons. How would you like to go and tell the mathematicians that they must change their way, and abjure countless errors, now that philosophy has discovered that there are no classes? Will you tell them, with a straight face, to follow philosophical argument wherever it leads? If they challenge your credentials, will you boast of philosophy's other great discoveries: that motion is impossible, that a being than which no greater can be conceived cannot be conceived not to exist, that it is unthinkable that anything exists outside the mind…

Perhaps Lewis is being somewhat droll here, but as well as making clamorous exaggerations, I think he makes valid and important points. In contemporary philosophy of science there is a visible emphasis on what may be called the sociological aspect. Rather than following the Carnapian ideal of neatly structured formal scientific theories, we are now more convinced that the actual *practice* of science should also have its mark in the philosophy of science. Overall, this is a healthy development, even though it has sparked off less than healthy theories where philosophy of science has become a bastardized form of sociology of science. But even before this development, philosophers of science were noticeably reluctant to suggest much revisionism in, say, physics. They would discuss the sense in which concepts like "force" or "electron" exist and function, but they would generally not make much of an effort to contribute materially to the actual theories of physics. The success of physics has spoken volumes to philosophers. The two most important developments in 20th century physics, quantum theory and general relativity, have had a vast influence on philosophy of science in general; one cannot find a modern book on philosophy of science where these two developments have not

made a profound impact. Both the methods and results of modern physics have been crucial in giving us the modern paradigm for science.

For mathematics this has been different. Even if we neglect the numerous straight revisionist attempts (intuitionism being the most famous of these), there remains a distinct history in which philosophical theories of mathematics have not been required to conform to the practice of mathematics. This is not to suggest the connection between mathematics and philosophy has not been a two-way street. Certainly mathematical discoveries like non-Euclidean geometries have had an important effect on the philosophy of mathematics. Yet this effect is not comparable to the one that discoveries in physics have had. To mention an example that is certain to become familiar in this work, Gödel shattered Hilbert's dream of complete and uniform formalism, yet formalism in philosophy has retained considerable popularity after Gödel's results. Of course the philosophy of mathematics rarely *conflicts* with the actual results of mathematics. The phenomenon is more complex than that. It simply seems to be the case that the philosophers of mathematics do not give the same kind of status for their subject matter as the philosophers of physics do. They are more inclined to think that their object of study is a separate theory, something to be interpreted – even revised – philosophically. The actual methods used by mathematicians, in particular, seem to carry little or no importance for surprisingly many philosophers of mathematics.

Here the quote from Lewis has the most relevance. We seem to have a great deal of humility toward the methods and practices of physicists, but in mathematics we reserve a different, much more powerful and revisionist, role to philosophy. It is hard to see the reasons behind the difference in approaches. Perhaps it is because most philosophers of mathematics are more familiar with mathematics than philosophers of physics are with their subject. Modern physics requires, as well as a great deal of expertise, access to a lot of expensive equipment. Mathematics, for the most part, only requires the expertise. In this way most philosophers cannot understand the nature of modern physical inquiry as well as the nature of mathematical inquiry. As a result, philosophers of

mathematics will be more confident to use their expertise to suggest revisionist theories about mathematics – or so goes one theory concerning the implicit consensus opinion. Whether the theory is right or wrong, its implication is hardly something we should be ready to accept. I think that philosophers should apply to the practice of mathematics the same kind of humility as they do to that of physics.

Is there any philosophical relevance to all this? After all, mathematical practice is not perfect, and it is sensible to think that philosophical theories of mathematics should be evaluated on their own merits. If a revisionist philosophy is needed, then that is what we should focus on. This is of course true, but way too simplistic. As a result of constantly surfacing mathematical revisionism in philosophy, it is easy to forget that none of the revisionist theories have had *even nearly* the kind of success that classical mathematics has enjoyed. I dare to say that extremely rarely, if ever, has *philosophical* revisionism meant development in the practice and results of mathematics. Whatever conventions we have in mathematics, and whatever their origin may ultimately be, they have proven to be highly successful. If we were to reject these conventions, it would have to be for a thoroughly convincing reason – and it is an understatement to say that no such reason has been established so far. Against this background, we can only use the starting point that our best contemporary theories of mathematics are the best theories of mathematics there are. [1] They are not perfect, of course, but it is safe to assume that the flaws are more likely to be resolved by mathematicians than by philosophers. Any revisionism over mathematical practices and concepts must be handled with this in mind.

One such concept is *truth*. It is one of the most fundamental underlying conventions in mathematics. Take a textbook of elementary logic, and one finds it full of expressions like "truth-

[1] This is not to say that this work is limited only to classical mathematics. Subjects such as Hintikka's and Sandu's IF logic and Kripke's arguments applying Kleenean logic will also be addressed, especially their implications for the question of truth.

tables" and "truth-preserving inferences".[2] It is the latter that we are after in mathematics when we define rules of *proof*. Rules of proof are, of course, the way we reach theorems from axioms and as such they are the most important tools of a mathematician. That is the implicit near-consensus we have about mathematics: it is a systematic pursuit of *true* sentences, starting from true axioms and using truth-preserving rules of proof. However, when we move to philosophy, this suddenly becomes highly controversial. One of the most important forms of revisionism in philosophy of mathematics of the latter part of 20th century has been extreme (strict) formalism (nominalism), and its ontological conclusion, Hartry Field's (1980) fictionalism. According to it mathematical objects do not exist, and the formal axiomatic systems that form the core of mathematics do not refer to anything outside them. In other words, for the extreme formalist rules of proof and axioms are *all* there is to mathematics. Instead of defining proof as truth-preserving inferences, we simply start from the rules of proof. In this way truth for the strict nominalist is *deflationary*, an empty concept: whatever we mean by it, it is already included exhaustively in the concept of proof.

When discussing extreme formalism, one question of course comes immediately into mind: without any outer reference, how do we know which axioms and rules of proof to accept? One main purpose of this work is to show that we do not. In this work that is called the problem of *theory choice*, and I will try to show it to be the most fundamental problem with strict formalist philosophy of mathematics. Simply put, I will argue that when taken to its logical conclusion, extreme formalism implies completely arbitrary mathematics: we would have no reason to prefer one set of axioms and rules of proof over another. That is a staggering conclusion, but we will see it is the only one that can be plausibly made if we reject *all* outer reference from mathematics. Fortunately it never comes to that, since mathematics without any outer reference does not make sense. We need to explain why we prefer some rules of proof and some axioms to others, and without any concept of reference this cannot be done. In this work I will argue that

[2] One can use Suppes 1959 as a point of reference.

without any outer reference, mathematics as we know it could simply not be possible: it could not have developed, and it could not be learnt or practised. Sophisticated formal theories are the pinnacle of mathematics but, philosophically, they cannot be studied separately from all the non-formal background behind them.

This way, what might seem like a completely formalist theory of mathematics turns out to be nothing of the sort. It could not have existed without a wide *pre-formal* background, which we will see when we examine mathematical practice in general.[3] Formal systems are not of the self-standing type that extreme formalism seems to claim. My purpose in this work is to show that the formalist program uses the actual practice of mathematics as a ladder that they later discard. This by itself is of course perfectly acceptable, and it mirrors the way we strive for formal axiomatic systems in mathematics. What is not acceptable is how they refuse using the ladder.

When it comes to the question of truth and proof, this could not be any more relevant. The deflationist truth of extreme formalism equates mathematical truth with formal proof. However, as we will see, that strategy requires that we take mathematics to concern only formal systems. Once we look at the wider picture, we see that outer criteria are needed to avoid arbitrariness. Theory choice must be explained, and this requires reference outside formal systems of mathematics. Philosophers have tried to explain this by a wide array of concepts – usefulness, assertability, consistency and conservativeness, to name a few – but ultimately none of them have been satisfactory. The only plausible way to answer the problem of theory choice, I will argue, is by appealing to *truth*.

---

[3] What I refer to as pre-formal mathematics in this work is more often discussed as *informal* mathematics in literature. The choice of terminology here is based on two reasons. First, I want to stress the order in which our mathematical thinking develops. We initially grasp mathematics through informal concepts and only later acquire the corresponding formal tools. Second, the term "informal mathematics" seems to have an emerging non-philosophical meaning of mathematics in everyday life, as opposed to an academic pursuit – which is not at all the distinction that I am after here.

1.2 Another approach

In the end, all of the above comes down to the question of reference. If we follow extreme formalism in that mathematical theories have absolutely no outer references, we will end up with the position that mathematics is arbitrary fiction. Deep down, under this interpretation, going through a mathematical proof is similar to solving a Sudoku puzzle. Although this goes against the image most of us have about the nature of mathematics – as well as all the practical applications – the formalist program has one clear strength: it avoids the daunting ontological problems we are faced with in the philosophy of mathematics. If we accept that mathematical theories have references, the understandable consensus is that we must specify what these are. On this matter, however, non-formalists have found very little to agree on. Platonism, structuralism, empiricism, naturalism and many other suggestions have been presented – and all of them have been shown to be problematic in one way or another. The conclusion for strict formalists has been that references in mathematics are not possible, and mathematics must be a fiction. In particular – against the main thesis of this work – mathematical truth is deflationary.

The approach for the extreme formalist, hence, is to minimize the ontological commitments in order to make mathematics as philosophically unproblematic as possible. In this work I want to suggest another approach, one that is necessitated by the failure of extreme formalism. While ontologically minimal, extreme formalism makes mathematics impossible as a *human endeavour* – which is much more alarming than any intricate philosophical problems. In a nutshell, I will argue that if extreme formalism were correct, mathematics could not have developed in the first place – nor could it be practised today. It must not be forgotten that mathematics is a human endeavour just like all other sciences. If something is essential to mathematics as a human endeavour, we would seem to have good reason to believe it is also a factor in the philosophy of mathematics – or at least something we should expect a theory in philosophy of mathematics not to *conflict* with. As well as providing an explanation for the formal theories that are the core of mathematical knowledge, philosophical accounts of

mathematics must be able to explain why we prefer certain theories to others, why they are useful in practice, and how we are able to teach and learn mathematics. When it comes to mathematics as a science, this is of course something everybody is ready to agree on. In fact, it is so obvious that most philosophers of mathematics seem content not to grant any importance to it. For the majority of philosophers, mathematics seems to consist of formal systems – often using Peano arithmetic (PA) as the example – and the philosophy of mathematics concerns the ontological and epistemological status of these systems.

As central as those questions are, to me they only seem to cover half the picture. It is obvious that besides formal systems, mathematics as a human endeavour has a large informal element. Textbooks of mathematics are not written in completely formal languages and all kinds of informal examples are used in learning mathematics. The communication in mathematics is facilitated everywhere by informal elements. Indeed, it should be safe to say that in order to understand mathematics, we as human beings *must* use these informal elements. In addition, the history of mathematical thinking of course reveals that formal axiomatic systems of mathematics are a rather late development. The Peano axiomatization of arithmetic, for example, was only published in 1889, millennia after arithmetic was first used to great success. These informal – *pre-formal* – elements have made mathematics possible to use and learn whether we consider individual or the wider historical development.

Yet the pre-formal element has been largely neglected in the philosophy of mathematics. It has been widely assumed – and not just among formalists – that these are matters for psychology and sociology, and not of much interest to philosophers. In this work I must argue against that. These pre-formal elements are the very reason why mathematics makes sense to us. Not surprisingly, they also have a central position in the whole problem of mathematical reference. When we acknowledge that formal theories have been *designed* to correspond to our pre-formal mathematical ideas, we immediately recognize that the latter are in fact the reference of formal mathematics. Rather than think of, say, the natural numbers as defined by the axioms of PA as fiction, we can consider them

referring to our pre-formal notion of number – and arbitrariness is avoided.

That is the first stage of mathematical reference, and when we speak about the truth of formal mathematical theories, at this first stage we are concerned with them *corresponding to our pre-formal ideas*. Of course, in order to avoid arbitrariness, the pre-formal ideas themselves must have references, and that second stage is the question of Platonism, structuralism and other ontological theories. In a way, by introducing pre-formal thinking into philosophy we are admittedly only moving the problem of reference to another level. However, this is giving the strict formalist too strong a case. I will argue that the non-formalist does not need to specify her ontological and epistemological positions. All she needs to show is that *some* theory of reference – and truth – is needed in the second stage for a philosophical theory of mathematics to make sense. In this work I defend Alfred Tarski's (1936) T-scheme as a theory of truth fitting both of these two stages. Tarskian truth is *semantical* and the connection of formal and pre-formal mathematics seems to be a semantical one, as well: we understand formal sentences by what they *mean* pre-formally.

It will be seen that Tarskian truth in the first stage – over formal mathematics – is not deflationary. What Tarskian truth in the second stage refers to is a whole other question – but it is also one we do not need to answer in order to refute extreme formalism and deflationism. There exists a reference for formal mathematics, and when it comes to the question of truth and proof, it will be enough to complete the argument here to show that there must exist one for pre-formal mathematics, as well. If we examine mathematics as a wider phenomenon, we will see that there is only one philosophical theory of mathematics that conflicts with this – and that is extreme formalism with its irrevocable problems of arbitrariness. Other than repudiating that kind of strict formalism, I will argue, the deep ontological questions of the second stage can be left unanswered in a work about truth and proof.

1.3 Truth and proof

The denial of deflationism leads us to a difference in the concepts of truth and proof. For the strict formalist, truth and proof are of course equivalent concepts: a sentence of mathematics is true if and only if it is provable, and as such truth is a deflationary concept. But it is not often acknowledged that even for the Platonist this is the case for the most part – practically for all sentences. After all, proof is the way we establish true sentences in mathematics, whatever our philosophical leanings may be. The possibility of errors in our proofs aside, there are only two kinds of sentences where truth and provability do not coincide for the Platonist. The first type is formed by unsolved problems like Goldbach's conjecture. Platonists, and most other non-formalists, believe that even though we have neither proved nor disproved them, they are either determinately true or determinately false. The extreme formalist, on the other hand, believes that sentences only become true or false once we prove or disprove them; before that we cannot think of them having determinate truth-values.

Unsolved problems are of course the main challenge for many mathematicians, but as far as mathematical sentences go, they only form a minuscule set. For the most part, the difference between the formalist and non-formalist cannot be seen in the extension of truth, that is, the classes of true and false sentences. Both the non-formalist and formalist believe that we acquire true sentences by proof, and apart from revisionists like intuitionists, they also agree on the rules of proof. The main difference is that while the Platonist thinks our proof methods are *valid* means of finding out true sentences, the formalist believes that truth and proof are the same concept *by definition*. Still, the extensions of true and false sentences are almost completely the same for the formalist and non-formalist. Mainly, it is only in the intensions that there is a difference, in addition to the status of yet unsolved problems. While the latter is a philosophically interesting question, the former is *all* about philosophy. For the practising mathematician it does not make much of a difference whether we consider truth to be deflationary or not. This makes the intension of truth seem unproblematic mathematically: if we had a way of solving all the

problems of mathematics, the extensions of truth and proof would match. In such circumstances, it would be difficult to argue for a non-deflationary notion of truth based on anything other than purely philosophical basis.

That brings us to the second type of sentences that cause a difference between the non-formalist and formalist: undecidable sentences. While Goldbach's conjecture could be proved in the future, there are sentences that cannot be proved or disproved even in principle. Famously, Kurt Gödel (1931) proved that in *every* consistent formal system containing arithmetic there are such sentences. That already by itself is mathematically and philosophically highly interesting. Consistent formal systems are always incomplete. But the real philosophical catch is that such *Gödel sentences* can also be seen to be *true*. In short, given some very reasonable basic truth-theoretic assumptions, the Gödel sentences are true but unprovable. This way even the extension, as well as the intension, of truth will always differ for the formalist and non-formalist. That is why Gödel's incompleteness theorems are absolutely essential to the question of truth and proof in mathematics: they give us the only known *explicit* case of a difference between truth and proof. If that indeed were the case, it would already show that truth is a substantial, not a deflationary property.

However, when we say that the Gödel sentences are true, we are obviously talking about truth in a context different from proof in formal systems. From the first glance it is obvious that we mean semantic truth: looking at the construction of Gödel sentences we see that they have the semantic content: "this sentence is unprovable", which indeed is the case by Gödel's proof. That is what we mean by the truth of Gödel sentences: they are true through their meanings. But this is something seemingly very different from the rigid rules of proof we are accustomed to in mathematics, and it immediately raises two questions. First, if not in the original formal systems, in what kind of expanded systems do we establish the truth of Gödel sentences? Second, are we entitled to call such semantic properties *truth* in mathematics?

The apparent truth of Gödel sentences was already noted by Gödel himself, but he left open the question of the underlying

conditions concerning truth. It is from the later work of Stewart Shapiro (1998) and Jeffrey Ketland (1999) that we get an exact argument concerning the whole process of establishing the truth of Gödel sentences. These are called *semantical arguments* in the literature, based on the fact that the truth of Gödel sentences is established in a Tarskian semantic expansion of the formal system. In this work it is through them that we will get into the bottom of the question of truth and proof in mathematics. Keeping in mind the existence of pre-formal element in mathematics – and how it is semantical – we should be fully justified in taking on such a project.

1.4 Tarskian truth

So what can be our motivations for expanding formal systems of mathematics with Tarskian truth, or indeed to include a truth predicate in the first place? These are both important questions, since Tarskian truth is very problematic for the formalist project – even if we ignore the semantic arguments concerning Gödel sentences. Against extreme formalism, Tarski's T-scheme for truth presupposes a relation between languages and the objects they refer to. In Tarski's own words:

> *Semantics* is a discipline which, speaking loosely, *deals with certain relations between expressions of a language and the objects* (or "states of affairs") "*referred to*" *by those expressions*. (Tarski 1944, italics in the original)

This kind of relation can be of two kinds: either it is *language-to-language*, or *language-to-world*. In Tarski's conception the definition of truth for an object language L_1 must be given in a metalanguage L_2. The truth predicate for L_2, in turn, must be given in another language L_3. Famously, Tarski (1936) proved that no classical formal language could contain its own truth predicate, due to Liar's paradox.[4] As such, if we want to include a truth predicate,

[4] "This sentence is false."

we are committed to a hierarchy of languages. Moreover, if consisting only of formal languages, this hierarchy does not *collapse*: at no level will a language L_m provide a truth predicate for a language L_n, where $n \geq m$.

Hence, if one is restricted to classical first-order two-valued formal languages, he must also be committed to an infinite hierarchy of languages in order to include a truth predicate. More specifically, the strict *formalist* philosopher of mathematics is committed to this – for him all the relations between languages and objects are in fact relations of the language-to-language type, as there exists no outer reference for the sentences of mathematics. Tarskian truth turns out to be highly problematic for the formalist, and that is why we need to have a clear justification for its introduction. Perhaps the most common anti-Tarski argument goes: "why do we need such a semantic truth predicate in the first place?" That question is particularly relevant in mathematics. This is something we must be able to deal with, and in this work I will use the existence of pre-formal thinking to justify it. Tarskian truth, I will argue, introduces nothing new when we consider the full picture of mathematical thinking. Keeping this in mind, it is not surprising that strict formalism runs into problems like infinite hierarchies with it, remembering how it only considers one part of the whole phenomenon of mathematics.

If one is not committed to strict formalism, there are far less problems with Tarskian truth. In particular, the hierarchy of languages can be collapsed. There are two ways of doing this. One can either move from formal to *informal* languages – where Tarski's undefinability result does not hold in the strict sense[5] – at some point in the hierarchy, or one can hold some level in the hierarchy to be of the language-to-world type. Philosophically these two strategies are largely equivalent, since we seem to have no way of describing the world outside language. This makes the job a lot easier for the non-formalist. Rather than try to explain a problematic relation between mathematical languages and

[5] Of course the Liar's paradox also exists in informal languages like English, but at the same time we also get new tools to recognize it as a paradox.

mathematical reality, we can concentrate on characterizing the connection between our formal and pre-formal mathematical languages. This will already give us enough tools to go through everything needed for a Tarskian expansion to formal systems. Moreover, I will argue, explaining that connection in fact *requires* Tarskian truth.

1.5 Reference

However, we cannot escape the language-to-world relation completely. In fact, the above strategy only works if such a relation is assumed to exist. If we do not postulate any relation between a language and some reference outside the language – that is, names and the objects they denote – there does not seem to be any difference between formalism and non-formalism. Why should we worry about an infinite hierarchy of languages if there is no alternative? If languages are all there is to mathematics, surely there is nothing ontologically or epistemologically problematic in defining new languages, even *ad infinitum* – in principle, even though obviously not in practice. Seen this way, the infinite hierarchy of languages would simply be a fact concerning classical formal languages. For the strict formalist this is certainly extremely inconvenient – after all, we are talking about an *infinite* hierarchy here – but it would not be enough to refute his position, if formalism were otherwise feasible.

What, then, is the biggest problem for the formalist? It turns out to be the exact thing that is also the main strength for him. As we know, strict formalism requires no ontological commitments. Only formal mathematical systems exist, and mathematics is the study of them, and only them. Ontologically this is obviously as economical a position as there can be. But we should not be blindly concerned only about the ontological status. There are two important ways in which the lack of reference is devastating for the formalist program.

The first one of these is the relation to informal languages. In this work I will stress the importance of pre-formal thinking, and the role of Tarskian truth in its connection to formal theories. In

this way, mathematical truth is postulated primarily as the Tarskian relation between formal and pre-formal mathematics. What proof is to formal mathematics, truth is to pre-formal. We deal with mathematical proofs syntactically, but at the same time we as human beings think about them semantically. We cannot deny pre-formal thinking, and its need for semantical truth. However, this alone is not enough to show a substantial difference between truth and proof. Even though the existence of pre-formal mathematics cannot be reasonably contested, there is always the possibility that when it comes to truth, it is essentially *superfluous*; whatever we can achieve with truth, we could also achieve with proof alone. This is where the semantical arguments of Shapiro and Ketland have the most importance. They show that Tarskian truth as an expansion to formal languages is not *conservative*. Adding Tarski's T-scheme to a formal system **T** will give us a *new* true sentence, one that wasn't true by proof in **T**. This is of course the Gödel sentence of **T**. With Tarskian truth we establish new truths, and that will prove to be extremely problematic for the strict formalist, especially since the introduction of pre-formal thinking into philosophy also serves as the perfect justification for introducing Tarskian truth.

The second problem that the lack of reference causes for formalism is one that does not require semantical arguments, or indeed any sophisticated philosophical devices. It is a fact that mathematics is very important to other sciences, perhaps even – as in Quine's (1995, p. 40) thinking – indispensable. Science as we know it could not exist without mathematical theories. But the matter goes even deeper than that. It could be plausibly claimed that *human thinking* as we know it could not exist without some mathematical knowledge. There are numerous mathematical applications in everyday life that have nothing to do with the modern theories of physics, or indeed any other science. Even the least mathematical among us constantly express ideas with quantities and geometrical objects. Yet we can be surprisingly quick to forget that we only employ a very small fraction of all the possible mathematical theories, whether in science or in everyday life. The sum of 2 + 2 could be any natural number, yet we only use theories where 2 + 2 = 4. But if mathematics has absolutely no

reference, what reason do we have for picking one theory over another? It must be remembered here that this reference does not have to mean anything resembling a Platonic universe of mathematical ideas. Simply put, if we believe that 2 + 2 = 4 rather than 2 + 2 = 3, we must believe in *some* kind of reference.[6]

So there must be something, whether we know it or not, that makes us think of the reference of 2 + 2 as the same as that of 4, rather than that of 3. Obviously it is the Peano axiomatization (or some other axiomatization of arithmetic) that does the job for mathematicians, and that has lead the extreme formalists to the strange conclusion that Peano axioms could be all there is to arithmetic. In this way mathematical theories are thought of only as conventions, with nothing to tie them to any outer reference. But this means ultimately arbitrariness and the Peano axioms are definitely *not* arbitrary: they were carefully designed to correspond to our best arithmetical knowledge. If they were arbitrary, either we could change them without it damaging our arithmetical thinking and its applications, or we have by sheer chance stumbled, from all the infinite number of possible axiomatizations, upon one that works so brilliantly for us.

Mathematics without any reference means arbitrary theories, and this must never be forgotten. One common extreme formalist strategy is to claim mathematics to be a fiction, but praise it as a "useful" fiction.[7] But by itself this means nothing. Quite clearly mathematics is useful, but the real question is *why* some theories are more useful than others. I will argue that even the slightest tendency toward choosing one theory of arithmetic over another can only be justified on the basis of some theory of reference. Of course with actual mathematics we go far beyond such tendencies: most mathematicians would be ready to assert mathematical theories to be *true*, and the objects denoted in them to exist. But even if we do not accept such viewpoints, it is impossible to accept arbitrariness. As far as mathematics as a human endeavour is

---

[6] It must be noted that I do not mean to use "some" as a hedge word here. My point throughout this work is that the relevant dichotomy is reference against *no* reference, rather than no reference against *Platonist* reference.

[7] See Field 1980, p. 15.

concerned, I claim that arbitrariness is the worst possible situation for a philosopher, and it is one of my main purposes here to show that this is exactly what the strict formalist will end up with. In this work I will not try present a comprehensive epistemological and ontological theory of mathematics to *replace* formalism, and in that way the approach taken here is largely destructive. However, I will argue that to introduce Tarskian truth and reference, we only need the knowledge that *some* such non-formalist theory must be accepted – and that is equivalent to rejecting extreme formalism and the arbitrariness of mathematics.

What arguments can we have for strict formalism in which mathematics consists only of conventions? Formalists often point out the introduction of non-classical logics and non-Euclidean geometries as examples of the conventional nature of our mathematical knowledge. What were once thought to be necessary truths have more recently been found out to be, at least in the non-Euclidean case, just unnecessarily limiting postulates. However, this as such is no argument for conventionalism. Obviously most non-formalists do not claim that our mathematical knowledge is *perfect*. A many-valued logic could turn out to be less problematic than our classical two-valued one, but this is different from saying that the rules of classical logic are arbitrary. Likewise, non-Euclidean geometries turned out to have important physical applications, but this does not mean that Euclidean geometry is not correct in its own realm of planes and three-dimensional spaces. Mathematics develops, just like any other science. Physicists no longer claim that Newton's physics is true, but there is no denying that it is *closer* to truth than, say, Aristotelian physics. Why should we think any differently when it comes to mathematics?

1.6 Non-classical languages

While the overall attitude in this work is a non-revisionist one, this should not be thought to influence the arguments by default. Whenever a revisionist theory seems to have more potential when it comes to the question of truth and proof, it will be taken into consideration. For example, even though this work uses classical

mathematics and two-valued logic as its starting point, the results will not be limited to them. Since Gödel's incompleteness theorems have an important role, obviously the arguments based on them will not be applicable to mathematical languages in which the incompleteness results do not hold, or take a significantly different form. But that will only be one part of the arguments here. In the realm of classical logic Tarski's undefinability result shows us that formal languages cannot contain their own truth predicates. Gödel's first incompleteness theorem gives us an explicit case of an unprovable sentence, and with Tarskian truth we can establish that the unprovable sentence is in fact true. This is how the semantical argument goes, and it shows that mathematical truth is not deflationary in two-valued first-order languages. Consequently, it might seem that without Gödel's result the deflationist would not be harmed, and a many-valued or higher-order logic could provide him with a refuge.

However, the matter is not as simple as that. Although Gödel's theorem gives us an explicit case of the difference between truth and proof, the difference itself exists independently of Gödel's results. Truth is not deflationary, and Gödel sentences are a *symptom*, not the cause of it. In this Tarski's undefinability result is essential. If formal systems cannot contain their own truth predicates, we must commit to an infinite hierarchy of formal languages in order to include a truth predicate. This by itself is highly problematic for the formalist: for one, it would already demolish the esteemed ideal of presenting mathematics (or at least important parts of mathematics) as a *single* formal system. But here many-valued logics may have stronger potential. Saul Kripke famously used Kleenean logic to define a truth predicate for a language L, within L. More recently Jaakko Hintikka has claimed that his and Gabriel Sandu's IF-logic can escape Tarskian undefinability and the need for a hierarchy of languages. These are important developments in the question of mathematical truth, and they – as well as higher-order two-valued logics – will be given a detailed account in this work.

1.7 The basic theory of mathematics

In a work like this where mathematical truth is claimed to be the subject matter, it is important to clarify just *which* mathematical theories we are discussing, and how the results can be applied to other theories. Here arithmetic, in particular the first-order Peano axiomatization of it, is considered to be a suitable candidate for the first, most basic, theory of mathematics. The choice is only a matter of convenience, and no philosophical importance should be attached to it. Gödel used arithmetic to clarify the set of assumptions needed to go through his proof of incompleteness. Since Gödel's work is central for the purposes here, it makes sense to follow him and study formal mathematical systems containing arithmetic. In a strict sense, the subject here could be actually seen as *arithmetical* truth and not the wider concept of mathematical truth in general.

However, I claim that this is not the case. Arithmetic is used as an example, but the conclusions should be more or less directly applicable to other mathematical subjects. From time to time I will use logic, set theory and geometry in examples, which should widen the scope, but mostly this work will remain a study of arithmetical truth. Should this be considered a weakness? Certainly arithmetical truth by itself is important enough as a subject, but I am also confident that most of this work concerns a wider range of mathematical theories. Pre-formal mathematical thinking, for example, is a general phenomenon. In fact, aside from the Gödelian arguments, nothing in this work concerns *only* systems containing arithmetic – and most of the interesting mathematical systems contain arithmetic, anyway. This question will not be given much attention later on – I trust that the reader will agree about the projection of the results here into other fields of mathematics. If not, I am happy to concede that the scope of this work is in fact arithmetical truth.

Nevertheless, arithmetic as the primary mathematics is by no means a consensus choice, and a word should be said about that. It could be claimed that the semantic argument, for example, cannot be used if we do not choose arithmetic as our basic mathematical theory. Basically, there are two competing choices for primary

mathematics: logic and set theory. Of course we could first define natural numbers via sets, or follow Frege's and Russell's logicist ideal, but that would not change anything of substance here. We would still be defining natural numbers, and the conclusions would be exactly the same.[8] In this work I do not claim that the objects of arithmetic – rather than those of set theory or logic – exist, or that the sentences of arithmetic are more likely to have objective truth-values. As was said, arithmetic is used for Gödel's sake – nothing further is meant by that choice.

1.8 The limitations of the approach here

Above I have presented the rough philosophical and mathematical framework for this thesis. Later on, more exact formulations will be given for all the central concepts of this introduction. However, some important themes – like pre-formal mathematical thinking – are too complex and intractable to be given exact definitions. Any such effort would be artificial, and even if successful, more or less achieved by smoke and mirrors. Even though pre-formal thinking and the reference of mathematics will be crucial concepts in this work, I do not claim to give anything close to a comprehensive and consistent description of them. As for pre-formal thinking, it undoubtedly requires a psychological theory behind it to be fully clarified. For the reference of pre-formal sentences, we would need to answer all the metaphysical and epistemological problems in philosophy of mathematics. In other words, to complete the theses here, I would not only need to explain the epistemology and ontology of mathematics, but the psychology of it, as well.

Fortunately, no such thing is needed for the task at hand: the theses in this work are valid if we agree that there is *some* form of semantical pre-formal thinking, and mathematics has *some* reference. Obviously the best way to show the latter is by showing the opposite viewpoint, extreme formalism, to be absurd. As for the former, I think only a minimum of evidence is needed, and that evidence is available to all of us with access to textbooks of

[8] See Enderton 1977 for ways of doing this.

mathematics – *any* textbooks of mathematic. It is simply a fact that mathematics is universally taught with the help of all kinds of informal tools, from the use of visual aids to verbal explanations. As I see it, this is actually much better evidence for pre-formal thinking than any single psychological theory could be. Theories about learning mathematics have been notoriously fallible, but no feasible psychological theory can suggest that human beings process mathematics completely formally.

Simply put, the aim of this work is to show that mathematical truth is a substantial, not deflationary, property. I will try to show this with the very minimum of psychological, ontological and epistemological burden – which means that I will not use arguments based on any particular ontological theory in philosophy of mathematics, or any psychological theory. That will be seen in the negative nature that the answers to questions like mathematical reference will have. The main argument against strict formalism is that no reference makes no sense, and hence we must accept some theory of reference, whatever it may be. This could mean Platonism, but it could also mean a weak form of naturalism or empiricism. Although in the end of Chapter 6 I will introduce an outline of an ontologically and epistemologically unproblematic account of non-formalist philosophy, ultimately in this work I must leave that question open.

Similarly, for pre-formal mathematics I will suggest an account – including some tentative examples – but ultimately I want to leave the psychological questions of mathematics open. My account of pre-formal thinking may be faulty, and the examples mistaken, but that is not the point here. These are only introduced as quasi-heuristic means of clarifying my position concerning formal and pre-formal mathematics. Granted, if I believed there to exist thoroughly convincing psychological theories that were directly relevant to the subject here, I would be tempted to use them – even if it meant limiting the applicability of this work. The same goes for philosophical theories of ontology and epistemology of mathematics. The lack of such theories in both fields makes that temptation easy to resist – hence I will argue for my position from the very minimum of evidence. For reference in mathematics this means trying to refute extreme formalism. For pre-formal thinking

it means simply noting the fact that all mathematical communication between human beings includes informal elements.

1.9 The structure of this work

The main thesis of this work is that mathematical truth is a substantial property, and hence truth and proof are different concepts. We will need quite a few steps to arrive at that conclusion, and what follows is a rough outline of the structure that the argumentation takes. In Chapter 2 I present a brief philosophical and mathematical background for the work, including definitions of the central philosophical concepts and introduction to the relevant content of Gödel's incompleteness theorems and Tarskian truth. In defining such a central philosophical concept as realism as "anti-formalism", I have tried to make the formulations suitable for this work in particular. I hope that any unorthodoxy in the definitions will be seen to be justified later on when the whole context is revealed.

Chapter 3 consists of an overview of the recent debate concerning semantical arguments for the substantiality of truth. The approach here has been to make a critical semi-chronological presentation of the arguments from all sides: Hartry Field, Stewart Shapiro, Jeffrey Ketland and Neil Tennant. Only the aspects of the debate relevant to this work are considered, and all along the presentation I will be taking a critical part into the discussion. In the end of the chapter I will present my own conclusions, according to which the semantical argument of Shapiro and Ketland is valid and by expanding formal systems with Tarskian truth we can establish the truth of Gödel sentences. Thus there are true sentences that are not provable, and truth and proof are different concepts. However, Tennant is also correct in stating that there are other ways than Tarskian truth to arrive at the same result. That is why the key question is not the validity of semantical arguments, but the *plausibility* of the expansion we use. Here Tarskian truth beats Tennant's arbitrary-looking soundness principle – in fact, I will argue that Tarskian truth is not an

expansion at all, but rather a very natural part of mathematics once we recognize that in addition to the formal part, human beings also use pre-formal mathematical thinking.

In Chapter 4 the phenomenon of mathematical thinking is studied in a wider context in order to justify the conclusion made in the end of Chapter 3. The purpose is to put formal mathematical systems into their own proper place: as a crucial achievement and perhaps the ultimate tool of mathematics, but far from being the complete picture. Mathematics is a human endeavour, and we must not ignore the way mathematics is practised, learnt and taught. We as human beings use pre-formal – semantical – mathematical thinking all the time, and this enables us to understand mathematics. Human beings do not process mathematics completely formally as computers do. We comprehend mathematical ideas in our pre-formal thinking, and the formal theories are a way of making these ideas maximally unambiguous. *Proof* is of course the method by which we acquire new theorems in the formal systems, but the rules of proof cannot be arbitrary. They have been designed to correspond to our pre-formal ideas of *truth*. It is in this domain of pre-formal thinking that we see the truth of Gödel sentences. As the semantical arguments show, Tarskian truth is all we need for that, and it corresponds well with the pre-formal thinking in mathematics. That is why the semantical arguments are valid, and mathematical truth is substantial. Of course this would be the case even without Gödel's incompleteness theorems and the semantical arguments; their importance lies in giving us an explicit sentence to study the problems with.

All the arguments up to and including Chapter 4 presuppose the use of classical two-valued first-order logic. From Tarski's undefinability result we know that such languages cannot contain their own truth predicates, and to include truth we must commit to a hierarchy of languages. In Chapter 5 we are concerned with other logics as the basis for mathematics. Three logical theories in particular are studied for the possibility of including a truth predicate for a language L within L: two-valued second-order logic, Saul Kripke's use of Kleenean many-valued logic, as well as Jaakko Hintikka's and Gabriel Sandu's IF logic. As the most

modern development, IF logic is given the most detailed treatment, and many of the problems with the other approaches can already be seen in it. In an IF-language one can indeed have a materially adequate truth predicate for that language. The problem is that we cannot *show* this predicate to be such within IF logic. The only successful approach would be to define, and to be able to recognize, the truth predicate completely within the used language. This problem is called, after Philippe De Rouilhan's and Serge Bozon's suggestion, the *monolingual speaker problem*. In order to avoid hierarchies of languages, a completely monolingual speaker of a language would have to be able to establish everything needed for the truth predicate for that language. This is not possible in any of the three approaches. In addition, all of the approaches have their own specific problems, ranging from the intractable concept of logical consequence in second-order logic to the use of set theory in Kripke's argument. The hierarchies of languages cannot seem to be escaped, even with many-valued or higher-order logics. However, in the end of the chapter I will argue that with the help of pre-formal languages we can collapse the hierarchy at any point. Moreover, this is what we *actually* do when talking about truth in mathematics.

For deflationist truth there is one remaining haven: extreme formalism, or strict nominalism, that Field's fictionalism suggests. If we insist that formal systems and their rules of proof are all there is to mathematics, we cannot discuss Tarskian – or any other – expansions that require reference for formal mathematical sentences. In such context pre-formal mathematics would be philosophically superfluous, and all the arguments in this work vacuous. However, with *all* the other philosophical theories of mathematics there is room for reference, and hence also room for Tarskian truth. In Chapter 6 I will try to show that extreme formalism as a philosophical position is untenable. The most important question here is ultimately that of reference. If we deny all reference for formal mathematics, we have no way of answering the question of theory choice: why we prefer some formal systems to others. Without any reference, formal mathematics can only be seen as arbitrary fiction. That would not only make the scientific applications of mathematics a miracle, but also render mathematics

as a human endeavour impossible. Extreme formalism is an impossible point of view once we look at mathematics in a bigger picture. In addition, I will argue that the milder forms of nominalism, such as geometric strategies and modal reconstructivist approaches, are not really nominalist at all in the strict sense we are concerned with in this work. In all of them mathematics includes a reference, however weak it may seem, and as such they are compatible with Tarskian truth. In the end of the chapter I will suggest a rough outline of ideas for a non-formalist philosophy of mathematics that does not run into the epistemological and ontological problems of Platonism. This is not meant to be a fully constructive argument, but rather a framework sketched in order to illustrate that the non-formalist options range much wider than the usual Platonism-nominalism dichotomy suggests.

In this work my argument is that mathematics without any reference for the formal sentences does not make sense, and the existence of reference – together with pre-formal thinking – allows us to use Tarskian truth in an unproblematic fashion. However, there exists an interesting argument to the opposite direction: that the analytic truth of arithmetic implies that there exists a reference for it, even a Platonist one. This *neo-Fregean* argument of Crispin Wright and Bob Hale is the subject of Chapter 7. Neo-Fregeanism has many technical problems which will be discussed here, but the most crucial weakness is the epistemological rationalism it requires. The analytic truth concerning natural numbers that Wright and Hale use is in fact, I claim, a *definition* of natural number. For one that does not accept epistemological rationalism, it is extremely problematic to infer objective existence from a *linguistic fact* that a definition essentially is. Instead of Platonism, we are in danger of succumbing into its exact opposite: extreme formalism where mathematics consists *only* of such linguistic facts. Hence the neo-Fregean strategy of truth-first, reference-second fails. Neo-Fregeanism is only successful when we can have antecedent justification that there exists an objective reference for natural numbers. Øyvind Linnebo has proposed a Fregean strategy where *semantic values* of numerals (names of natural numbers) work as this reference. This way, we can avoid both Platonism and

formalism. Augmented with a theory about the origins of mathematics, I propose this to be a promising route to take in the wider ontological and epistemological questions concerning the philosophy of mathematics.

Chapter 8 deals with various slightly tangential subjects not addressed in the first seven chapters. These include non-standard models, another semantical argument and other Gödelian arguments. In the very final chapter of this work the concept of substantiality (robustness) of truth is discussed. In it I have tried to make explicit an underlying argument of this work: we do not need to know the exact nature of mathematical truth in order to be able to talk about it. In fact, from this work one will not find comprehensive arguments for Platonism, empiricism, naturalism, structuralism or any other metaphysical and epistemological theories of mathematics. Yet the study on truth and proof here should not be on any weaker basis than in more complete philosophical pictures of mathematics. Aside from the substantiality of truth, that is the main thesis (sort of *meta*thesis) of this work: we can know there is a difference between truth and proof without knowing what truth exactly is. Simply put, if such a difference did not exist, mathematics as we know it could not be possible.

2. The Background

2.1 The Problem of terminology

As is the case in many areas of philosophy, there is worryingly little consensus on much of the central terminology in the philosophy of mathematics. Terms like realism and Platonism on one side of the spectrum, and formalism and nominalism on the other one are used interchangeably. There is a noticeable tendency for the nominalists to define their position negatively as "non-Platonism", and for the realists to define theirs as "non-formalism". This is usually augmented by targeting the extreme version of the opposing viewpoint. Perhaps some of this can be attributed to more general human frailties, but certainly much of it stems from the subject matter. While there is an almost universal agreement concerning the accepted methods of working mathematicians, the philosophers of mathematics have found remarkably little to agree on.[9] Hartry Field, for example, is a fictionalist. According to him mathematics is just a fiction; mathematical objects do not exist and mathematical sentences[10] do not have objective truth-values. Thus it seems that when we say that a mathematical sentence φ is true, we are only really saying: "from our accepted axioms, with our accepted rules of proof, we can derive φ". Nothing in this refers to anything objective, if we do not count the very weak sense in which mathematical conventions

[9] This also includes criticism on those accepted methods, as intuitionism and revisionist logics have shown us.

[10] Throughout this work, mathematical *propositions* will be talked of as *sentences*. There is some tradition of using the word "proposition" in philosophy and the word "sentence" in mathematics, the difference essentially being that a sentence s is always in some specific language L, while a proposition p is the *content* of s that can also be expressed in other languages. This way s is the sentence putting forth the proposition p in the language L. With the formal languages we have in mathematics the difference largely vanishes, since we are almost always working in the context of a specified language. In any case, it will not be of much importance in the problems considered in this work, and whenever sentences are essentially language-dependent, this will be acknowledged.

can be considered to be objective. This point of view is called extreme (strict) formalism. Granted, mathematics is a *useful* fiction, but a fiction nonetheless. Or so Field argues.[11] Roger Penrose, on the other hand, is a Platonist. He believes that there exists an independent world of abstract mathematical objects and relations between them, and by practising mathematics we are able to gain knowledge of this world.[12] When we say that a sentence φ is true, we are making a statement about the state of affairs in that world. If we are correct, the sentence φ corresponds to that state of affairs.

Such diverse viewpoints are of course nothing unusual in the history of philosophy. What makes the matter interesting is the fact that Field and Penrose are both contemporary and actively publishing philosophers.[13] With such extreme points of view still around, it is hardly surprising that most of the intermediate positions seem to be covered, as well. Therein lies the terminological trap: when a nominalist speaks about objectivism in mathematics, there is a good chance he is referring to the Penrose-type Platonist realism. Similarly, a realist is bound to argue against the Field-type extreme formalism. Most of the actual philosophical viewpoints, however, belong somewhere between the two extreme positions. This has the unfortunate consequence that more moderate points of view get less exposure, as most of the criticism is concentrated on the radical positions. At its worst this picking of easy targets has an effect on whole arguments, but at the very least it is a problem in understanding the terminology. What a nominalist means by "realism" will not always coincide with the realist's own definition, and there remains little hope for any debate to be fruitful.

To avoid problems of that kind, I want to be clear about the central assumptions and terminology in this work, although the

---

[11] See Field 1980 (first chapter) and 1998. While Field is clearly a fictionalist over mathematics, there is some controversy whether Field's philosophy can be considered to be extreme formalism. We will return to that in Chapter 3.1.

[12] See Penrose 1989, pp. 146-151.

[13] Penrose is of course primarily a physicist, and perhaps secondarily a mathematician, but his later work is largely philosophical.

arguments presented here are compatible with most philosophical accounts of mathematics. In fact, the only notable exception is extreme formalism. As long as we are ready to agree that there is *something* more to mathematics than the formal axiomatic systems – that is, something objective – the arguments here can be applied. It need not be much – we certainly do not need to assume a whole independent world of abstract ideas – but it will be necessary to be able to speak of reference to something outside the formal mathematical theories. Whatever formal mathematics is, it is a formalization *of* something, and not a completely independent arbitrary set of axioms and rules of proof. That is the only assumption needed for now. It will not be used as an axiom, though, and Chapter 6 of this work is about justifying this assumption – that is, arguing against the position of extreme formalism. Until then it is my purpose to show that the only philosophical viewpoint that contradicts with the arguments in this work will be extreme formalism. This will cause both explicit and implicit tendency to use seemingly realist-flavoured terminology. However, that by itself must not be confused with advocating Platonism, or even some milder form of realism. Nor should it be thought that the arguments here are somehow dependent on the used terminology. It is simply a terminological requisite for this work that we are able to speak of concepts like existence, objects and reference in a non-formalist way. In addition, it must be remembered that there is also a deep-entrenched custom of using realist terminology among practising mathematicians. That has been a factor in the choice of terminology here, as well.

2.2 Platonism

Although mathematics existed before ancient Greece, it is from Plato that we have got the first more or less systematic philosophical account of mathematical thinking. Remarkably, it is also quite a popular one today, especially among working mathematicians. But when a philosophical term lives for more than two millennia, it is bound to go through significant changes in meaning. That has also happened to Platonism: it is not uncommon

to call mathematical realism "Platonism" even though our mathematical and philosophical thinking has no resemblance to that of Plato's. That is why throughout this work Platonism simply means the position of Plato with regard to the philosophy of mathematics.

Mathematics in Plato's time was a direct continuation of the Pythagorean tradition, and his philosophy of mathematics is properly understood only in that context. For the Pythagoreans, mathematics was a curious mixture of science and religion. The now traditional paradigm for mathematical thinking, that is, the *a priori* pursuit of necessary universal truths, was already present in the Pythagorean philosophy.[14] In this sense, the status of mathematical thinking for the Pythagoreans – with the obvious differences in scope and formalism – was largely the same as it is for us. Their philosophy of mathematics with its mystical and religious character, however, is a whole other matter. Essentially, numbers were gods for the Pythagoreans, and as such something eternally beyond our physical world.

The influence of the Pythagorean tradition on Plato's thinking was overwhelming. When Plato updated the mysticism of Pythagoras to his own brand of rationalism, the nature and importance of mathematics changed little. In Plato's philosophy gods (in this respect) were changed into a world of *ideas*.[15] They were abstract, that is, completely non-spatial and non-temporal in their essence. To mention the most obvious example, geometry – the paradigm of mathematics at the time – was concerned with abstract objects, while physics was concerned with their spatial resemblances.[16] In this sense, any real knowledge was always of the mathematical type, and it was completely *a priori* in its nature. Hence, in this sense, Plato's philosophy had not drifted far from that of Pythagoras. Mathematical knowledge was the model that all real knowledge followed. While it did not concern gods any more, ontologically and epistemologically it was not far off.

---

[14] See Jones 1980, pp. 34-39 and Boyer 1985, pp. 65-66; pp. 115-127.

[15] Here Parmenides and his idea of eternal existence were a great influence on Plato, as is visible in the dialogue *Parmenides*.

[16] See *The Republic* 527a-b.

When we use the word "Platonism" in the philosophy of mathematics, this history should be remembered. Often Platonism is used as a synonym for mathematical *realism*, which I understand here as the wider philosophical position that mathematical concepts and truth refer to something objective – something outside the work of human mathematicians. But mathematical Platonism, properly understood, is the philosophical viewpoint that mathematical objects exist in *an ontologically independent world of abstract ideas*. Of course postulating a whole eternal world for the objects of mathematics is maximal in its ontological commitments, and thus highly problematic for most modern philosophers. For Plato, however, this was obviously no problem. In his philosophy, *all* knowledge – not only that of mathematical truths – concerned such a world. This is an important point to remember. In Plato's philosophy mathematical knowledge was the model for all real knowledge, and it did not differ in character from any other forms of real knowledge. Of course here a modern Platonist in mathematics thinks very differently. For him to think that mathematical knowledge concerns an ontologically independent world means that mathematics is totally different from most – perhaps all – other types of knowledge. As we will see, that is a highly problematic position, and not something that most realist philosophers of mathematics would be ready to agree with.

With the gradual death of Platonism in general ontology and epistemology, one would have expected to see its disappearance from the philosophy of mathematics, as well. Instead, mathematics has proved to be the last refuge for a Platonist. Not only are there still active Platonist philosophers of mathematics, but more importantly, Platonism has also remained the archetype for realist philosophy of mathematics – up to the point where the two terms are used interchangeably. This is an equivocation we must be able to banish from the philosophy of mathematics for good. To believe that sentences of mathematics have objective references should in no way imply that these references are objects or relations in a Platonist world of ideas. That is why throughout this work I will speak of Platonism only as it was characterized by Plato himself: the position that mathematical objects exist in an ontologically independent world of ideas, and the truth of mathematical

theorems depends only on the state of affairs in that world. While Platonism is obviously one form of realism, mathematical realism can mean a number of different positions, many of them much more ontologically economical. If we reject Platonism, we do not need to advocate the view of mathematics as a fiction consisting of (ultimately) arbitrary conventions.

2.3 Realism/objectivism

Perhaps the weakest form of mathematical realism is defined by W.V.O. Quine's (1966) famous *indispensability* argument. According to him, mathematics is an indispensable part of scientific theories, and mathematical objects exist in the same way as scientific objects do. Consequently, a mathematical theorem is true in the same way as a theorem of, say, physics is true. Quine's position is notoriously vague in its holism, but it seems sensible to call Quine's indispensability argument realism when it comes to mathematical objects, and the truth of mathematical theorems. Of course his brand of realism is quite different from Platonism, and many mathematicians would be quick to dismiss Quine's understanding of the nature of mathematics. Perhaps most importantly, in Quine's account the connection between mathematics and empirical sciences could be conceived as a two-way street: new empirical findings in science could change the truth-values of mathematical statements. Certainly most mathematicians would not be ready to agree with this. Nevertheless, in Quine's theory mathematical objects exist as independently of human conventions as anything does.[17] In other words, in the Quinean interpretation, mathematical realism is just a branch of scientific realism in general. As I see it, when we brand a philosophy of mathematics "realism", we must include both the Platonist and the Quinean variation. In fact, we must include all

[17] Obviously one could claim that *nothing* exists independently of our thinking in Quine's philosophy. This criticism is not wholly unjustified, but Quine's account still seems to be best understood as a form of realism.

philosophical accounts that contain *any* reference to something outside our mathematical thinking.[18]

This makes realism a vast and varied field, but there does not seem to be any other way of doing the taxonomy. The only possible definition seems to be the weakest one: realism is the philosophical viewpoint according to which at least some mathematical objects exist independently of our thinking. That is called *realism in ontology* in the literature. Since the topic of this work is mathematical truth, we will be more concerned with *realism in truth-value*, that is, the position that the truth of mathematical statements does not depend only on our thinking.[19] But first we must try to clarify what we mean by the vague concepts "mathematical objects" and "mathematical sentences". This is no small matter: in fact, a definition would be too much to ask. Mathematics has expanded vastly in the last centuries and what we consider objects of mathematics are different from, say, what Newton did. Yet the nature of mathematical objects need not be any different. The best solution here is picking one area of mathematics as the paradigm case for us to study. The most popular choices for such "first mathematics" in the literature are logic, set theory[20] and arithmetic (number theory). For reasons that will be evident later on, I choose the last one. From now on, unless otherwise mentioned, mathematical objects will mean the objects

[18] The term "objectivism" could be better in this respect. In this work realism and objectivism are used synonymously.

[19] Whatever difference there may be between realism in ontology and truth-value, they should not matter in this work. See Shapiro 1997, pp. 36-38 for clarification over the two types of realism. Although realism in ontology may seem like a much stronger position, it should be remembered that since mathematical objects need not be of the Platonist type, the difference is not necessarily significant.

[20] Unless otherwise mentioned, by set theory throughout this work I refer to the usual Zermelo-Fraenkel (ZF) set theory expanded with the axiom of choice (ZFC). It must be noted that we can define arithmetic in ZFC, so at least set theory as a choice of basic mathematical theory could in no way change the arguments here.

of arithmetic, first-order Peano arithmetic (PA) to be exact.[21] These objects are the natural numbers (0, 1, 2, 3, …). Similarly, mathematical statements will mean the sentences of arithmetic. With this, and a final touch of replacing the vague term "our thinking" with something more concrete, we should be ready to give the definition for mathematical realism.

> ***Realism (objectivism)*:** there are mathematical sentences that are true if and only if they refer accurately to some entities or relations independent of the work of human mathematicians.

What this definition is saying is that according to realism, some mathematical statements have objective truth-values. Of course this does not need to concern all mathematical sentences. Complex numbers, for example, are often not considered to exist, even in realistically inclined circles. Here our choice of arithmetic as the

---

[21] Peano arithmetic was constructed by Giuseppe Peano in 1889. It is the theory of natural numbers defined by the following five axioms:

(1) Zero is a number.
(2) If n is a number, the successor of n is a number.
(3) Zero is not the successor of any number.
(4) If two numbers have equal successors, they are themselves equal.
(5) If a set S of numbers contains zero, and for every number in S its successor is also in S, then every number is in S. (This is called the induction principle.)

The Peano axioms are often also called *Dedekind-Peano* axioms due to Richard Dedekind's earlier similar work in the axiomatization of arithmetic. The most famous other axiomatization of arithmetic is the one presented by R.M Robinson in 1950, and denoted **Q** in literature. The main difference is Robinson's exclusion of the induction principle as an axiom. Although **Q** is a weaker theory than Peano axioms, the results considered in this work follow from both axiomatizations. Instead of the second-order axiom (5), we can also present the induction principle as a first-order induction schema, thus getting a first-order axiomatization of arithmetic, called PA in literature. From now on, unless otherwise mentioned, we will be concerned with first-order Peano arithmetic.

paradigm case seems helpful: if there are mathematical sentences that have objective truth-values, statements concerning natural numbers seem like a safe set to start from.

Realists believe that there is *something* outside the work of mathematicians that makes mathematical sentences true or false; that it is not a mere convention. It is important here to use the weaker concept "entity or relation independent of the work of human mathematicians", instead of a full-fledged Platonist ontology. Aside from a Platonic idea, this could mean, among other things, similarity in the physical structure in the brains of human beings (a kind of naturalism), or the Quinean view that mathematical facts are just one class of scientific facts in general. What is meant here by "the work of human mathematicians" is basically the end product of mathematical thinking: what mathematicians write on paper. This does not include the thought process and creativity behind that end product. Simply put, a textbook of algebra is a work of human mathematicians, while all the thought process behind that writing is not. This is a necessary clarification to make. Mathematical objects *could* be dependent on our thinking, but still be something more than conventions. If, for example, human beings turned out to have a common physical disposition toward thinking in certain mathematical ways – like most of us have a physical disposition toward seeing colours – this must be considered to be something outside the work of human mathematicians. Mathematical sentences would have objective truth-values, and we would have to be realists over mathematical truth.

2.4 Formalism/nominalism

As the complement of the above definition of realism, formalism means anti-realism in the philosophy of mathematics.[22] According to formalism, mathematics is only a human creation that does not refer to anything objective, and the paragon of human creation in mathematics is the *formal system*. The concept of formal system is most often associated with David Hilbert. In the intuitionism battle of the early 20th century, Hilbert proposed reducing the (for the intuitionists) problematic methodology of abstract inferences and ideal statements in mathematics to the universally accepted finitistic proof methods and real statements. Real statements mean finitistically meaningful statements of the form $\forall x(f(x) = g(x))$, where f and g are some simple functions, usually interpreted to mean primitive recursive.[23] A mathematical system (that is, a set of mathematical sentences, or axioms and rules of proof) consisting only of real statements will be called a formal system.[24]

Hilbert wanted to show that just like complex numbers were conservative over real numbers, abstract inferences were conservative over finitistic proof methods. The way that the use of

---

[22] It could be (quite justifiably) argued that the array of philosophical theories is much wider than the realism/formalism dichotomy I endorse in this work. The approach here does not mention such philosophical theories as constructivism, empiricism, intuitionism and fictionalism, which are usually considered to be alternatives to formalism and realism. However, the classification here serves a purpose. The subject of this work is mathematical truth, and the main question concerning that will be whether truth refers to something objective or not – that is, whether we consider mathematics to be realist or formalist under the definitions here.

[23] Primitive recursive functions are the smallest set of functions that include the zero-function $f(x) = 0$, the successor function $f(x) = x + 1$ and the projection function $f_i^n(x_1,...,x_n) = x_i$, and that is closed under composition and recursion. A set of functions S is closed under composition if every composition of the functions in S is also included in S. Similarly, a set is closed under recursion if every function formed by recursion from the functions in S is included in S.

[24] The terms "theory" and "formal theory" are often used for the same concept.

complex numbers did not lead to any new algebraic identities concerning *real* numbers, Hilbert wanted to show that the abstract proof methods and ideal statements could not be used to derive any real statements that could not be proved with finitistic proof methods. This way, for all the abstract methods used by mathematicians, there would always be a completely formal, finitistic equivalent. In essence, were this *Conservation program* of Hilbert to be successful, mathematics could be completely reduced to formalism.[25]

It might seem obvious that concerning the foundations of mathematics, Hilbert was a formalist. But this is not so simple. It must be remembered that Hilbert was trying to *defend* all the non-formalist methods that working mathematicians use. His quest on this occasion was not creating a new foundation for mathematics, but rather defending the classical mathematics against the intuitionist "putsch". In this sense, Hilbert's program does not imply any ontological theory about the existence of mathematical objects. However, even though he as a prolific working mathematician did not reject the use of ideal statements and abstract proof methods, he thought that mathematics *could* have a completely formal foundation. Whatever means mathematicians have of proving theorems, Hilbert believed that there is always a completely formal, finitistic, equivalent. Simply put, all the contemporary mathematics could be reduced to mere deriving of strings of symbols from other strings of symbols according to specified rules of syntax – for that is what a formal system is essentially about.[26] This should also give us a clear conception of the central idea of formalism: ultimately, mathematics is always equivalent to rules of symbol manipulation. Any trust in Occam's razor would seem to imply that this is indeed *all* that mathematics really is, even if that is something Hilbert himself would not have agreed on.

---

[25] See Smorynski 1977, p. 822 for the mathematical details, and Detlefsen 1986 for a philosophical overview. The original work can be found in Hilbert 1970.

[26] This includes the obvious feature of formal systems that they can always be presented in completely formal languages.

That last viewpoint is the extreme position of formalism, but some weaker interpretations of Hilbert's program are at least as valid. Formalism concerning mathematics, it can be argued, consists of at least three different points of view. First, we have the stance that mathematics should be presented, as far as possible, in formal axiomatic systems for maximal clarity and deductive power. All mathematicians and philosophers of mathematics are likely to accept this version of formalism. Second, there is the position that formal systems are the objective of mathematics. Theories are not truly mathematical until they can be axiomatized. Finding out the best axiomatizations is of course one of the main concerns of mathematics, but unlike in the first type of formalism, in this second type *only* formal systems are considered to be acceptable presentations of mathematical theories. This is more or less the Hilbertian (1970) idea of formalism and, although more limiting, still not a difficult one to accept. Indeed, this is the goal most mathematical theories strive for. The third, and by far the most controversial, type of formalism is the philosophical doctrine that mathematics as a subject only concerns formal systems, without any reference to anything outside them. While the milder types of formalism are ways of presenting mathematical methodology, this third type is highly committing philosophically. For Hilbert formal systems were the ultimate tool, perhaps even the essence of mathematics. But for him it was not *all* there was to mathematics. [27]

---

[27] Although quotes like the following certainly suggest that Hilbert's formalism was of a considerably strong flavour:

> If the arbitrarily given axioms do not contradict each other through their consequences, then they are true, then the objects defined through the axioms exist. That, for me, is the criterion of truth and existence. (In Meschkowski 1973, p. 56, quoted in Smorynski 1977, p. 825)

While this reserves a crucial role for formal systems, Hilbert all along speaks of existence and truth of mathematical objects. I believe this quote expresses Hilbert's optimism in his Conservation program, and the role he hoped formal systems would play. However, he did this in order to defend mathematics, not to introduce a limiting system of strict formalism (see Reid 1970, pp. 155-157).

This third type, what I call *extreme*, or strict, formalism, is the type we are concerned with in this work. From now on, unless otherwise stated, formalism in the philosophy of mathematics will mean this extreme type, for which we are ready to give a definition:

> ***(Extreme) Formalism:*** to say that a mathematical sentence is true involves no reference to any entity outside formal systems. Hence, a mathematical sentence is true in a formal system **S** if and only if it is provable in **S**, and mathematical truth cannot be discussed in any other context.

When we remember that formal systems are the ultimate work of human mathematicians, we clearly see that this definition of formalism is the definition of anti-realism as presented in the previous chapter. Rules of proof are the way we reach theorems from axioms, and for the extreme formalist this is the only way of acquiring, and recognizing, true sentences. Keeping in the wider tradition of realism and anti-realism over concepts, formalism is often called *nominalism* in the literature. The possible difference between the two terms is not likely to be important here, and the two terms will be used synonymously. Some nominalists, such as Charles Chihara (2005), are ready to extend mathematics beyond formal systems, which would imply that nominalism and formalism are two different positions. In addition, nominalism over physical concepts is a different matter from nominalism over mathematical ones. For these reasons formalism seems like the preferable term, although it carries the potential confusion with Hilbert-type moderate formalism. This will be a subject later on in this book (Chapter 6.3), but for now it suffices to say that I do not think that these differences are important, and there should not be any problems using the terms nominalism and formalism interchangeably. Basically, either mathematics refers to something outside the work of human mathematicians (ultimately something outside formal systems) or it does not. It is hard to see any neutral

ground, and I do not see how a nominalist could accept such a reference and still remain a nominalist.[28]

2.5 Soundness and completeness

Realism in truth-value and formalism are both answers to the question: what does it mean to say that a mathematical statement is true? The obvious follow-up question is how we can find out *which* mathematical statements are true? Of course the whole point of mathematics is to use rigorous rules of proof to obtain new theorems, and – whatever we mean by truth – the trust in these rules of proof is founded on them *preserving truth*. This is the standard way of describing our logic, and hence, our mathematics. Although the modes and methods of mathematical practice are largely universal, there still does not exist a full consensus on the accepted rules of proof (that is, the accepted inferences in logic). Intuitionism in the late 19th and early 20th centuries was based on exactly such a conviction: Henri Poincaré and later L.E.J. Brouwer argued that the rules of proof used in mathematics were in fact flawed.[29] More recently, similar arguments have been made based on many-valued logics. Nevertheless, the motivation behind all these logics is that they are supposed to preserve truth. Whatever rules of proof we accept, we believe that with them we will end up with true theorems from true axioms.

What does that mean in the framework of realism and formalism? Obviously rules of proof belong to the domain of formal systems. Indeed, formal systems consist only of axioms and rules of proof. Hence, as far as truth and proof are considered, the formalist position is quite simple: rules of proof and axioms are all there is to mathematics. Any talk of proof preserving truth is

---

[28] Here I take the direct approach to understanding nominalism: as the word itself suggests, instead of mathematical objects only their *names* exist. This position is also called fictionalism in literature. Nominalism in mathematics is sometimes also understood specifically as the position that *sets* do not exist.

[29] See Dummett 1977 for a good introduction on intuitionism.

redundant, for truth is *equivalent* to proof. Whatever our rules of proof are, according to the formalist position, they must *always* preserve truth. As such, truth would be an empty, *deflationary*, property – in extreme formalism truth "deflates" completely into proof.

The realist position is a more difficult one, and philosophically much more interesting. If we believe that there exists an outer reference for the statements of mathematics, we must clarify what the relation between this reference and our formal systems is. The question can be divided into two parts. First of all, we should want our rules of proof to preserve truth, that is, to be *valid*.[30] In addition to the rules of proof being valid, we want the axioms of the formal system to be true. These two conditions put together are called the criterion of *soundness*. A formal system is sound if and only if *only true sentences (theorems) can be derived in it*. Secondly, we should want our rules of proof to be such that *all* such theorems can be derived with them. This is the criterion of *completeness*, and it is the converse of soundness. A formal system is complete if and only if *all true sentences (theorems) can be derived in it*, that is, for every sentence φ in the language of the formal system, either φ or $\neg\varphi$ is a theorem.[31] At this point we should not worry ourselves about the nature of truth, or the epistemic status of axioms. Whatever truth may be, soundness and completeness are the two features that the realist would hope from our formal systems. One direct consequence would be that such formal systems would also be *consistent*: for no theorem φ could its negation $\neg\varphi$ also be provable.

What about the formalist? The mathematical essence of Hilbert's program was to show that formal systems could be

---

[30] To be exact, this means that by the rules of proof it is impossible for the axioms to be true and a theorem to be false.

[31] In logic and mathematics there are many different forms of completeness. The one given here is used for reasons that become apparent in the next chapter. The main idea is that every well-formulated sentence we can construct in a language must be a theorem, or else its negation must be a theorem. This is a very natural criterion to have, especially when we remember that we are dealing with sentences of arithmetic, that is, statements concerning natural numbers.

shown to be complete and, equivalently, consistent. Had it succeeded, it would have presented a strong argument for formalism: once we agree on the axioms and rules of proof, we could (in principle) have automata filling in the formal systems with theorems, and do the job completely. For the formalist soundness comes automatically, and with completeness formal systems would be just the kind of perfect tool Hilbert wanted them to be. It would not have solved the problem of truth and proof for good – after all, we would still need to agree on the axioms and rules of proof – but at least it would have eliminated the possible occurrence of true but unprovable sentences. Whatever the set of all true sentences is considered to be, it would be equivalent to the set of all provable sentences once we find the proper formal system. Metamathematical and philosophical questions aside, this would make it very hard to distinguish between truth and proof as concepts. The question of realism and formalism would still exist in the question of theory choice (that is, the choice of axioms and rules of proof), but the subject matter of formal systems would be sound and complete. For the formalist of Hilbert's (the second) type this would have been the ultimate success. The extreme formalist would still have questions to answer, but her case would definitely have had extra strength, as well.

2.6 Gödel's incompleteness theorems

Of course Hilbert's dream was not to be realized. His optimism for finding a consistent, complete and totally formal basis for mathematics was dealt a devastating blow in 1930 when Kurt Gödel presented two theorems that proved Hilbert's program to be impossible. Gödel showed that all consistent formal systems of arithmetic are in fact *incomplete*. Gödel's two incompleteness theorems are as follows[32]:

[32] For the proofs of the theorems, see Gödel 1931 or Smorynski 1977, pp. 826-828.

First Incompleteness Theorem: Let **T** be a formal system that contains[33] arithmetic. Following the so-called fixed-point theorem, in the language of **T** there can be constructed a sentence φ : $\varphi \leftrightarrow \neg \mathrm{Pr}_T(\overline{\varphi})$, where Pr_T is the provability predicate of **T,** that has the following traits[34]:

(1) If **T** is consistent, then $\mathbf{T} \nvdash \varphi$.

(2) If **T** is consistent[35], then $\mathbf{T} \nvdash \neg\varphi$.

---

[33] This means that there is a known embedding from **PA** to **T**. An embedding from some system **S** to **T** is a function that preserves the arithmetic characteristics (such as addition, multiplication and ordering) of the objects of **S** in **T**. For example, if $f : \mathbf{S} \rightarrow \mathbf{T}$ is an embedding from **S** to **T** and $a + b = c$ in **S**, then $f(a) + f(b) = f(c)$ in **T**. This embedding property is needed for the logical operations, among other things.

[34] The notation $\overline{\varphi}$ means the natural number that is the *code* of the sentence φ. For his proof, Gödel needed a way to encode sentences of formal systems into natural numbers. This is also the reason for the condition of containing arithmetic. Such an encoding is called *Gödel-numbering* in literature, and there are various ways of doing it. For details, see Gödel 1932, p. 157.

[35] To be more specific, this includes an additional condition that **T** should be *ω-consistent* (see Gödel 1932, p. 236 and Smorynski 1977, pp. 851-852). This concept of Gödel means, in the case of formalized arithmetic, that the following two conditions are never simultaneously satisfied for any φ:

(1) $\mathbf{T} \vdash \exists x \varphi(x)$

(2) $\mathbf{T} \vdash \neg\varphi(\overline{0}), \neg\varphi(\overline{1}), \neg\varphi(\overline{2}), \ldots$

In other words, this means that a sentence cannot be true for some x at the same time as it is false for every x. Obviously, this is what we should expect from any mathematical system, so this additional condition should not worry us. Indeed, J. Barkley Rosser proved in 1936 that the requirement of ω-consistency can be dropped from Gödel's proof. For Rosser's Theorem, see Rosser 1936 or Smorynski 1977, pp. 840-841.

> *Second Incompleteness Theorem*: Let **T** be a consistent formal theory that contains arithmetic. Then $\mathbf{T} \nvdash Con_T$, where Con_T is the sentence asserting the consistency of **T**.

In other words, all consistent formal mathematical systems containing arithmetic contain sentences that can neither be proved nor disproved within that system. In addition, such a system cannot prove its own consistency. We remember that Hilbert's Conservation program was based on the belief that the abstract proof methods and ideal statements that mathematicians use are conservative over the finitistic proof methods and real statements of formal systems. This Conservation program is equivalent with his *Consistency* program of showing that such a formal system containing the abstract proof methods and ideal statements is consistent.[36] Gödel showed this latter program to be impossible: if such a formal system **T** is consistent, we cannot prove this consistency within **T**. At the same time he demolished the Conservation program. In short, Gödel proved that the formalist program could never be *complete*. This is the first important philosophical conclusion of his incompleteness theorems. The second one is that such a troubling Gödel sentence φ, while unprovable, is nevertheless *true* under very reasonable truth-theoretic assumptions.

The second conclusion will be the subject of the next chapter. Let us now look at the direct implications of the first one. It is clear that the second of Gödel's theorems demolishes Hilbert's program. We know unassailably that all consistent formal systems containing arithmetic are incomplete, and in addition unable to prove their own consistency. Because of this Hilbert could not succeed in justifying the abstract proof methods and ideal statements. But it also seems to have the graver consequence of making our very rules of proof doubtful. After all, completeness and consistency are two of the main qualities we would hope to include in our concept of proof. In extreme formalism, mathematics consists only of axioms and rules of proof. Thus the implication of Gödel's theorem to formalism seems to be that,

---

[36] See Smorynski 1977, p. 824.

taken as one consistent all-including formal system, *mathematics* is incomplete. In addition, if such a system were indeed consistent, we cannot possibly know this.

This alone is a remarkable result. The subsequent discussion on Gödel's theorems and formalism has been very active and varied, up to the point where the direct consequence of incompleteness is not often fully acknowledged. But there seem to be no two ways about it: if extreme formalism is correct, and all our mathematical knowledge can be presented as a single formal system, Gödel's incompleteness theorems have the inescapable consequence that mathematics is either inconsistent or incomplete. When switching from proof and disproof to the concepts of truth and falsity, this obviously means that under the formalist interpretation there are mathematical sentences that are neither true nor false, which goes against the law of excluded middle. For a realist, it is of course only the proof method (and the axiomatization) that is incomplete. For a formalist, it is the whole of mathematics.[37] At least tentatively, the realist conclusion seems much easier to accept. But we will get to the bottom of this soon, and in Chapter 4.2 these considerations will be given a detailed treatment.

2.7 Is the Gödel sentence true?

Now we know that the sentence φ is undecidable in the formal system **T** – but there is more to this. In Gödel's proof the unprovable Gödel sentence is formulated to be self-referential with the usual diagonal[38] procedure as $\varphi \leftrightarrow \neg \mathrm{Pr}_T(\overline{\varphi})$, where Pr_T is the provability predicate of **T**. In words, this means that φ is true if and only if it cannot be proved in **T**. But that φ cannot be proved in **T** is exactly what the part (i) of the first incompleteness theorem tells us. So looking at the meaning of φ, it must be *true*. This already

[37] If we give up the requirement of single formal system, we still end up with the very strong consequence that whole areas of mathematics are either inconsistent or incomplete.

[38] The method of self-reference also used in Liar's paradox, Russell's paradox and Cantor's diagonal slash.

makes the Conservation program impossible. The sentence φ is a real statement (this can be seen from the way it is constructed, plus from the fact that it is a sentence of a formal system), and it is true. Yet Gödel showed that we could not prove its truth with formal proof methods, that is, finitistic derivations. This is a direct counter-example to the Conservation program: our abstract proof methods and ideal statements give us the opportunity to see at least one additional real statement to be true, which was exactly what Hilbert wanted to show to be impossible.

However, so far we have not specified what it means to say that φ is true. Clearly we have jumped ahead in the matter, and assumed that we have defined what we mean by truth. As will be seen, this is not such a simple question. Our pre-philosophical intuitions immediately suggest the truth of φ, but truth without definition is an empty concept here. In fact, the definition for truth, and the need for the concept of truth, turns out to be one of the more important questions in the philosophy of mathematics, and most of this work concerns that problem. Meanwhile, however, it certainly looks like φ (usually called the Gödel sentence[39] G of **T**, or G(**T**), in the literature) is true – because φ *says so*. If it is indeed true, but not provable – and all this under reasonable assumptions – that would already be enough to show that truth and proof are different concepts. It would also imply that extreme formalism could not suffice; that there are sentences of mathematics which are true, but the truth of which escape the strict formalist interpretation. At the very least, Gödel sentences give us an explicit case to study – if there is a true unprovable sentence, G seems like a worthy candidate.

[39] We talk here about "the Gödel sentence", but actually every such formal system has infinitely many unprovable sentences like the one given here.

2.8 Gödel sentences and Tarski

Now there are two questions we have to ask about the truth of $G(\mathbf{T})$. First of all, since not by proof in **T**, how do we establish that the Gödel sentence is true? For now, we should be confident of having a good idea what proof is. What do we know about *truth*? The second question concerning the truth of $G(\mathbf{T})$ follows directly from the first one. Once we clarify how we see its truth, we have to find out *where* we see its truth, that is, in what kind of mathematical (or other) system do we establish that the sentence $G(\mathbf{T})$ is true? Obviously it is not in the formal system **T** itself, when we follow the formalist conception that provability equals truth in formal systems. But **T** is the only system we have examined in proving Gödel's incompleteness theorems. Are we changing the game by talking about other systems, and if so, what consequences does this change have?

Both of these questions are central to the philosophy of mathematics, and thanks to Gödel's work, we now have an explicit sentence through which to look at the matter. It is obvious that we do not see the truth of $G(\mathbf{T})$ as proof of G in **T**; this is the very thing that Gödel proved. The way $G(\mathbf{T})$ was constructed, it seems just as obvious that we establish its truth through the *meaning* of $G(\mathbf{T})$, that is, semantically. There seems to be very little doubt over this.[40] We said that "$G(\mathbf{T})$ states that it cannot be proved in **T**", and since it indeed cannot be proved in **T**, it must be true. This is easy to understand because we can change positions from the formal sentence $G(\mathbf{T})$ into its semantic content. In the process, however, we also quite clearly switch from *truth as proof* (that is, syntactical truth) to *semantical truth*.

How does this fit into our previous domain of formal systems? One solution could be denying us the right to talk about mathematical truth beyond that of proof in formal systems. In the traditional formalist view, truth in formal systems is defined as provability; truth and proof are the same concept. The ones who maintain this position can claim that Gödel only showed that

[40] See Gödel 1958, p. 241 for his own argument to this direction.

formal *systems* are incomplete, not that the formal concept of *proof* is incomplete. The basis for this line of thought is the (correct) observation that we do not establish the truth of $G(\mathbf{T})$ in **T**, but in another, broader, system. Because in extreme formalism formal systems are all there is to mathematics, there is no way we can see the truth of $G(\mathbf{T})$ in a broader system, as well as no way of seeing the truth of "$G(\mathbf{T})$ in **T**". For a proponent of such strict formalism, we simply have no justification of speaking about truth outside the context of formal proof. Even though formal systems are always incomplete, we have no *other* way of establishing truth-values of sentences, and hence the formal proof method can in fact be complete, although only in this reduced sense. This way, all true sentences can still be considered provable; neither $G(\mathbf{T})$ nor $\neg G(\mathbf{T})$ is provable, *ipso facto* neither of them is true.

However, there are three big problems with this line of thinking. The first thing that strikes the eye is the apparent emptiness of the argument. If proof defines truth completely, of course all true sentences are provable, and vice versa. Yet if this turned out *not* to be the case, we could never find it out with the strict formalist way of thinking. Without any outer reference, *all* formal systems are complete in the reduced sense. No matter what our formal systems are, they exhaustively prove all the true sentences. But surely this is too drastic a result for most formalists. Not all formal systems can be equally good. Outer reference cannot be ruled out this easily, and we cannot dismiss the possibility of true but unprovable sentences.

Secondly, the conclusion that neither $G(\mathbf{T})$ nor $\neg G(\mathbf{T})$ is true goes against the most basic premise of two-valued logic. Gödel's proof is established in two-valued first-order logic, and this reduced sense of completeness would conflict with the choice of logic it was based on. This way, in order to include a new concept of completeness, we would need to go deep into revising our notion of logic. Such strategies can have potential – they will be the subject of Chapter 5 in this work – but they go beyond Hilbert's conception of formalism in mathematics.

Thirdly, following Gödel's proof, it seems that we immediately *do* establish the truth of $G(\mathbf{T})$. We have not specified what we mean

by truth, let alone have a theory for it, but it seems instantly obvious that if there is a consistent formal system **T** containing arithmetic, the sentence *G*(**T**) is true in some system "**T** + theory of truth". Rather than deny this, we should concentrate on *explaining* it, even if the truth of Gödel sentences turned out to be an illusion. Since we clearly use the semantic content of *G*(**T**) in the process, we should start unravelling these problems from the semantic notion of truth introduced by Alfred Tarski (1944).

Tarski's theory of truth[41] was based on his "Convention T" being materially adequate, that is, giving all the true sentences of an object language. This means T-instances of the form:

"*P*" is true if and only if *p*.

Here we have to distinguish between the two languages that "*P*" and *p* are in. "*P*" is a sentence in the object language, and *p* tells in our metalanguage the proposition expressed by "*P*". The classic example is:

"Snow is white" is true if and only if *snow is white*.

Here *snow is white* is in the metalanguage (in this case English) and "Snow is white" can be replaced by any sentence expressing the same proposition. This depends, of course, on the object language.

---

[41] There are many ways of addressing Tarski's scheme in literature. Some call it a definition of truth, while others call it more cautiously an "adequacy condition". The latter line of thinking is that Tarski's scheme does not define truth, but rather gives us a form that all cases of true sentences take. As far as mathematical truth is considered, I think we should be safe in using both terms. If a scheme is adequate for enlisting all the true (as well as false) sentences, it should be more than satisfying as an account of mathematical truth. In a mathematical sense we could call it a definition, although an *implicit* one. Truth would be defined by all the instances of true sentences. What the concept "truth" ultimately means is a metaphysical question and potentially a very complicated one. This will be discussed later on.

For example, in the object language of Italian, the T-scheme takes the form:

"La neve è bianca" is true if and only if *snow is white*.

Now we can use the T-scheme to our Gödel sentence *G*(**T**), remembering that the informal semantic content of *G*(**T**) was "*G* cannot be proved in **T**":

G(**T**) is true if and only if *G cannot be proved in **T***.

This seems to illuminate our "seeing" the truth of *G*. We can now take notice of some of the upcoming problems. It is indeed the case that *G cannot be proved in **T***, but it must be noted that this is a sentence of our metalanguage, English. It is not in **T** anymore, and one might argue that we are changing the game by changing the language. So while it seems that *G* is indeed true, it is not true in **T**. However, if we do not equate truth with proof like the formalist does, will it make sense to speak of "truth in **T**" in the first place? It seems that we only have "proof in **T**", while truth is always in "**T** expanded with a theory of truth"? The fundamental question here seems to be whether we are prepared to expand formal systems with a theory of truth in the first place. If we are, we seem to be able to establish a true sentence that is not provable. If not, we must equate truth with proof, and neglect the apparent semantical truth of Gödel sentences. But that means dispensing with truth altogether, and dispensing with semantical thinking in mathematics, as well as facing all the problems we noted above. In the next chapter we will see explicitly how formal systems can be expanded with Tarskian truth in order to establish the truth of Gödel sentences, and all the questions here will get an unambiguous treatment.

Before that, we must consider the possibility of defining truth within **T**, in which case we would have no need to expand the formal system. One of the most important results concerning truth and logic is Tarski's (1936) *undefinability* theorem of truth. Tarski

showed that interpreted formal languages[42] (in the realm of classical first-order logic[43], and having enough expressive power to satisfy the fixed-point theorem of self-reference) could not include their own truth predicates. To be exact, they cannot be extended with a truth predicate without the occurrence of liar's paradox.[44] This is an important result, because it shows us that the Tarskian theory of truth *must* be formulated in the metasystem (and the metalanguage). Moreover, Tarski showed with this undefinability result that a truth predicate in classical formal languages commits us to an infinite hierarchy of languages. The truth predicate for an object-language L_1 is given in a metalanguage L_2, and the truth predicate for L_2 in another language L_3. Crucially, this hierarchy never collapses: at no level will a language L_m provide a truth predicate for a language L_n, where $n \geq m$.

Without this result there would have been the possibility that Tarski's T-scheme is equivalent to some predicate of the object system, which would conflict with the undefinability result, and hence free us of the need for metalanguages. In that case, the T-scheme would have been obsolete in formal languages. But it is not, and we see that in order to speak of the truth of sentences like $G(\mathbf{T})$, we have to step outside the system they were formulated in. This makes expanding the formal system necessary if we want to include *any* account of truth[45] – that way it should also help us a great deal in justifying the use of a Tarskian theory of truth.

The only alternative is dispensing with truth altogether, and that approach has many difficult problems. For one, it can be

---

[42] Unless otherwise mentioned, throughout this work we will be concerned with fully interpreted languages, that is, languages in which all sentences are either true or false through their meanings. Since the focus here is on arithmetic, this means that we are discussing standard models of PA. In Chapter 8.1 I will consider the significance of non-standard models in this context.

[43] In Chapter 5 I study the possibility of defining a truth predicate within the object language in other logics.

[44] Informally, "This sentence is false".

[45] Aside from the one where truth is categorically defined as a translation of proof – but this is also something we can only do outside the formal system, or otherwise Tarski's undefinability result is contradicted.

considered that such a truth predicate $Tr(\overline{x}) \leftrightarrow x$ *exists*, even in classical formal languages where it causes a paradox. We can define it completely formally and reach Tarski's undefinability result. In this way, it is not the introduction of truth predicate that gives birth to a hierarchy of languages that does not collapse. Instead, the truth predicate is a way to *show* that there exists such a hierarchy of languages that does not collapse. To avoid this, the language in question would have to have a specific ban of formulating a truth predicate, which is something not expressible formally within the object language.

We see that it is very difficult for the formalist to outright deny us the use of Tarskian truth. Even more importantly, in this work I will argue that Tarskian truth is in fact philosophically the *proper* theory of truth for mathematics. We will see that when we consider mathematics in a wider picture than just the formal systems, the formalist conception of truth as proof – and nothing else – is already highly insufficient, Gödel or no Gödel. To make sense of the reference between formal and what I will call *pre-formal* mathematics, Tarskian semantical truth is the perfect vehicle. Indeed, it will be seen that without a semantic conception of truth and its appeal to reference we would have no way of escaping arbitrariness in mathematics. These matters will be discussed in detail later on in this work. First we must see what kind of damage Gödel sentences and Tarskian truth do to the extreme formalist case.

3. The Semantical argument

3.1 Field's nominalism

The current discussion on what Neil Tennant (2002) aptly calls "Deflationism and The Gödel Phenomenon" consists mainly of two debates. In the first one, Hartry Field (1999) argues that mathematical truth is a deflationary, not substantial, property. Applied to formal systems, truth means proof, and only proof, and as such the property "truth" is empty. However, Stewart Shapiro (1998) uses Gödel's incompleteness theorem to argue that this is not the case: with a semantical notion of truth we can see that there are true sentences that are not provable – the Gödel sentences – and hence the two concepts are not one. This is called the *semantical argument* against deflationism in mathematics.

The second debate started with Jeffrey Ketland's (1999) answer to Field, and was followed by Tennant's (2002) reply to him (and Shapiro). The two discussions are similar and largely overlapping, but it is important to examine both. My approach here is to take a critical look at these arguments as what they are: parts of debates. This should give us a comprehensive and detailed overview of the competing viewpoints. The part considered in this chapter concerns only ordinary first-order two-valued logic. Arguments based on second-order logic and many-valued logics will be examined in Chapter 5.

As we know from Tarski's (1936) work, when we limit ourselves to classical first-order logic, formal mathematical theories cannot include their own truth predicates. This was an extra incentive for Tarski to pursue his semantic theory of truth. In fact, Tarski (1969, pp. 418-423) already thought that Gödel's incompleteness theorems, together with his own semantical conception of truth, showed that truth and formal proof are different concepts. He believed that a meta-theory (in a metalanguage of the formal system) could be constructed so that we can explicitly say that $G(\mathbf{T})$, a sentence of the object-system **T**, is true. This meta-theory would of course have to include the object-theory (the formal system **T**), but also Tarski's truth definition.

This goes well with everything that was said in the last chapter of the previous chapter, when we remember that **T** must be consistent. Against the deflationist thesis, for consistent formal systems containing arithmetic, Tarskian truth seems to give us the ability to establish a new true sentence.

One of the main opponents of this position has been Field. His most famous work is the book *Science Without Numbers* (1980), targeted against the contention of Hilary Putnam (1971) and Quine (1966) that mathematics is indispensable for scientific theories. In Quine's famous indispensability argument mathematics is considered to be an integral part of science, something without which science as we know it would be impossible. Quine held this to be a convincing argument against mathematical nominalism. Field's argument against indispensability was that mathematics is in fact *conservative* over nominalist scientific theories, and as such Quine's argument fails. His example was showing that Newtonian mechanics could be presented in a way that did not include mathematics.[46] In this strict nominalist – *fictionalist* – account no ontological commitment to mathematical entities is made, and all talk of them is essentially fiction. Hence, when applied to formal systems, it makes no sense to speak about the truth of mathematical sentences outside the proof of them. Mathematics consists merely of conventional axioms and rules of symbol manipulation, and what we mean by truth is defined exhaustively[47] by the rules of proof. This is deflationism about mathematical truth.

At this point we should look at the various terminological questions that Field's philosophy presents us with. Nominalism and fictionalism are without doubt views that can be attributed to Field, as apparent from Field 1980. After that the matter becomes a bit more vague. In Field 1980 (p. 1), he writes:

---

[46] We will return to some of the merits and weaknesses in Field's program later on in Chapter 6.4. See also Shapiro (1983) and Hale 1987, pp. 102-122 for critical assessments of Field's program.

[47] Meaning that all cases of true sentences are given to us by the rules of proof.

> In defending nominalism I am denying that numbers, functions, sets or any similar entities exist.

In the same book he continues (p. 15):

> What makes the mathematical theories we accept better than these alternatives to them is not that they are true [...] but rather that they are more *useful*. [...] Thus mathematics is in a sense empirical but only in the rather Pickwickian sense that it is an empirical question as to which mathematical theory is useful. (Italics in the original.)

In Field 1998 (p. 295), he writes:

> Anti-objectivism has considerable plausibility for the typical undecidable sentences of set theory. It has much less plausibility for the undecidable sentences of elementary number theory

As I see it, these quotes suggest three different philosophical theories. The first quote seems to describe strict fictionalism and extreme formalism. The second one implies that Field is actually not an extreme formalist, since there are empirical reasons for us to accept some mathematical theories over others. Finally, the third quote suggests that Field is an extreme formalist over set theory while suggesting a mild form of realism – or at least remaining cautious – concerning arithmetic. Add to that the fact that in the first quote numbers and sets are lumped together as something fictitious, and it becomes clear that we can make several interpretations of Field's philosophy.

The second quote seems particularly problematic. If all mathematical entities are fiction, how can one explain the usefulness of certain mathematical theories? We will return to this problem in Chapter 4.1, but for now it suffices to point out that empirical usefulness seems to conflict with the whole concept of fictionalism. After all, if there are good reasons for thinking that a sentence φ is useful, can these same reasons not be seen to suggest that it is also *true*? Moving to the third quote, how do we distinguish between natural numbers and sets as mathematical objects, and which do we consider to be philosophically more

important?[48] These are all difficult questions that deserve closer examination – which they will receive in Chapter 6 of this work. However, at this point we must make a choice and attribute a specific philosophical viewpoint to Field. Since he is most adamant in his fictionalism, nominalism and the rejection of substantial truth, I will consider them to be the central elements of his philosophy. Also, because he (Field 1999) defends deflationism in the debate we are about to go through, I believe we can attribute to Field the philosophical position that proof and truth are the same concept, that is, extreme formalism.[49]

When we compare Tarski's and Field's views, it must first be pointed out that in Field's philosophy there is no need for a Tarskian definition of truth in mathematics. It should come to us as no surprise that he sees little philosophical worth in the notion of semantic truth, or semantics in the philosophy of mathematics altogether. For Field these are completely heuristic devices; philosophically they are superfluous.[50] One implication of this is that undecided (or undecidable) sentences cannot be thought to have determinate truth-values. Under this conception, Fermat's Last Theorem[51] only became true in 1994 when Andrew Wiles proved it; before that it did not have a determinate truth-value.

---

[48] See Field 1998 (pp. 294-306) for his analysis on the question.

[49] If mathematical entities are fiction and mathematical truth plays no role, which I consider to be the two most important facets of Field's philosophy as far as the subject of this work is concerned, I do not see a problem with this characterization of Field. Indeed, in Field 2001, p. 317 (see Chapter 6.6 of this work), for example, he quite clearly suggests that mathematics is not about anything, which I can only understand as extreme formalism. We will return to his philosophy later on in this work, but for now, I believe the terms extreme formalism, fictionalism and deflationism can all be used for the philosophy of Field.

[50] See Field 1999, footnote 3.

[51] The statement according to which for an integer $n > 2$, the equation $a^n + b^n = c^n$ has no solutions when a, b, c are non-zero integers.

Needless to say, this kind of viewpoint must hold truth and proof to be the same concept.[52]

Unsolved problems, like Fermat's Last Theorem used to be, are only one kind of undecided sentences. The other kind can of course be seen in the undecidable Gödel sentence *G*(**T**). The important difference becomes apparent when we think about the truth-value of Fermat's Last Theorem before Wiles' proof. While we had a good guess – like we now have a good guess about the truth of Goldbach's Conjecture[53] – we had no way of establishing the truth of it. Hence, it is a sentence that was, in the extreme formalist account, undecided for a long time and became true in 1994. Gödel sentences, however, are different. As Tarski proposed, and our informal account in Chapter 2.7 suggests, *G*(**T**) is *at the same time* undecided in **T** and true, although the latter only in an expanded system. If this indeed turned out to be the case, and we would be justified in making the expansion, it would obviously contradict with Field's deflationism.

This discussion relates closely to the one on the general nature of truth. In the deflationist account of Paul Horwich (among others)[54] truth is not a robust, metaphysically substantial property. Deflationism has an interesting history in philosophy, and its most famous original proponents include Frank Ramsey, A.J. Ayer and Rudolf Carnap[55]. It comes in many guises and names; one deflationist is not necessarily bound to agree with another. There is also a difference between deflationism over truth in general,

---

[52] See Field 2001, pp. 332-343. The point of view that Field opposes to (at least for set theory) – that sentences can be thought to have determinate truth-values even though we have no guaranteed way of establishing them – is called *semantic realism* in literature. The most notable opponent of semantic realism has been the intuitionist Michael Dummett. See Dummett 1978, Chapter 1 for his position, and Hale 1977 for criticism. We will return to semantic realism later in this work.

[53] The statement according to which every even integer greater than 2 is the sum of two primes.

[54] See Horwich 1998, 1-8 for an outline of a general deflationist account of truth.

[55] Carnap 1956 provides an illuminating account of deflationism in its original form.

scientific truth and mathematical truth. All this makes deflationists a heterogenic group, perhaps more so than is usually recognized. However, following Quine (1990), one central claim of the more recent deflationism is that Tarskian theories of truth are *disquotational*, that is, the T-scheme of Tarski only adds quotation marks around the sentence and then calls it a definition (or an adequacy condition) of truth. To say that "snow is white" is true (in the object language), according to them, asserts nothing more than saying *snow is white* (in the metalanguage) does, and the mentioning of truth is redundant. Take the T-scheme:

"Snow is white" is true if, and only if, *snow is white*.

For the deflationist, simply saying "snow is white" carries all the information that the T-scheme does, and the whole need for different languages and definitions of truth vanishes. Hence the name deflationism: the assertion of the *truth* of a sentence is deflated into the mere assertion of the *sentence*, thus making truth a redundant concept.[56]

When it comes to mathematics, the natural way to deflate truth would seem to be equating it with proof.[57] After we prove a sentence φ in a formal system **T**, saying that φ is true does not give us any new information. Whether we agree on deflationism generally or not, in the philosophy of mathematics it seems to have a lot going for it. After all, provability is the thing we are after in formal mathematics, and truth could be considered merely a translation of this. If formalism is all there is to mathematics, and we could explain why we prefer some formal systems to others,

---

[56] Here one must distinguish between redundant in the sense that the phrase "…is true" can be left out, which was the original deflationist view of Ramsey (1927), and what Michael Williams (1999) calls "extended" disquotational truth where truth serves a purpose, but is still deflationary. Field's approach is the latter.

[57] This is the approach I will take, but there are various ways of understanding deflationism, even about mathematics. See Ketland 1999, pp. 69-79 for a good overview of the different types of deflationism in mathematics.

this would seem to complete the philosophical picture with very little ontological burden. However, there remains the problem of Gödel sentences, where a theory of truth *does* seem to give us new information on true sentences. This so far vague argument has been made exact, independently of each other, by both Shapiro and Ketland. We will now move on to them.

3.2 Shapiro's semantical argument

Shapiro and Ketland have been the main modern critics of deflationism in mathematics. Shapiro's argument is based on the claim that if deflationism were correct, truth would have to be *conservative* over formal theories, that is, adding a truth predicate to a formal theory could not give us any new true sentences. To be exact, if we add a theory of truth to our formal system **T**, this new augmented system **T'** must not allow us to derive the truth of any theorems that were not derivable from **T** alone. If truth were *not* a conservative property, then there would be some theorem γ the truth of which can be derived from **T'** but not from **T**. This would mean that it is logically possible that all the axioms of **T** are true but γ would be false in it. However, it would not be logically possible that all the axioms of **T'** were true but γ would be false in it. Thus, truth would have to be a robust, and not an insubstantial, property; after all, the addition of truth predicate made $\neg\gamma$ impossible. The move from **T** to **T'** clearly added something to the system.[58]

Shapiro's argument up to this point seems to be valid. If this indeed turned out to be the case, then deflationism runs into trouble – for we can now see that adding Tarski's truth definition to a consistent formal system **T** makes us able to see the truth of the Gödel sentence $G(\mathbf{T})$. Let us take a consistent formal system containing arithmetic **T** (in a language ***L***), and add Tarski's T-scheme (or if that fails, some other adequate truth condition) as an

[58] See Shapiro 1998, pp. 497-507 for the full argument.

axiom[59] to the system. We can move to examining this new system **T′** (in a language *L′*). Now **T′** would be conservative over **T** when it comes to proof, because no new rules of inference were included in the augmentation. Hence, if all the axioms and theorems in **T** were true, then in **T′** with an adequate truth condition we would be able to establish their truth. So in the system **T′** it would hold that:

(1) $\mathbf{T'} \vdash \forall x(\mathrm{Pr}(x) \rightarrow Tr(x))$,

where $Tr(x)$ is the truth predicate of **T′** and Pr(x) is the proof-predicate of **T**. Effectively, (1) says that all the theorems of **T** are true in **T′**. Now let us take any contradictory statement $\overline{\Lambda}$ of **T** (like $\overline{0=1}$, which is clearly contradictory in arithmetic). Clearly one of the T-sentences would tell us that $\mathbf{T'} \vdash \neg Tr(\overline{\Lambda})$, for **T** is assumed to be consistent. Because it is the case that $\mathbf{T'} \vdash \neg Tr(\overline{\Lambda})$, we get from (1) that $\mathbf{T'} \vdash \neg \mathrm{Pr}(\overline{\Lambda})$. But $\neg \mathrm{Pr}(\overline{\Lambda})$ is equivalent with the sentence Con_T, expressing the consistency of **T**. Now if the conservativeness claim of the deflationists were true, it would follow that $\mathbf{T} \vdash Con_T$. However, Gödel's second incompleteness theorem states that $\mathbf{T} \nvdash Con_T$. This leaves us with two options. Either our initial assumptions formalized as (1) are wrong, or **T** is not conservative over an adequate truth predicate. Looking at the way (1) was formulated, the initial assumptions do not seem to have anything troubling in them. Hence, the problem must be with the conservativeness assumption, and deflationism fails. This is Shapiro's semantical argument against deflationism in a nutshell.

---

[59] This is possible either by adding all the T-sentences as axioms, or by Tarski's own way of adding the truth definition as a set of rules. The latter would have the obvious advantage of always being finite. See Tarski 1944 for details. An important matter concerning the truth-predicate will turn out to be whether we want it to apply the schema of mathematical induction to formulas containing the truth predicate. Shapiro's argument rests on the assumption that this is indeed the case. We will return to the potential problems concerning this shortly. See Halbach (2001) for more on the different ways of expanding formal systems with a theory of truth.

Of course, through all this we have been talking about a consistent formal system **T** containing arithmetic, and at this point we must return to one of the central assumptions behind Gödel's incompleteness theorems: consistency. For Field (2001, pp. 343-350; 2006), that is a recurring problem in claiming *G*(**T**) to be true, and thus with the whole semantical argument: we can only know the truth of *G*(**T**) if we can *know that the system **T** is consistent*. This is something, Field argues, that we cannot convincingly state of our mathematical systems, and hence the semantical argument does not hit its target.[60]

This should not surprise us. Gödel's second incompleteness theorem states that consistent formal systems containing arithmetic cannot prove their own consistency. If we follow extreme formalism in equating proof with truth, we should be pretty confident in claiming that no consistent formal system containing arithmetic can ever be established. Gödel *proved* this. So strictly speaking, Field's argument here does not add anything that we did not already know. However, it *does* remind us of an important assumption. We assume that Peano arithmetic is consistent, or at least that there could exists some other adequate formal system of

[60] To be exact, here Field is concerned with "our fullest mathematical theory". Although this concept includes the original Hilbertian idea of single-system formalism, it is more convenient to use formalized (Peano) first-order arithmetic instead for the sake of simplicity. Making this choice, I accept the possibility of proving sentences of PA *outside* PA, that is, essentially in our fullest mathematical theory. However, such a theory would have its own Gödel sentence, and we would be back at the starting point. The differences between single-system formalism and milder versions of it will be a subject in the next chapter. For Field's treatment of the concept of "fullest mathematical theory", see Field 1998. His tentative suggestion for the concept (p. 302) is "the set consisting of all our explicit mathematical beliefs, plus perhaps those mathematical sentences we could easily be brought to believe explicitly, plus perhaps their logical consequences". This should give an idea of what Field means, as well as of the vagueness that the concept ends up having.

arithmetic that is consistent. Since we obviously cannot prove it, this assumption is the best we have.[61]

How big a problem is this for the semantical argument? The consistency of PA is something we cannot hope to establish indisputably due to Gödel's proof, and that way the truth of Gödel sentences might seem to escape us forever. However, it must be remembered that we are not talking about just any old property of formal mathematical systems. If arithmetic turned out to be inconsistent, surely it would present us with much graver problems than incompleteness does. In inconsistent systems we can prove any theorem, including Con_T, $G(\mathbf{T})$ and $\neg G(\mathbf{T})$. Moreover, if we do not believe in the consistency of our mathematical theories, we cannot believe in any of the proofs we acquire with them, whether it is Fermat's Last Theorem or an elementary theorem about addition. This way, the lack of consistency proof for PA should not be thought to imply in any sense that it is somehow likely that PA *is* in fact inconsistent. In the strict formalist sense, Field's criticism still remains unanswered – since we cannot know of consistent formal systems that they are consistent – but could we ever *practise* arithmetic without believing it to be consistent?

Field (2001, p. 349) acknowledges that it is of course reasonable to *hope* that our theory of arithmetic is consistent, and hence $G(\mathbf{T})$ true, but this does not make for a strong case for it actually being so. Strictly speaking Field may be right, but here I must strongly disagree with the strict sense being the *relevant* sense. I think that Shapiro's argument carries only the slightest of burdens because of the unprovability of consistency. In fact, it seems that, overall, Field damages his own case more than that of Shapiro's. It is true that the semantical argument does not apply if there cannot be such things as consistent formal systems containing arithmetic, that is, if PA and all other adequate axiomatizations of arithmetic are inconsistent. But here something very strange must happen for us to end up with the belief that there are in fact *no* consistent

[61]Some mathematicians, however, are ready to accept Gerhard Gentzen's (1936) ordinal analytic proof of this. It should also be noted that there are efforts to express consistency in ways that avoid the second incompleteness theorem. See Detlefsen 1980 for an example.

formal systems containing arithmetic. Field admits that it is reasonable to hope, and even have "positive belief", that arithmetic is consistent; I claim that we *must* have positive belief for it to be so. After all, that we can claim 2 + 2 = 5 to be false follows from the very same belief in consistency as the truth of $G(\mathbf{T})$.

If we go deep enough in our skepticism and doubt the consistency of all systems of arithmetic, what would be the point of discussing arithmetical truth and proof in the first place? Indeed, this is particularly damaging for the extreme formalist case. For the realist there always remains the possibility that our formal systems are fundamentally inconsistent, but we can trust them as long as they somehow give us *true* sentences. But for the strict formalist, proof is the only method of acquiring and recognizing true sentences. If we do not believe that the rules of proof and axioms are consistent, we cannot believe in any of our arithmetical statements, whether it is that 2 + 2 = 5 is false or $G(\mathbf{T})$ is true. Clearly there must be limits to how far we can go in the fear of inconsistency.

Certainly there does not exist a generally accepted proof for the consistency of PA – that is not the question here. However, as far as human knowledge goes, the consistency of arithmetic and the Peano axioms have to be among the least unreasonable assumptions. While we should be safe in making the assumption, of course it is still important to make it explicit. In this work, PA is assumed to be consistent, as is commonplace in mathematics and the philosophy of mathematics. To be more precise, we are assuming that *some* axiomatization of arithmetic is consistent. Perhaps PA is inconsistent after all, but the arguments throughout this work are not tied to it, but rather the more allowing concept of "consistent (and adequate) formal system containing arithmetic". Essentially, we are assuming that there is some way of correcting PA, or another theory of arithmetic, to be consistent. As I see it, following the account above, this amounts to nothing more than assuming that we can state that 2 + 2 = 5 is false.

Finally, we must remember that Gödel's incompleteness theorems do not hold for inconsistent formal systems. Obviously we would never even know that $G(\mathbf{T})$ is not provable if **T** were not assumed to be consistent. Shapiro's argument would not be valid,

but there would not be grounds for any such argument in the first place, since Gödel's proof of the existence of undecidable sentences would not apply. Field's point is that we cannot unassailably know that our formal systems containing arithmetic are consistent, and in this he is correct. But of course we can *assume* that there are consistent formal systems, and Shapiro's argument retains the exact same weight. If there can exist even one formal system containing arithmetic that is consistent, then by expanding it with Tarskian truth we establish a true unprovable sentence. Of course in philosophy and mathematics one can never be too careful, but this is definitely one question where the burden lies in showing all formal systems of arithmetic to be *inconsistent*. Safe to say, if the opposite were true – that is, if all axiomatizations of arithmetic were inconsistent (or inadequate) – mathematics and the philosophy of it would have much more serious problems than the ones considered in this work. To deny the truth of $G(\mathbf{T})$ based on the lack of consistency proof is to deny the possibility of talking about truth and falsity of *all* arithmetical statements, whether we understand the concepts as deflationary or not. That should be reason enough to move on to other issues.

There is one more thing we need to address here. One possible way around Shapiro's semantical argument is that we construct arithmetic in another formal system, and we can prove the Gödel sentence in this new system. One candidate for this is set theory. But this approach does not provide us with anything new. First of all, it is not at all certain that we could prove $G(\mathbf{T})$ in set theory (for example). But more importantly, even if we could, this new system would still contain arithmetic, and thus have its *own* Gödel sentence $G(\mathbf{T})'$ which could not be proved. If a consistent formal system has enough expressive power to contain arithmetic, Gödel's incompleteness theorems apply, and there is no way around them.

So far Shapiro's argument seems to be very powerful. One must always be careful not to make fantastic claims based on Gödel's incompleteness theorems, as we will see in the final chapter of this work, but it cannot be said that Shapiro makes that error. He makes an explicit argument of what we already considered to be intuitively obvious earlier: semantic truth gives us a way to see truths that we cannot acquire with formal proof. These are very

special kind of, and in mathematical practice rather uninteresting truths, but truths nevertheless. Mathematical deflationism seems to be in deep trouble if we include Tarskian truth, and Shapiro has shown us exactly why this is the case.

3.3 Counterarguments beyond consistency

At this point, I think we can be confident to set aside the problems concerning the possible inconsistency of arithmetic. Let us now move on to other criticisms of Shapiro's argument. Field's (1999, pp. 533-534) other answer to Shapiro is concerned with a perceived ambiguity between the notions of metaphysical and *expressive* thinness. To see this we must examine what Field means by these concepts, and what his basic concept of deflationism is. Field claims that instead of "metaphysical thinness"[62], Shapiro argues against "*expressive* thinness", and that is something Field does not disagree with. Truth is, in Field's (ibid., p. 533) account, a device for making "fertile generalizations". These are of the kind:

> Everything the Pope has said so far has been false. Hence, everything that the Pope says is false.

Thus we commit ourselves to the truth (or falsehood, rather) of remarks that we have not heard. Since we do not know what these remarks will be, we cannot disprove them, and so our original sentence has more expressive power than what could be attained without the concept of truth (falsehood). This is what Field accepts as the main feature of the notion of truth: it has expressive power that we would not have without it, and that is why truth is not *expressively* thin.

Deflationism, following that account of truth, is the notion that:

---

[62]Another deflationist Jody Azzouni gives a rather eloquent and telling description for metaphysical thinness: "truth is not a metaphysically substantial property, and it has no nature." (Azzouni 1999, p. 540.)

> …truth is a purely logical notion applicable only to sentences we understand, and serving solely as a device of generalization…(ibid., p. 534).

The example of the Pope's utterances illuminates this viewpoint well, although the concept of "understanding" looks somewhat obscure, at least when it comes to mathematics. We will return to that, but let us grant Field this definition and see what follows.

Field accepts Shapiro's argument that **T′** is not conservative over **T**, but points out that the argument depends on the way we add the truth predicate to **T**. Perhaps it is best to quote Field on what he means by this (only the symbols are translated to fit the ones used in this work). The following is Field's conception of the truth predicate:

> (i) [Let us assume] that for each of the finitely many atomic predicates p of **T, T′** licenses the usual general rule about how the truth of any sentence of form $p(t_1,...,t_n)$ depends on the denotations of $t_1,...,t_n$; and that for each method of composing sentences out of simpler sentences, **T′** licenses the usual general rule how the truth of compound depends on the truth of the components. (ibid, p. 534).

What Field refers to is basically the usual Tarskian scheme of truth, with induction over the structures of formulas containing the truth predicate. Shapiro accepts this, and a truth definition like (i) is indeed conservative over **T**.[63] But according to Shapiro (1998, pp. 497-498), this is not sufficient, for it lacks the property of *mathematical* induction. We also want our truth definition to include generalizations over truth, such as "all the theorems are true". This requires mathematical induction over formulas containing truth, and is not included in a notion like (i).

So something more is needed for Shapiro's account to succeed. Field (Field 1999, p. 535) formulates it as follows:

---

[63] See Halbach 1999 Lemma 2.1 for proof that Peano arithmetic with full induction over *formulas* involving the truth predicate is conservative over PA.

(ii) [Let us assume] that **T′** allows mathematical induction on formulas containing the truth predicate.

This is the decisive step in Shapiro's argument, for he made use of the statement "if all the axioms and theorems in **T** were true, then in **T′** with an adequate truth condition we would be able to establish their truth", that is, $\mathbf{T'} \vdash \forall x(\mathrm{Pr}(x) \rightarrow Tr(x))$.

It is the truth predicate containing both (i) and (ii) that Shapiro wants, and with it we can know arithmetical truths that we otherwise could not, those of Gödel sentences. That is why, Shapiro argues, truth is not a metaphysically thin concept. Field (ibid., p. 536), however, thinks that this is only a case of making a generalization like the one he did with the utterances of the Pope. As such it would not be a problem for the deflationist, since truth is certainly not claimed to be expressively thin. What Field and Shapiro disagree on, according to Field, is whether such generalizations are metaphysically thin.

Before we go any further, a couple of ambiguities must be clarified. Field's original concept of deflationism was the viewpoint: "…truth is a purely logical notion applicable only to sentences we understand, and serving solely as a device of generalization…", aided with the example of the Pope's utterances. There are two difficulties in this. Firstly, there is an obvious difference between mathematical and *empirical* induction. Mathematical induction is hardly similar to the case of the Pope, for we *know* for sure that if the induction principle holds, then proof by mathematical induction is just as valid as any other method of proof. Of course, while empirical induction at its best gives us probable knowledge, mathematical induction over natural numbers is in fact completely deductive. Generalizations over the Pope's utterances are not generalizations in the same sense, and the difference is an important one here. We commit to the falsehood of the Pope's future statements, but of course there remains the possibility that our commitment is mistaken. That is also the reason why truth is not expressively thin. However, when proving a sentence $\varphi(n)$ with mathematical induction, there is no possibility of $\varphi(n)$ being false for some n. With mathematical

induction we *know* the future cases to be true, while with empirical induction that is just a tentative prediction. As such empirical induction is merely a heuristic device, and can be conceived as metaphysically (although not expressively) thin. But mathematical induction is one form of mathematical *proof*. With it we do not make predictions concerning the truth of sentences, we *establish* them. The difference here is important, as it could very well lead to the difference between expressive and metaphysical thinness. In this sense, when it comes to mathematical truth, Field's example of the Pope's utterances is highly misleading.

Secondly, the notion "sentences we understand" needs to be clarified. In fact, when it comes to formalist mathematics, I can only conclude that Field means something along the lines of being well-formulated or provable. Any semantical account of understanding should not enter into the picture here, because that would right away warrant a semantical account of truth, which is of course the very thing that Field is trying to avoid. If Field does mean provability, then truth would mean generalizations of the type "we can prove $A(x)$ for $x = a$, $x = b$, $x = c$. Hence $A(x)$ is true for $x \in S = \{a,b,c\}$". But this property is not a very interesting one, and hardly the kind of truth mathematicians are after. It *is* a metaphysically thin notion of truth, however, and it resembles the falsehoods of the Pope's utterances. However, in this case truth would also be *expressively* thin. We must now extend the above example indefinitely, that is, "we can prove $A(x)$ for $x = a$, $x = b$, $x = c$. Hence $A(x)$ is true for $x \in S = \{a,b,c,d,e,...\}$". When the set S is infinite, we seem to have an analogy to Pope's utterances: we commit to truth of future cases. But where in mathematics do we see truth used in such manner? Clearly this is not a proof, but a conjecture. Goldbach's conjecture has been showed to be hold for all natural numbers $n \leq 10^{18}$, but no mathematician would claim it to be true for all n based on that. Such empirical induction does not belong to mathematics. However, that does correspond to the way Field uses truth to predict the Pope's utterances, and it makes truth a metaphysically, but not expressively, thin property. Unfortunately for Field, there does not seem to be anything in mathematics that corresponds to such a notion of truth.

In any case, mathematical induction is a whole other matter, and over formulas containing the truth predicate it gives us the truth of $G(\mathbf{T})$, a sentence of PA. Moreover, it does not do this by conjecture. If truth is still metaphysically thin when it unassailably gives us a new truth of PA, then it is beginning to seem that *proof* is also metaphysically thin – a point of view perhaps acceptable in extreme formalism, but certainly not one that is likely to convince doubters. It amounts to claiming that truth is metaphysically thin because *everything* in mathematics is metaphysically thin. Such a concept of metaphysical thinness is hardly interesting anymore, although it could very well be just what Field is after in his fictionalism. If that is the case, however, it is hard to see the point of going into any debates on truth and proof in the first place.

So it seems that mathematical induction, not the different notions of thinness, is the key question here. As we saw from Shapiro's argument, it was the principle $\mathbf{T}' \vdash \forall x(\mathrm{Pr}(x) \rightarrow Tr(x))$ that made $\mathbf{T}'$ a more powerful system than **T**. This is not a T-instance of a sentence of **T**, so the theory of truth needs to do more than enumerate all the T-instances. Namely, it must be able to express that it *does* enumerate all the T-instances. There are two arguments for this. According to Ketland, this is a direct consequence of the Tarskian theory of truth, for Tarskian truth is defined as giving us *all* cases of true sentences.[64] The other one is the point (ii) that Field argues against, mathematical induction over sentences containing the truth predicate.

By now we can forget the Pope's utterances as a false lead for the deflationist. As far as the point (ii) of Field is concerned, the situation looks somewhat more promising for him. Rather than deal with metaphysical and expressive thinness, he can question the justification for mathematical induction in the first place. Indeed, Field's (1999, p. 537) other argument is that we can expand **T** with a truth predicate also in ways that are conservative – as long as these ways do not include the principle of mathematical induction over formulas containing the truth predicate. Field

[64] See Ketland 1999, pp. 79-91 and Tarski 1956 (Definition 23) for more on this.

argues that nowhere in a theory of truth should we include mathematical induction over it. In short, truth is truth and arithmetic is arithmetic: mathematical induction is a property of the latter, why should it also be one of the former?

For Shapiro the principle of mathematical induction is essential to mathematical truth, and this was used in his argument against the conservativity of truth. It is essential to truth because we want to be able to show that if the axioms of a system **S** are true, and the rules of inference of **S** preserve truth, then all the theorems of **S** are true. But Field (ibid., p. 538) claims that this contention is not acceptable. He states that because we want to be able to show that, the induction principle is something just added to the theory of truth for our convenience. Therefore, Shapiro's argument is wrong to rest on such weak assumption. What Shapiro would need, according to Field, is to show that the induction principle follows *only* from the nature of truth.

Of course for Field truth is deflationist, so the induction principle will not be likely to follow from truth only – by definition *nothing* follows from deflationist truth only. In fact, he (ibid.) correctly points out that the induction principle follows from a property of arithmetic, namely that natural numbers are ordered and always have finitely many predecessors. As such, the induction principle would be essential to *all* predicates of arithmetic, not just truth. Since the induction principle would not be an essential quality of *truth*, but rather a quality of *arithmetic*, Shapiro's argument would fail. After all, we could formulate truth as another predicate – one that does not conform to mathematical induction.

However, here Field seems to neglect the main part of Shapiro's argument. Obviously truth is not *only* a predicate of arithmetic. With truth we could establish the truth of Gödel sentences, which we could not do in PA alone. This should be enough to convince us that truth is not a predicate of arithmetic. In fact, already by Tarski's undefinability result, no predicate of PA is equal to the truth predicate of PA. That arithmetical truth has the property of mathematical induction *common* with the arithmetical predicates should not be a problem. Indeed, we must remember that we are talking specifically about *arithmetical* truth here. Surely there

cannot be much controversy in including the induction principle in the concept of truth when it comes to arithmetic. For Bertrand Russell (see Russell 1920, pp. 20-28) this was actually the definition of natural numbers: natural numbers are the set that satisfies the principle of mathematical induction. If natural numbers can be defined as the set that satisfies the principle of mathematical induction, we should not have any problem including mathematical induction over the *truth* of them. In fact, whatever we want from arithmetical truth, for it to satisfy one of the Peano axioms (for that is what mathematical induction is) seems to be one of the weakest possible assumptions.[65]

There does not seem to be any reason why the notion of arithmetic truth could not draw from arithmetic. In fact, the opposite viewpoint would seem quite peculiar. Volker Halbach (2001, pp. 187-188) has pointed out that *nothing* in a theory of arithmetical truth depends *only* on the nature of truth. All the T-sentences depend also on the nature of numbers. This is an important point: the class of the true sentences of PA of course depends on the nature of natural numbers. As it turns out, mathematical induction is a crucial property of natural numbers. To claim that the truth of natural numbers could not include mathematical induction over formulas containing the truth predicate seems totally unacceptable from this background.

However, all this must not be confused with truth being a *property* of the natural numbers, like Field seems to claim in his counterargument. Truth is still a property of the expanded system of PA *added with Tarskian truth*; it simply includes one important property of PA, that of mathematical induction. It seems that Shapiro is partly guilty of causing this confusion. He argues (1998, pp. 500-501) for the induction principle as something that follows from arithmetic. He does this to show that adding the truth predicate to arithmetic does not change the subject, because we can use the induction principle to the truth predicate. He quotes Shaughan Lavine (1994):

---

[65] A similar point has been made by Hyttinen & Sandu (2004, p. 417).

> …to define a property of natural numbers is to be willing to extend mathematical induction to it…

Obviously we should be ready to accept this, but it indeed helps to make truth prone to Field's treatment of it as just another *property* of the natural numbers. If truth were (in principle) indistinguishable from other properties, Field would have a case that truth is indeed metaphysically thin. But this is not the case. When we moved from **T** to **T′** (and from ***L*** to ***L′***), we added something new to the system. Shapiro wants to show that we did not change the game, and hence ends up implying that arithmetical truth is a predicate of the natural numbers. Of course it is a predicate *concerning* natural numbers, but it is not a predicate of **T**; it is one of the expanded system **T′**.

It seems obvious by now that a theory of arithmetical truth should be able to draw from both truth and arithmetic, and when we talk about arithmetical truth, we are always concerned with the expanded system **T′**. Field claims that truth itself should be enough, while Shapiro sometimes seems to give the impression that arithmetic itself is enough.[66] We must expand the scheme of mathematical induction from **T** to **T′** for the semantical argument to hold, and this way we are of course changing the game a bit.[67] But if the deflationist theory of truth is inadequate, as Shapiro argues all along, what is wrong in changing the game, especially as the change is this insignificant?

In another facet of the same subject, Jody Azzouni (1999, pp. 402-403) claims that what Shapiro is doing does not really concern deflationist truth, since the deflationist does not need such truths and generalizations as Shapiro's $\mathbf{T'} \vdash \forall x(\mathrm{Pr}(x) \rightarrow Tr(x))$. According to him, a theory of truth has no need to establish generalizations *about* truths. Sentences like this go beyond the scope of deflationary truth, and so Shapiro's argument does not hit its target. Perhaps one might be ready to consider Azzouni's criticism

---

[66] It must be stressed that these are not actual arguments of Shapiro, only impressions a that less than careful reader can get of them.

[67] See Ketland 2005, p. 78 for his argument to the same effect.

to be valid, but only if we had some reason to believe that deflationary truth is indeed what mathematical truth is – the success of Hilbert's program would have been at least a step toward that direction. For those who think that mathematical truth could be something else than deflationary truth, it seems rather limited that we cannot speak about truths in a theory of truth. In fact, $\mathbf{T'} \vdash \forall x(\mathrm{Pr}(x) \rightarrow Tr(x))$ in particular sounds like a very important theorem of arithmetical truth. While Azzouni grants that the theorem is obvious, he does not see any reason to regard it as a part of a theory of truth. It might not be, but it is hard not see it as something that should at least *follow* from an adequate and satisfactory theory of arithmetical truth.

Looking at Field's and Azzouni's criticisms, it might seem that the theorem $\mathbf{T'} \vdash \forall x(\mathrm{Pr}(x) \rightarrow Tr(x))$ simply looks too appropriate for Shapiro's purpose. However, there is always the possibility that the best solution also happens to be the most convenient one. That we can use the principle of mathematical induction like Shapiro does is admittedly very convenient, but that does not necessarily make it problematic. In fact, just the opposite seems to be the case. Considering the result "if axioms of a system **S** are true, and the rules of inference of **S** preserve truth, then all theorems of **S** are true", we would be hard pressed to find it unintuitive or otherwise problematic. As far as philosophical conclusions go, there cannot be many more plausible ones, and any satisfactory theory of arithmetical truth should include it as a consequence. Just because this result could not be acquired *within* **S** does not mean it cannot be justified by other means, like here by adding a Tarskian theory of truth.

For the deflationist there seem to be two options left. Either there is the path proposed at least implicitly both by Field and Azzouni of simply sticking to arithmetical truth as proof in PA. Truth cannot include mathematical induction – in fact, it cannot include anything except the T-instances of the theorems of PA. This is extreme formalism, and by default we are simply forbidden to add a non-deflationary theory of truth to it. Truth in such systems is only a translation of proof, and it is obviously deflationist. Other than noting the suspiciously arbitrary nature of

denying the use of mathematical induction over truth, in this context there is nothing more to be said about such formalist truth. When all the other counterarguments fail, the strict formalist can still claim that formal systems are all there is to mathematics, and hence acquire immunity against any strategy that includes expansions to formal mathematics. However, this denial of all expansions only works when we accept that formal systems *are* indeed all there is to mathematics. Unfortunately for the deflationist, when we think of mathematics as a human endeavour, we will find that such extreme formalism does not make sense. That will be a main subject for the rest of this work.

Aside from extreme formalism and the rejection of all expansions, the other remaining option for the deflationist is to find alternative means of establishing generalizations like $\mathbf{T}' \vdash \forall x(\mathrm{Pr}(x) \rightarrow Tr(x))$. It could be the case that we do not need a substantial theory of truth to arrive at that crucial generalization. This is an interesting possibility, and Neil Tennant's argument for it will be examined shortly. But before that, it must be pointed out that there is also a third option for the deflationist. Volker Halbach (2001, pp. 188-189) has proposed that even though deflationism runs into problems with conservativeness, it could still be correct on the grounds that deflationism does not *require* conservativeness. This is an interesting idea. Essentially, Halbach is saying that the deflationist cannot define truth in a way that would remain conservative, but this is no problem. For Halbach, it is enough for deflationist truth that it succeeds in its purpose, that is, providing us with generalizations of the $\mathbf{T}' \vdash \forall x(\mathrm{Pr}(x) \rightarrow Tr(x))$ kind. The fact that these generalizations are not conservative over the formal system **T** does not matter; that deflationist truth has substantial consequences is "merely a secondary effect". In this way, Halbach concludes, deflationist truth is "not innocent, but that does not mean that it is wrong".

Halbach's position is very different from Field's. While Field's argument for deflationism seems to be refuted by the lack of conservativeness, Halbach's is bound to prevail against just about any counterargument. In fact, the argument looks suspiciously *too* strong. Whatever follows from deflationist truth, it will not matter

as long as it also serves its purpose of enumerating all the theorems of arithmetic as true (and stating this). Thus establishing the truth of Gödel sentences is simply another "secondary effect", and not part of the deflationist truth. But surely this is too simplified. If our theory of truth has as its secondary effect new true sentences, on what grounds can we distinguish between this secondary truth and the primary one we were defining? The difference between primary and secondary effects seems very much arbitrary: after all, they both follow from the same theory of truth. In this way, Halbach's position resembles that of Azzouni's: they limit the domain of deflationist truth in ways that the arguments like Shapiro's cannot touch. It is misleading to say that deflationist truth defined this way is not innocent – I would say it has been granted full immunity from prosecution. The problem is that this immunity is given in an arbitrary fashion.

3.4 Why not deflationary truth?

Jeffrey Ketland (1999) has provided independently of Shapiro a similar semantical argument against deflationism. His argument is fundamentally the same as Shapiro's, so we will not go through its details. But Ketland has some interesting things to say based on the argument, and we should take a look at them. Most importantly, there is one question that we need to answer: when it comes to arithmetic, substantial or not, why should we prefer a Tarskian theory of truth to a deflationist one? This is where we ended up with the arguments of Field, Azzouni, and in a different manner, also Halbach. The deflationist can claim that even though Tarskian truth is not conservative, this is not a problem since the deflationist can insist that deflationary truth is the only *correct* theory of truth. To make full use of the semantical argument, we will need to show that deflationist truth is not acceptable as a theory of mathematical truth.

To this effect, Ketland points out that we should abandon deflationist truth based on the very fact that it is *incomplete*. As we know from Gödel's theorems, in a deflationist theory of truth we

cannot have such desirable theorems as "For any closed formula γ, $\neg\gamma$ is true if and only if γ is not true" in consistent formal systems containing arithmetic. A deflationist theory of truth fails in many such cases, Ketland (ibid., pp. 85-86) argues, because the formal system cannot prove its own consistency. Of course one such case is the Gödel sentence $G(\mathbf{T})$. The truth of $G(\mathbf{T})$ is forever outside the bounds of the formal system it was formulated in. According to Ketland, deflationary truth is simply not satisfactory, because so much of what we expect from a theory of truth does not follow from it. We must remember that the main purpose of Tarskian truth is to provide an *adequate* definition of truth – which is something that deflationist truth fails to do.

However, once we add Tarski's theory of truth to the system, this is no longer a problem. In a manner similar to Shapiro, Ketland (ibid., p. 87) shows that $G(\mathbf{T})$ is deducible from this strengthened theory: if **T** is the original consistent formal system containing arithmetic, and **TS** Tarski's theory of truth, then $\mathbf{T} \cup \mathbf{TS} \vdash G(\mathbf{T})$. This corresponds to Shapiro's argument, and the difference in details is not important. When we recognize the truth of the Gödel sentence, we obviously do not do it within the formal system. Gödel proved that consistent arithmetical systems cannot prove their own consistency and are incomplete. If we equate truth with proof – as deflationists do when it comes to mathematics – the deflationist theory of truth must be incomplete as well. This is obvious even without the semantical arguments. But like Shapiro, Ketland (ibid., pp. 87-88) argues that the Gödelian results show that a non-conservative, Tarskian, theory of truth "significantly transcends the deflationary theories". Only the Tarskian theory of truth is adequate, because with it we can deduce the truth of the Gödel sentence, as well as that of other sentences requiring consistency. For Ketland that is one main reason to prefer Tarskian truth to deflationism.

Both Ketland and Shapiro make a convincing case. Tarski's theory of truth transcends the deflationary theories, which certainly helps to make it more appealing. When all the deflationist smoke has settled, it clearly is a drawback for a theory of mathematical truth that it turns out to be incomplete. Keeping this

in mind, Tarskian truth over arithmetic seems much more appealing than deflationism. But even so, the deflationist might have one more weapon in his arsenal. He could claim that the Tarskian theory of truth, while neither redundant nor incomplete, is nevertheless *false* when it comes to mathematics. Mathematics is about proof in formal systems, one could claim, and proof in formal systems *only*. In this way, the Gödel sentence would not actually be true, and the substantialist would be making false claims all along. The truth of Gödelian sentences would be a mere illusion: we would not be justified in making any declarations over truth that go beyond proof in formal systems – and if we reject all expansions to formal systems, that indeed must be the case.

That is why we should also make it clear that there is another reason for us to choose Tarski's semantic conception of truth: the fact that human beings *use* semantical thinking when they deal with mathematics. This is something that both Shapiro and Ketland fail to emphasize adequately, and it will be one of the main theses of this work. I will argue that the existence – and importance – of semantical thinking in mathematics should by itself be enough to justify the use of a semantical theory of truth. The formalist program and its followers may have distracted us, but the fact is that there exists overwhelming evidence of *pre-formal* mathematical thinking. Both historically and in terms of individual development it is the most basic mathematical thinking we have. Mathematics as a human phenomenon is *not* simply symbol manipulation in syntactic formal systems. In the upcoming chapters I will argue that Tarski's theory of truth gives us an adequate conception of truth for this pre-formal mathematics. A deflationist one, for the reasons that Shapiro and Ketland have made clear, does not.

This should be, more than the incompleteness of deflationist truth alone, justification for us to expand formal systems with a semantical notion of truth. As was noted in the previous chapter, while deflationist truth is not an expansion over proof when it comes to formal systems, Tarskian truth clearly is. If we categorically reject all expansions to formal systems, there is not much the anti-deflationist can do. The deflationist can undoubtedly use this as an argument against semantical truth, and

it seems that when all the other counterarguments are dealt with, that is indeed what they will return to. Without any justification for the expansion, the semantical arguments seem to be powerless against what I call the *final deflationist thesis* (FDT):

(FDT): If formal systems are enough for mathematics, we have no reason to introduce expansions like semantical theories of truth.

It is impossible to deny the power of conceptual simplicity in FDT. If one gives Occam's razor the sort of power it is often given, the formalist has a seemingly convincing case. If we consider formal theories like PA to be successful, as most of us are likely to, why should we worry about *expansions* to them? Indeed, I do not see any reason why FDT as a philosophical thesis is not valid. I will argue, however, that it is not *sound*: formal systems are *not* enough for mathematics. Although PA might seem like everything there is to arithmetic, when we examine mathematics as the full picture – including the pre-formal mathematical thinking – that is not the case. In fact, in that full picture of mathematics, Tarskian truth is not an expansion at all. Tarskian truth carries no extra conceptual weight simply because mathematics as we understand it could not exist if extreme formalism were correct.

That question will get a detailed treatment in this work, but let us put it on hold for a while. As far as the semantical arguments are concerned, the main advantage of Tarskian truth in arithmetic is its adequacy – to be precise, the ability to establish the truth of Gödel sentences. This is Ketland's and Shapiro's argument, and the deflationist's best reply so far seems to be ignoring it, and insisting on formal proof being everything there is to mathematical truth. However, after all the considerations so far, there is still another way out for the deflationist, one that looks much more promising. Neil Tennant (2002) has argued against Shapiro and Ketland that semantical truth, and the semantical argument we have followed, need not be semantical after all. Tennant defends the deflationist point of view (even though he's not a deflationist himself) by claiming that whatever we can express with truth predicates in the semantical argument, we can also express without them. What

appears to be semantical in the argument could also be achieved by completely deflationist methods. Essentially, according to Tennant, the arguments of Shapiro and Ketland are valid, but they make the wrong point.

3.5 Tennant

Tennant (2002) does not argue against Shapiro or Ketland on the basis that their contention of semantical truth is wrong. Instead, he admits that it is more or less valid, but not the *only* valid one. Tennant aims to show that the semantical argument can also be reached in a deflationary way. He claims to present another way of recognizing the truth of Gödel sentences (or, to avoid reference to truth, recognizing that they should be *asserted*, rather than denied), one without any reference to truth predicates. This way, truth would not need to be a substantial property, and Tennant would refute what he calls the "substantialist dogma".

Tennant (ibid., pp. 562-564) correctly points out that this "optional way" of coming up with the semantical argument cannot be just adding the truth of Gödel sentences as a new axiom in the expanded system **T′**. The semantical argument is powerful because it tells us *why* asserting *G*(**T**) is the right thing to do, that is, why we informally see the truth of it. It is this point that Tennant wants to reach with his argument, only by different means. He must avoid the use of a truth predicate, but still be faithful to the deductive structure of the semantical argument. This rules out a few possible solutions, as Tennant remarks, including ones using the consistency of **T** as the first principle of expanding **T**.

Of course some kind of expansion to **T** is needed in order to establish the assertability of Gödel sentences. What Tennant (ibid., p. 573.) offers as the least powerful but still sufficient expansion to carry out the semantical argument is the following *principle of uniform primitive recursive reflection* ($UR_{p.r.}$):

($UR_{p.r.}$): Add to **T** all sentences of the form $\forall n(prov_T(\psi(\overline{n}))) \rightarrow \forall m\psi(m)$, for primitive recursive ψ.

Here $prov_T(y)$ is equal to the earlier provability predicate $\Pr_T(y)$. Three things are required of $UR_{p.r.}$. First, it has to be sufficient for carrying out reasoning similar to the one in the semantical argument. Second, it must not be stronger than what is needed, and certainly not stronger than the addition of a truth predicate. Third, it cannot have any implicit connection to a truth predicate.

Actually, $UR_{p.r.}$ is really just a specific case of the so-called *soundness principle*. Solomon Feferman (1998, p. 233) has proposed the soundness principle (*SP*):

(*SP*): $prov_T(\overline{\psi}) \rightarrow \psi$

as a possible expansion of **T** that would suffice. Since in this work we have been dealing with primitive recursive functions ever since Hilbert's program, I am confident that we can move from $UR_{p.r.}$ to *SP* without any danger of damaging Tennant's case. Feferman thinks of *SP* as the "means of expressing faith in the correctness of **T** without any new predicates at all". I think that this is an appropriate description, and one that sounds intuitively like a promising candidate to replace Shapiro's $\mathbf{T}' \vdash \forall x(\Pr(x) \rightarrow Tr(x))$. In *SP*, truth of the sentence is just replaced with the mere assertion of the sentence. Since this is exactly what the deflationist program is all about, we should be making the right choice of expansion.

The structure of Tennant's (2002, pp. 576-578) argument is as follows. Let φ again denote $G(\mathbf{T})$. We can start examining our original Gödelian construction $\mathbf{T} \vdash \varphi \leftrightarrow \neg \Pr_T(\overline{\varphi})$. To give us some tools to manoeuvre with, we can define the provability predicate in the following equivalent ways:

$$\Pr_T(\overline{\psi}) =_{df} prov_T(\overline{\psi}) =_{df} prov_T(z,\overline{\psi}) =_{df} \exists z(prov_T(z,\overline{\psi}))$$

Here z belongs to the set of derivations (codes of proofs) in **T**. Now let x be an arbitrary derivation. First Tennant (see ibid. p. 577 for details) shows that in the system expanded with soundness, from Gödel's incompleteness theorem we know that x cannot prove $\overline{\varphi}$ in **T**. To this Tennant introduces the following representability theorem:

$$\neg prov_T(x,\overline{\varphi}) \rightarrow \exists y(prov_T(y,\neg prov_T(x,\overline{\overline{\varphi}}))).^{68}$$

Applying that theorem, we get: $\mathbf{T'} \vdash \exists y(prov_T(y,\neg prov_T(x,\overline{\overline{\varphi}})))$. But we must remember that x was arbitrary, and hence the sentence holds for all x, that is, $\mathbf{T'} \vdash \forall x(prov_T(\neg prov_T(x,\overline{\overline{\varphi}})))$. Now equipped with his $UR_{p.r.}$ Tennant can conclude that in the expanded system $\mathbf{T'} \vdash \forall z \neg prov_T(z,\overline{\varphi})$. Since the Gödelian construction can take the following form:

$$\varphi \leftrightarrow \neg\exists z(prov_T(z,\overline{\varphi})) \leftrightarrow \forall z \neg prov_T(z,\overline{\varphi}),$$

Tennant finally arrives at $\mathbf{T'} \vdash \varphi$, where $\mathbf{T'}$ is $\mathbf{T} \cup \boldsymbol{UR}_{p.r}$. So $\mathbf{T} \cup \boldsymbol{UR}_{p.r} \vdash G(\mathbf{T})$.

Those are the important technical steps in Tennant's argument. However, as Ketland (2005, pp. 81-82) points out, there is really no need for them. Accepting the soundness principle will automatically enable us to prove the Gödel sentence $G(\mathbf{T})$. As we constructed it, $G(\mathbf{T})$ is true (or assertable) if and only if $G(\mathbf{T})$ is not a theorem of **T**. If we assume that **T** is sound, then $G(\mathbf{T})$ is true if it is a theorem of **T**. Hence $G(\mathbf{T})$ is not a theorem of **T**, and $G(\mathbf{T})$ is true. So the soundness assumption implies $G(\mathbf{T})$, and we have no need for the more complicated argumentation of Tennant.

Now what should we make of Tennant's argument? Granted, he does not use reference to a truth predicate anywhere in it. But he

---

[68] See Tennant 2002, p. 560 for the theorem. This is based on the idea that if something is not provable, this unprovability can be proven.

does use his $UR_{p.r.}$ which enabled him to talk about the assertability of all the sentences of **T**. Since it is a form of the soundness principle, the strength of Tennant's argument ultimately lies only on the strength of that assumption. Unfortunately for Tennant, soundness is in fact a very strong assumption. For Shapiro and Ketland, the soundness principle *followed* from Tarski's theory of truth. Tennant simply *assumes* soundness to hold. Which assumption should we consider more acceptable? This question will be examined in the next chapter.

Before that, however, there is another question we must ask about Tennant's argument: how exactly is it an argument for *deflationism*? As I see it, and as Tennant himself mildly suggests (ibid., p. 579), the question is not really about choosing between substantialism and deflationism, but rather between substantialism and *prosententialism*. We know that *SP* is not derivable from **T** (without making **T** inconsistent), as Löb's theorem, following directly from Gödel, tells us.[69] Since **T** is assumed to be consistent, there is no problem, and *SP* is a true expansion of **T**. But how strong an expansion is it? A deflationist cannot say things like "all the theorems of **T** are true", as Shapiro pointed out (or, rather, "all the theorems of **T** can be asserted"). But equipped with *SP*, the deflationist can show in **T'** that any theorem of **T** can (and should) be asserted. This is not the same thing, but just how big is the difference? Tennant (2002, p. 574) claims that it is not much; what the anti-deflationist says about **T** in the metalevel is *shown* by the deflationist, and *SP* states this.

However, there is one problem here. It is different to state *explicitly* that "all the theorems are *x*" from presenting a scheme that *shows* "for any theorem it holds that *x*". The difficulty lies in that *SP* may consist of infinitely many instances, and there might not be an exhaustive way of stating this finitely. To replace the schematic principle of *SP*, there are two possible ways of stating the involved principle explicitly, according to Tennant. One could either:

[69] See Tennant 2002, p. 574 or Löb 1955 for details.

(i) use an explicit truth predicate, and say "All **T**-theorems are true"

or

(ii) use a prosentential device, saying, instead something like "For every sentence that **T** proves, *tthat*."

Here *tthat* is a prosentential extension of English, referring to some new quality that captures the property of "all **T**-theorems being assertable".[70] Essentially, this is a way to remark that our language may not have all the expressive resources that are needed in order to speak about meta-mathematics. All this may sound far-fetched, but that is not necessarily the case. As Tennant (ibid., p. 575) points out, philosophers have been ready to change the fundamental concepts of classical logic. Surely introducing new concepts is not any more radical? [71]

Ultimately, as far as deflationism in considered, the argument may come down to choosing between truth and some other expression *tthat*. At this point the first question we must ask is why we could not use the concept that already exists in our language, instead of inventing artificial ones? If *tthat* is simply a translation of *truth*, the prosentential argument can hardly be any better than the Tarskian one. Indeed, there is a big problem in such prosentential devices behaving *too much* like Tarskian truth. One very problematic matter with Tennant's soundness principle turns out to be the very feature it was introduced for: it performs the same function as Tarskian theory of truth in establishing the assertability of Gödel sentences. While Tennant only looks at the desirable results of his approach, there is also an important undesirable result for the deflationist. As we saw, adding the soundness principle gives us the new assertable sentence *G*(**T**). Even if we

[70] Robert Brandom (1994) has written extensively about such prosentential account of truth.

[71] There is a parallel to this in the famous Tonk-discussion of A.N. Prior and Nuel Belnap, concerning the introduction of new logical connectives. See Prior 1960 and Belnap 1962 for details.

avoid all reference to truth, the soundness expansion causes our *assertability property* to be substantial. The whole point of mathematical deflationism was of course that no expansions to formal systems are needed – but if we do allow expansions, certainly they cannot cause truth/assertability to be substantial. Whatever we want from our assertability property, substantiality would have to be very low on the list. In any case, it cannot be a tool for the deflationist any more.

If Tennant were correct, it would certainly make for an interesting case. But even so, that is no reason for us to abandon our underlying contention of semantics in mathematics. It does not really make a case for the deflationism of Field's type, a fact that Tennant is quick to acknowledge himself. As far as Field is concerned, Tennant's argument must be equally unacceptable, because it uses the consistency of formal systems and the results of mathematical induction in the same way. Indeed, Tennant is not a deflationist, and believes that truth is a substantial property. As he points out (2005, p. 89), his was an effort to show that Shapiro's and Ketland's arguments do not suffice to refute deflationism. But that does not mean that deflationism is correct.

Tennant's argument seems to have an air of translation around it, and that should immediately alarm us. Prosentential devices must be something other than direct translations of the existing concepts in order to be useful. This means that they should at the very least not lead to the exact same properties as the concepts they were meant to replace. But Tennant's argument does *exactly* that. We will see the deep problems of translation-based solutions in philosophy when we deal with various nominalist theories in Chapter 6. Tennant's argument seems to fall into this category. Tennant undoubtedly does give an alternative for a Tarskian account of truth, but only by replacing truth (not directly, it must be noted to Tennant's advantage) with other principles that were not included in the original formal system **T**. But one must question the motivation behind such endeavours. Surely everything would be less complicated and more intuitive by including an account of truth, a concept we are already familiar with.

3.6 Why soundness over truth?

However problematic such prosentential approaches may be, the real burden of Tennant's argument concerns the *reasons* why we should expand formal systems with soundness principles rather than Tarskian truth. To justify a soundness principle, it would obviously have to be a weaker expansion. Ketland (2005) concedes that Tennant's argument is valid, but he contends that assuming the soundness principle is in fact a stronger logical commitment than introducing the notion of truth. Ketland points out that there is a logical difference between accepting each theorem of arithmetic, and accepting the proposition that *all* the theorems of arithmetic are true/assertable. The latter (the soundness principle) is logically stronger than the former, which is the approach we take with Tarskian truth. While both approaches have the equal result of us accepting that "all the theorems of PA are true", with truth we *arrive* at the stronger notion – with the soundness principle we *assume* it. This way Ketland (ibid., p. 75) sees that there is no reason to choose the soundness principle over truth as our choice of extension for **T**.

Hence Ketland (ibid., 76-77) thinks that Tennant has misunderstood his and Shapiro's argument.[72] For him there is nothing strange in that a soundness principle will suffice to carry out the semantical argument (or the reflection argument, as Ketland calls it). It is the *justification* of the soundness principle that is important, and that is what his and Shapiro's arguments are all about. While Tennant adopts his reflection principle without any justification at all, Ketland and Shapiro were trying to show how such principles *follow* – with the help of mathematical induction over formulas containing the truth predicate – from a truth-theoretic expansion to formal systems like **T**. This is an important matter, since Tennant's counterargument is based on semantical

[72] As Ketland points out, Solomon Feferman (1991) had introduced largely the same argument already in 1991. It is only for the sake of convenience that we have concentrated on Shapiro's and Ketland's versions of it; Feferman's argument was not followed by the kind of discussion that the later ones did.

truth not being the *only* way to establish the truth (assertability) of Gödel sentences, which is something that Ketland did not claim it to be.[73] What Ketland wants to emphasize is that the soundness principle (and its variation, Tennant's reflection principle) is available for him and Shapiro because they justify it with a Tarskian theory of truth. It is not available for Tennant because he just assumes it without any argument. Thus Tennant's counterargument fails.

That is an important point, and I will argue that Ketland is ultimately correct. But still, in one sense, Tennant *is* right. Gödel's incompleteness theorems do not immediately imply a substantial notion of truth. There are other alternatives in expanding the formal system, one of which is the soundness principle. In an extremely important way, that changes the question. We no longer need to ask whether we can establish the truth (or assertability) of Gödel sentences. We undoubtedly do, and there are various ways of expanding the formal systems to do this. The question now becomes: which expansion is the most *plausible* one?

I think we can learn a lot about the weakness of Tennant's position from one harmless-looking claim in his article. Tennant (2002, pp. 560-561) points out that our talk of self-reference when it comes to the Gödel sentence is misguided. Thus it is, according to him, incorrect to say that $G(\mathbf{T})$ states its own unprovability. Rather, in Gödel's construction we just have proofs from φ to $\neg \mathrm{Pr}_T(\overline{\varphi})$ and

[73] It must be noted that Ketland also claims that if a Tarskian theory of truth is even *one* way of recognizing the truth of $G(\mathbf{T})$, it is enough to refute deflationism, since deflationist truth was supposed be conservative. If deflationism over mathematical truth were correct, we could not be able to prove $G(\mathbf{T})$ from the truth-theoretic extension of **T**. However, I do not agree with Ketland here. Since there are other ways of establishing the truth of $G(\mathbf{T})$, this comes down to the question whether we are justified in expanding formal systems at all. If we are, then the question becomes which of the expansions is the most plausible one, as Ketland initially argued. This second line of thinking is much less attractive, because it seems to *presuppose* that Tarskian truth is in fact a correct theory of mathematical truth.

from $\neg \mathrm{Pr}_T(\overline{\varphi})$ to φ. On this account φ is only a fixed point for the predicate $\neg \mathrm{Pr}_T(x)$. Of course this much we knew already, and Gödel's fixed-point theorem is instrumental in the proof of the incompleteness theorems. But why could we not talk about the Gödel sentence being self-referential? Surely Gödel formulated it knowing very well that self-referential sentences cause problems. Just because he found the syntax for formulating such a sentence without using imprecise terms such as "self-referential" does not change the fact that Gödel's whole proof was done with such a self-referential sentence in mind. This might seem like an unimportant detail, but it prepares for the sort of argument that Tennant has in mind: any semantical reference is forbidden. But could Gödel ever have formulated the incompleteness theorems if he did not have a semantical understanding of the central concept of self-reference? We should keep in mind that Gödel knew exactly how paradoxes are reached in mathematics. This is particularly important because it mirrors the larger question of mathematical thinking: perhaps the semantical argument can be translated into non-semantical terms, but that should not be understood to imply that semantical thinking does not *exist*, or that the introduction of semantical truth is not correct. We must not forget that Tennant is re-formulating (or perhaps even translating) an existing argument – one which is built on a semantic notion of truth. To reach its goal, Tennant's argument would need to achieve something beyond that: it should be in some way *better* than the semantical arguments of Shapiro and Ketland.

That is an important question, because Tennant's (2005, pp. 91-92) main counter-criticism against Ketland is that the truth-theoretic principles need justification as well. This is something that Ketland seems to overlook, and Tennant believes that justification for the soundness principles is easier to present. To this effect, he points out that from a Tarskian theory of truth we get stronger results (for example, concerning the consistency of the *expanded* system) than from the soundness principle, or from his own uniform primitive recursion reflection principle. Hence Tennant claims that his $UR_{p.r.}$ produces the *weakest* expansion of **T** that is able to prove the Gödel sentence $G(\mathbf{T})$.

This is Tennant's (2005) final point: a Tarskian theory of truth is perfectly acceptable (we must remember that Tennant is not a deflationist himself) as a means of recognizing the truth of the Gödel sentence. However, unlike Shapiro and Ketland think, it is not the only means. And unlike Ketland later claims, Tennant argues that it is not the least strong assumption, either. Tennant points out that he can perfectly well assume the soundness principle (or $UR_{p.r.}$; it matters little in our discussion) without a theory to back it up, because it is a weaker assumption than the one Ketland makes in introducing Tarskian truth. The soundness principle only assumes that we *trust* our arithmetic systems, while Ketland and Shapiro assume them to be *true*. Or so goes Tennant's (ibid., pp. 93-95) argument.

3.7 Conclusions

What can we make out of all this? I think there are two important points to be made here, in addition to the ones that have been mentioned so far. First of all, Tennant seems to identify the strength of an *assumption* with that of its *consequence*. Surely a theory of truth could be a weaker assumption than a soundness principle, even if it did lead to stronger results concerning the expanded system – just like Ketland argued on the difference between accepting PA plus truth and accepting the soundness of PA. As far as the logical strength of the expansions is concerned, there really does not seem to be much of a difference between truth and soundness. In fact, as far as the sentences of PA are concerned, which is the most important aspect in the whole matter, they are equal. However, there is another, much more important sense in which to distinguish between the strengths of assumptions here. Tennant is concerned only with the *logical* strength of an assumption, but we must also consider the *epistemological* strength. Whatever the logical status between the expansions of truth and soundness may be, we must look at other aspects of justification behind them. Surprisingly, this matter is largely ignored in the current discussion, even though we are ultimately concerned with the epistemological and metaphysical status of the truth predicate.

For minimal logical strength, we could assume that each theorem is given to us by direct observation of the physical world. This assumption has absolutely no logical strength, but it is obviously an extremely strong assumption as far as the epistemological aspects of philosophy of mathematics are concerned. Logical strength alone cannot be the deciding factor when we are choosing between expansions.

This brings us to the second – and I think, the most important – point. Just what arguments do we have in favour of the justification for each alternative, the soundness principle and a semantical theory of truth? The former, at least for Tennant, seems to be based only on the contention that it can do the same job that truth does, plus it is logically weaker. But how does it stand metaphysically and epistemologically? Where do we get the conviction that we should assert all provable sentences? Tennant claims that this is just equivalent to stating that we trust our formal systems. This is of course true, but by itself not satisfactory. Obviously we trust our formal systems, but Tennant's soundness principle applies to *all* formal systems, not just the ones we want to accept in mathematics. To be of philosophical value, such a soundness principle would need to have enough expressive power to differentiate between assertable and *non-assertable* formal systems. As it is now, Tennant's approach only moves that question to another level: it cannot explain why we want to assert certain formal systems and not others.

This weakness becomes obvious when compared to Tarskian truth. When we claim with Shapiro that $\mathbf{T'} \vdash \forall x(\mathrm{Pr}(x) \rightarrow Tr(x))$, we are making a statement that the theorems of **T** are *true* in a non-deflationary way: that the axioms of **T** are true, and the rules of proof valid. Importantly, for most formal systems **S** we would *not* be making this claim, since we do not trust all formal systems to give us true mathematical sentences. However, when we claim with Tennant that $prov_T(\overline{\psi}) \rightarrow \psi$, we are in fact stating that in addition to **T**, for *all* formal systems **S** provability implies assertability. How can a soundness principle alone differentiate between an assertable and a non-assertable formal system? The soundness principle is suspiciously arbitrary in its nature: if

Tennant wants to apply it to the formal system **T**, what is there to prevent us from applying it to any formal system **S**? This question can be answered only by going into meta-mathematical considerations over acceptable formal systems – but there we face the same problem all over again.

In fact, how can we have any criteria of what results we want to assert over others, without any reference to truth, or something else beyond formal systems? What reasons do we have for accepting some rules of proof and axioms over others? I think these are important questions that the deflationist must answer, and if not truth, one needs to introduce some other criteria in order to save mathematics from arbitrariness. It is much too easy to follow mathematical practice in claiming arithmetic to be sound, without taking any notice of the fact that axiomatic arithmetic is *designed* to fit our pre-formal conception of what properties natural numbers should fulfil. Obviously we have not simply by accident stumbled on the Peano axiomatization of arithmetic. This is a point I want to make in this work. When we want to make sense of the epistemology and metaphysics of mathematics, we must study the whole phenomenon of mathematical thinking – not just the end product that formal systems are.

Here we must return to the concept of pre-formal mathematical thinking. Just like pre-formal mathematics precedes formal mathematics both historically and psychologically, I will argue that a notion of truth precedes a notion of proof. In fact, I contend that truth is the feature of pre-formal mathematics that formal mathematics is designed to capture by proof. This is why I think Ketland is correct: the notion of truth is a more natural, more plausible assumption to make. This is the case historically, psychologically, pragmatically and epistemologically. No matter which way we look at mathematics, we cannot escape the impression that we are after very special kind of systems and sentences, and not the kind of arbitrary formal systems that the deflationist position leads to. The ball is definitely in the deflationist's court to show how the concept of proof, as we have it in arithmetic, could have developed without any reference to truth, that is, ultimately without reference to any reason for asserting certain theorems over others.

If we accept a Tarskian account of truth, then truth *is* a substantial notion. The Gödel sentence is enough to show that there is an extensional (as well as the intensional) difference between truth and proof. There are other ways of establishing the truth (or rather, assertability) of Gödel sentences, but a semantical notion of truth is the most plausible and natural one. The reason for this is that we already *have* the notion of semantical truth, and it corresponds to our pre-formal mathematical thinking. This is one aspect of the problem that Shapiro, Field, Ketland and Tennant have all largely neglected. But I think it is an important one. In fact, it is one of the main theses of this work, and I will argue for it in the next chapters.

To put the structure of my argument more formally:

(1) In any consistent formal system **T** containing arithmetic there are sentences that can neither be proved nor disproved in **T**. This is Gödel's first incompleteness theorem.

(2) Such Gödel sentences *G*(**T**) can be seen to be true in a system **T′**, where **T′** is an expansion of **T** that provides us with the soundness principle of **T**, that is $prov_T(\overline{\psi}) \rightarrow \psi$.

(3) There are (at least) two possible solutions to include the soundness principle in **T′**. First, a semantical Tarskian theory of truth can be introduced (following Shapiro and Ketland). Second, a form of the soundness principle itself can be directly introduced (following Tennant).

(4) Hence, us being able to establish the truth of *G*(**T**) does not directly imply that the notion of truth is substantial. *G*(**T**) can be asserted without any reference to truth, as Tennant showed by introducing a soundness principle.

(5) However, the introduction of a soundness principle without any reference to truth is problematic. We would not seem to have any reason to assert one sentence over another, and as

such all formal systems would be equally sound. Consequently, the axioms and rules of proof would be ultimately arbitrary. This is the main problem of extreme formalism.

(6) Introducing a semantical notion of truth carries no such burden of arbitrariness, because by doing it we are committing to the truth of assertable sentences.

(7) Therefore, introducing Tarskian truth is a much more natural and plausible extension than the soundness principle. In fact, it is no extension at all when we consider the whole phenomenon of mathematical thinking. Our pre-formal mathematics *already includes* the concept of truth, which the Tarskian account satisfies.

(8) Because with a Tarskian notion of truth we can establish (in **T'**) that *G*(**T**) is true, and it is the most plausible of the competing strategies, truth and proof do not have the same extension, that is, they are not the same concept.

(9) Therefore, the concept of truth is substantial, not deflationary.

The important point about (9) is that it does not follow directly from Gödel's incompleteness theorems. It follows from us already having a substantial notion of truth. Gödel's first theorem is important because it gives us an *explicit* case of seeing this substantiality of truth. Therefore, the existence of our pre-formal mathematical thinking, together with Gödel's incompleteness theorems, makes an extremely strong case for the substantiality of truth.

The debate over semantical arguments is still young and might take interesting new turns. The possible problems and solutions cannot be exhausted by the arguments we have just examined. But at the moment it certainly looks like the semantical argument really *does* give us a reason to think that deflationism either fails, or it has to be changed too much for it to remain appealing. This is

not because we could not assert the Gödel sentence without the notion of truth. Tennant is correct in claiming that a soundness principle will suffice for that purpose. However, his approach includes serious problems. Either one assumes the soundness principle without analysis, or he must present assertability conditions that account for the soundness. In neither case can we refer to anything outside the realm of formal mathematical systems, because if we do, we would be equally justified in expanding the systems with truth. The former case implies arbitrary trust in formal systems, so we must move into assertability conditions as the alternative to truth. We will see many serious problems with this approach.

In the next chapters, I will argue that no expansion to mathematics is needed in order to carry out the semantical argument, as long as we acknowledge that mathematics does not consist only of the formal part. The pre-formal part exists, as well, it is absolutely essential for us to be able to practise mathematics, and it is philosophically important. This is something neither Ketland nor Shapiro have used in their arguments, and which should be acknowledged now. All along, they are concerned with finding the right *expansion* to formal mathematics. I will argue that this makes the substantialist project needlessly difficult. We should start by considering mathematics as a *whole*, not just the important tip of the iceberg that is the formal part. Then we will see that Tarskian truth is in fact no expansion at all: it fits astonishingly well with our pre-formal mathematical thinking. As we will see, it also fits well with just about any philosophical account of mathematics, whether it is Platonism, structuralism, empiricism or moderate formalism. The only one it does not fit together with is extreme formalism, where we categorically reject the possibility of expanding formal systems in the first place. These will be the main subjects of the remainder of this work.

4. Formal and pre-formal mathematics

4.1 Assertability and arbitrariness

As we have seen, if we want expand a formal system with a soundness principle, we must justify this somehow. Tennant's claim was that we accept the soundness principle if we trust our notion of arithmetic proof. Although that initially seems acceptable enough, the matter is not that simple. We have not arrived on the rules of proof by accident, as becomes obvious from the briefest of looks at the history of mathematics. It would be all too easy to assume the soundness principle without any justification. In fact, we could straight away assume the assertability of all undecidable sentences and find an even more simple way of asserting the Gödel sentences.[74] However, the only justification for doing this could be that we *know* them to be true. Similarly, in assuming a soundness principle, we are assuming implicitly that we know our arithmetical systems to be sound. But we can only do this because we know what the systems are and what kind of work has been done to make them possible.

Without this knowledge we would only be dealing with some arbitrary set of axioms and rules of proof, and surely not *all* such systems are assertable. The problem with soundness principles is that they only work if we already assume the formal system in question to be sound. But not all formal systems of mathematics can be equally sound, and without any assertability conditions we cannot have any basis for asserting one such system over another. In this work I call this the problem of *theory choice*: with a satisfactory philosophical account we must be able to explain why we have chosen some mathematical theories over others. I will

[74] Of course we could not actually do this, as there are undecidable sentences of which we have no way of knowing whether they are true or false. The most famous of these is the Continuum hypothesis: the sentence according to which there does not exist a set of numbers with cardinality strictly between that of natural numbers and real numbers can neither be proved nor disproved from the axioms of Zermelo-Fraenkel (ZFC) set theory, as shown by Gödel (1940) and Paul Cohen (1963).

argue that Tarskian truth serves that purpose well.[75] If we hold formal arithmetic to be sound, that is because it has been developed to be such. Unfortunately for the deflationist, however, this process of development is contaminated by references to truth all over, starting from the rules of logic as "truth-preserving", and ending with arithmetic giving us "true" information about the physical world. There is no way a mere soundness principle will do as an expansion: we cannot use a ladder, throw it away at the top and pretend that it never existed. The real question is *not* whether PA, or some other formal mathematical system, is sound. It is whether the deflationist can distinguish between the soundness of PA and that of some arbitrary formal system **S**, and still remain a deflationist.

Of course one popular criterion of theory choice used to be consistency. However, not only are there consistent formal systems that we do not use in mathematics, but more importantly – by Gödel's proof – we cannot prove the consistency of the ones that we *do* use. Paradoxically, consistency is at the same time too weak and too strong a requirement for formal theories to use as the criterion of theory choice. Certainly avoiding inconsistency can be – and is – *one* criterion of theory choice, but by itself it does not suffice.

So we see that there is only one option for the deflationist if he wants to use soundness principles: he must present assertability conditions in order to escape the problem of arbitrariness looming in his argument. He must present us with reasons *why* we use the

---

[75] This is of course not to suggest that Tarskian truth by itself is sufficient for solving the problem of theory choice. Rather, Tarskian truth is *compatible* with having a solution to the problem, while a deflationist theory of truth is not. When we add Tarskian truth to a formal system **S**, we claim that the theorems of **S** are true, but we are obviously only prepared to do this for the formal systems that we hold to be true. Thus the criterion of theory choice is something external to Tarskian truth, which is hardly surprising. Tarski's theory of truth is a formal philosophical tool and should not be thought to help us in establishing the true axioms of mathematics. But it *does* give us the formal tools to handle the problem of theory choice, which is something that – I will argue in this chapter – the extreme formalist account of mathematics fails to do.

rules of proof we do, and why we have chosen the axioms we have. Importantly, he must do all this without any reference to truth – that is, ultimately, without any reference to objects outside formal mathematics. We must remember that truth is a much wider concept than the Platonist interpretation of it suggests. The problem starts with the very basic rules of logic. If we have $a \rightarrow b$ and a, why do we infer b instead of $\neg b$? It seems like a hopeless task to answer this question without *any* reference to notions outside formal mathematics, unless one accepts the option of pure arbitrariness.

Crispin Wright (1992, pp. 95-107) in his neo-logicist program has studied the possibility, as well as the problems, of assertability conditions. Just why are we justified in concluding b from $a \rightarrow b$ and a? One idea suggested by Wright and others is that the rules of proof are such deep-entrenched standards in the mathematical practice that they need no outer justification. In other words, we are warranted to use them because of the very fact that they *are* rules of proof. In Wright's terminology, such deep-entrenched statements are *superassertible*; they would "survive arbitrarily close scrutiny" (ibid., p. 48, Wright 1987, pp. 295-302). I think that there are two ways of interpreting this somewhat frustratingly vague characterization: either it can be thought of as conventionalism, or it can be interpreted as leaving the whole question concerning the origins of rules of proof and axioms open. Both interpretations have their problems. The former is obviously quite possible to do, but if we are to use conventions as the criteria for assertability, why should we neglect the convention of talking about the *truth* of mathematical sentences, and replace it with the non-convention of assertability? Just *what* conventions should we accept, and more importantly, how did we arrive at these conventions? Conventionalism has the troubling potential of becoming a philosophical explanation of everything. All current human knowledge can be thought of as conventions, which it undoubtedly is in a weak sense. It is a convention to think that the general theory of relativity explains the universe better than Newton's mechanics does. However, outside post-modernist circles, one is not likely to find philosophers who that think this is *only* a convention, not based on anything objective.

It goes without saying that many mathematical statements are deep-entrenched conventions. However, even the most fundamental of our conventions have origins, and the real question concerns those origins of the conventions. What we know about the origins of our rules of proof is that they were intended to find out true mathematical sentences. From the Pythagorean times to most working mathematicians today, mathematics has been widely understood as a distinct pursuit of true sentences, aimed of course at understanding mathematics itself, but also at explaining, and predicting, the physical world in mathematical terms. The success of this venture has given the mathematical rules of proof justification, and it has made perfect sense to call the accepted theories true. When we dispense with truth, we also seem to dispense with that justification. If not aimed at finding out truths, can mathematical conventions be thought of as anything else than arbitrary statements that we have somehow chosen to believe about things like numbers and geometrical objects – in essence no different from such non-mathematical endeavours as numerology that are seemingly about the same objects? Today, almost all scientific theories rely heavily on mathematical tools. Could it really be the case that all these tools – so incredibly helpful, perhaps indispensable, in explaining the physical world – are in fact arbitrary, products of pure chance?

How about the second interpretation: that we need to leave the questions of the origins untouched. Again, this is possible to do, although it basically leads to the demolition of philosophy of mathematics. But again, why dispense with truth? That is already a deep-entrenched mathematical practice, while assertability is not. This will be the main problem with any argument similar to Wright's: if we use the accepted status of rules of proof as an argument for their justification, how can we neglect the accepted status of truth as the grounds for these rules? Any minimalist argument drawing from conventions must be careful with this. Minimalism is *not* a convention of mathematical practice. In fact *Platonism* comes closest to being the philosophical convention of mathematicians, at least as far as the working language of mathematics in considered.

Accepting assertability without any analysis does not look like a fruitful approach, and hence the assertability conditions must be somehow justified. Perhaps the first effort to find assertability conditions was Arend Heyting's (1931) semantic of proof conditions. Michael Dummett (1977 and 1978) has also taken on such a project. One big problem with both of these accounts is that they are intuitionist assertability conditions, and not directly applicable to the classical mathematics considered so far in this work. That fact notwithstanding, neither of these accounts manages to avoid the problem of the origins of these assertability conditions. Intuition is the key concept in both of them, and even if one were ready to follow their account of mathematical intuition, there would still remain the question what causes our mathematical intuitions to be what they are. They simply do not give us any explanation *why* some sentences are assertable but others are not. That explanation would of course have to be completely independent of truth, *including* all the conventions that are based on the concept of truth. That last remark is very important. Otherwise, we could just take the easy way out: translate the word *true* as the word *assertable*. Quite obviously we cannot be after such translations in a deflationist pursuit of mathematics.

Keeping this in mind, it seems that the challenge for the deflationist is nothing less than creating a complete account of mathematics from the scratch. Perhaps this sounds too hard, but for the initial purposes he only needs to justify the axioms and rules of proof of basic mathematics – say, set theory or arithmetic. And if that sounds like too arduous a task, it is only because that is the price any such radical revisionist must pay. As of now, I have not stumbled upon a comprehensive account of assertability conditions that would be anywhere close to being an adequate replacement for a theory of truth. Indeed, most accounts defending assertability seem to run into the translation problem: they add nothing to the subject expect a new terminology dispensing with truth by translating it as assertability. But how we could ever hope to solve philosophical (in this case both epistemological and metaphysical) problems by translation?

What about usefulness and conservativity in physical theories as the criteria of assertability, which are the ideas behind Field's formalism? At first this approach has a certain appeal to it since it seemingly avoids the problem of arbitrariness: instead of any purely mathematical criteria, it is now a *physical* criterion that our theory choice should follow. As long as mathematical theories are useful in physical theories, we should accept them. In Field's (1980, p. 15) words:

> What makes the mathematical theories we accept better than these alternatives to them is not that they are true […] but rather that they are more *useful*: they are more of an aid to us in drawing consequences from those nominalistic theories that we are interested in. If the world were different, we would be interested in different nominalistic theories, and in that case some of the alternatives to some of our favourite mathematical theories might be of more use than the theories we now accept. Thus mathematics is in a sense empirical, but only in the rather Pickwickian sense that it is an empirical question as to which mathematical theory is useful. (Italics in the original.)

Clearly not all mathematical theories are equally useful in the accepted physical ones, so arbitrariness is avoided. Or is it? When we take a closer look at the matter, it becomes obvious that this is all just an illusion. Take Field's example of Newtonian mechanics. We are talking about a physical theory that actually *required* the development of calculus. Now let us think of the usefulness and conservativeness of calculus in Newtonian mechanics. Field (1980, 1989) shows that calculus is indeed conservative over Newtonian mechanics[76], ends up with deflationism and fictionalism in mathematics, and finally admits that calculus should be preferred because it is useful. But what he really has done is showing that a physical theory that required calculus can be *afterwards* presented in a calculus-free notation. In this way, of course it is calculus – and not some rival theory – that Field shows to be conservative over Newtonian mechanics. That calculus is useful in it is even more obvious. Does this tell us that we have a criterion for theory choice

[76] Or claims to have shown. See Chapter 6.4 for problems with Field's approach.

in mathematics? Quite clearly not, because with a different mathematical theory as the foundation, Newton (or someone else) would have ended up with a different theory of physics.

So the problem of arbitrariness is not avoided. How can we know beforehand which theories of mathematics to use in physics, not to mention which theories to use in all the simple non-scientific applications?[77] Clearly conservativeness and usefulness cannot be the answers here. Whether we manage to show our current mathematical theories to be conservative over physics or not, there remains the problem of accepting certain theories as the ones to use in physics in the first place. In a way, the question is moved from the theory choice between mathematical theories into choice between *physical* ones. Since modern physics is so thoroughly "contaminated" by mathematics, any study of the existing theories is redundant. If mathematical theories are conservative over physics in the first place, they are bound to be conservative over the physical theories they were used in, even *developed for*, as in Field's own paradigm case. That they are useful in such theories is too obvious to have relevance on the question of theory choice.

Some sort of counterfactual inference must be made here. What if our theories of physics had been developed differently, in a way that conflicts with calculus, or perhaps totally without mathematics? Clearly we would have no justification for using conservativeness as an argument for accepting calculus. But as it happens, we do accept calculus. However this is explained, it cannot be the case that conservativeness is the criterion of theory choice here. *All* mathematical theories are conservative over some theory of physics, although obviously most of them are that in a totally uninteresting fashion. If our best current physical theories had been developed in a non-mathematical way, Field could have a point. But conservativeness over thoroughly mathematical

[77] In the Quinean-Fieldian tradition of discourse one almost gets the idea that mathematics is used solely in developed scientific theories. It must be remembered, however, that we are using mathematics even when counting apples – and that has nothing do with being conservative over theories of physics. We will return to this question of "non-scientific" applications later on in this work.

theories of physics cannot work as the criterion of theory choice in mathematics.[78]

Either we must accept arbitrariness or we must make more out of Field's notion of usefulness (and perhaps conservativeness) being, in a weak sense, an empirical question. However, in that case mathematics would be an essentially empirical science, and it would make sense to speak of references for mathematical sentences. Clearly these would not be the kind of references that we usually associate with mathematics, but it would make perfect sense to ask whether such useful mathematical theories are in fact *true* in the same manner as the theories of physics are. In Chapter 6 we examine such problems more closely – let it suffice now to note that there is a strong air of contradiction in first claiming mathematics to be a fiction, and then presenting objective empirical criteria of theory choice for it. Usefulness and conservativeness over physical theories certainly seem to fall under that category. It does not suffice philosophically to say that some theories of mathematics are more useful than others, while maintaining that there is nothing that mathematics refers to – that is, nothing to *cause* this usefulness. Essentially, this is saying that we have criteria of theory choice, but they are not based on anything. It is comparable to simply saying that from two competing mathematical theories we use the "better" one. Why do we do this? Because "goodness" is the quality that decides the theory choice. The real question is, of course, *why* certain theories are better – or more useful – than their alternatives, and why could we not call this reason *truth*? Objective empirical criteria together with fictionalism are hardly the answer, and arbitrariness is even less so.

---

[78] As Shapiro (2000a, footnote on p. 232) points out, many mathematicians hold that while Field's program is *Science Without Numbers*, it is not science without *mathematics*. In this way, Field's program is thought to mirror Newton's mathematical structure of space and time, which makes it thoroughly mathematical. Reading through Field's book, this impression is hard to avoid.

4.2 Undecidable sentences and formalism

From all the considerations in Chapter 3, we should know enough about the semantical arguments and their position in the philosophy of mathematics. For the strict formalist, all the possible ways out seem to point in one direction: arbitrariness. But there is something ominous in that even for us non-formalists: after all, mathematicians actively pursue and study formal systems. Clearly we are all concerned with formal systems in mathematics, whatever our philosophical leanings may be. When we argue against extreme formalism, we must be careful not to affect the more moderate modes of formalism, which are indispensable for mathematics. That is why we need to take another look at formalism and the Gödel sentences.

To avoid a common lay misunderstanding, it must be emphasized that the Gödel sentence $G(\mathbf{T})$ is undecidable in the formal system **T**, but that does not imply that there are *absolutely* undecidable sentences: sentences undecidable in *all* formal systems.[79] With a different axiomatization and an alternative Gödel-numbering we could prove $G(\mathbf{T})$, while another sentence $G(\mathbf{T})'$ would be undecidable. That is the nature of Gödel's first incompleteness theorem. It reveals the incompleteness of single formal systems, but not that of all formal systems. Or does it? If we are committed to formalism in a strict sense, should we not campaign for a *single* formal system containing all of mathematics? That is most likely what Field (1998) means when he talks about "our fullest mathematical theory", and it seems to correspond to the original Hilbertian goal of providing a completely formal basis for mathematics. In that kind of single-system formalism there would indeed be fundamentally undecidable sentences, provided that the formal system is consistent. In addition, if we could establish the consistency of the system, we would know these

[79] Or to put it another way, undecidable independent of the formalization chosen. Let it be noted that all along we will be concerned with formal systems to which Gödel's incompleteness theorems apply, that is, consistent formal systems containing arithmetic.

sentences to be true.[80] Basically, we would have arrived at absolutely true arithmetical sentences that could never be proved. This is a much stronger result than the semantical arguments give us. In them, we were only considering single formal systems, not *all* of formal mathematics.

It should be noted that the question of absolute undecidability is treated here in a manner that already presupposes that we have agreed on the meaning of the concept. There are at least two important questions concerning this. First, as Gödel (1946) notes, there is a difference between decidability in the formalist sense and in terms of mathematical *definability*. While all formal systems are incomplete, and as such contain undecidable sentences, Gödel suggested the possibility that a concept of (transfinite) definability may enable us to show that every theorem expressible in, say, set theory is also decidable, thus achieving completeness in another way. Second, what do we mean by "absolute" here? Gödel's suggestion above is a set-theoretical one taking ordinal numbers as the primitive terms and, as noted by Gödel himself, in that sense not absolute. Can we ever hope to agree on the concept of absoluteness when we disagree on so much else concerning the foundations of mathematics? To avoid getting the original question muddled up in these kinds of problems, important though they are, I want to stick to the formalist concept of finitist undecidability here. The reason for that is simple: I believe that it is closest to what Field interprets mathematics to be, and it is Field and extreme formalism that I am arguing against. We can imagine our "fullest mathematical theory" as a single formal system, and provided that it is consistent, conclude that there are undecidable Gödel sentences in it. For now, we are concerned with absolutely undecidable sentences in this sense.

What could these sentences be? We know that from the axioms of Zermelo-Fraenkel set theory one can prove neither the

[80] It is hard to figure out how exactly this could be done, but that should not be considered a problem. If such a full mathematical theory did exist, it would most likely be at least generally accepted to be consistent.

Continuum Hypothesis nor its negation.[81] Are the undecidable sentences of arithmetic of this nature, or perhaps the type of Goldbach's conjecture? The short answer seems to be: we have no idea. Here we must remember the double role that the natural numbers play in Gödel's proof. They are used as codes for formulas, but at the same time they are also natural numbers in the usual arithmetical sense. The undecidable Gödel sentence is still also a sentence of PA, a statement concerning natural numbers. Hence, there is a property concerning natural numbers that we can neither prove nor disprove in PA.

If we consider the formalist ideal of having one "fullest" formal mathematical theory, this would mean that, as well as there being undecidable sentences in the theory, there would also be the corresponding sentences of arithmetic that would neither be true nor false. Natural numbers, those among the simplest and most primary of our mathematical concepts, would hold properties that are *fundamentally* undecidable. This is a startling prospect, and one could make at least three possible conclusions out of it. First, it is possible to infer that while mathematics is about formal systems, such *all-purpose* formal systems are not what we are after. Second, one can claim that our conception of formal systems in general is flawed. Third, one can accept that our mathematical knowledge will always have these particular gaps.

The third option can be ruled out right away. With different axiomatizations and different ways of carrying out the Gödel-numbering we end up with different undecidable sentences. Of course we are far from actually having any candidates for such

[81]This has been used by Putnam (1980) as an argument against objectivity of mathematical truths. Against that, it should suffice to point out here that objectivity should not be confused with the completeness of formal systems: we can have many ways of formulating undecidable sentences, and some of them might even be absolutely undecidable. Even if that were the case with the Continuum hypothesis, it would not imply that *all* mathematical sentences are similar in having no objective truth-value. Even Field (2001, p. 319) agrees that arithmetic truths can be thought to have objectivity.

fundamental formal systems, but one can imagine that a choice between the competing systems would not be unanimous. Such debate would then concern which sentences of arithmetic we want to render *absolutely undecidable*. The absurdity of such a situation should speak for itself.

The second conclusion is a fact, but we can make many different things out of it. Formal systems *are* flawed – Gödel proved that – but do we have anything better? Should we abandon using formal systems in mathematics? Obviously not: axiomatic formal systems are in a way the ultimate achievement of mathematics, and just about everything we do in mathematics strives for such maximally unambiguous mathematical formalism. The fact that formal systems are incomplete is unfortunate for the formalist in the *philosophy* of mathematics, but for the working mathematicians it means very little. We know our axiomatizations of arithmetic to be incomplete, yet mathematicians do not give this a minute of thought when they are proving theorems of arithmetic. We cannot prove that our formal systems of arithmetic are consistent, but we see no problem in using contradictions as means of refuting assumptions. However, there is very little controversial in this: for all practical intents and purposes, arithmetic *is* complete and consistent.[82] The possibility of running by chance into a self-referential sentence like $G(\mathbf{T})$ is negligible – which also works as a reminder of just how special a case the Gödel sentences are. If $G(\mathbf{T})$ had turned out to be provable, and PA complete and consistent, nothing would have turned out differently in the arithmetical practice.

Thus we are left with the first option. Formal systems should be used in mathematics as before, but we should not even make an effort to present the whole of mathematics as a *single* formal

[82] Of course Gödel's theorems do not state that PA is inconsistent, only that if consistent, we cannot establish this consistency within PA. PA could very well be consistent, and based on many considerations it indeed is. Gentzen's (1935) proof is the most famous of these. But if consistent, PA is incomplete – that much Gödel proved.

system.[83] Of course this is very much the status quo in mathematical practice. The logicist program of Frege (1884), Whitehead and Russell (1956) has been largely abandoned, and a uniform all-encompassing formalization of mathematics is something of a pipe dream. That is the situation within working mathematics: different fields of mathematics are more and more isolated from each other. But what should philosophers of mathematics make out of it? I think there is a lot to learn. Different areas of mathematics develop independently, although from a common background of accepted inferences, methods and notations. If our actual fullest mathematical theory were presented as a single work, it would certainly not be a single axiomatization. It would be a collection of axiomatizations, pieced together with a variety of informal meta-level explanations. In order for people to be able to *understand* this collection, the parts concerning the axiomatizations would also contain an abundance of informal clarifications and explanations. When we are talking about "our fullest mathematical theory", we are talking about such a work – and we know this because that is how presenting mathematical theories is *actually* done.[84]

Formal systems themselves are not problematic, and they survive the threat of Gödel's incompleteness theorems. But we must realize the proper place of formal systems in the mathematical practice. Extreme formalists must end up endorsing single-system formalism, because they dismiss the role of all

---

[83] Haskell Curry (1954, p. 204) has argued for this viewpoint on a formalist basis. In his "empirical formalist" account we can consider mathematics to consist of several (incomplete and possibly contradictory) formal systems. Furthermore, mathematics can include metatheoretic propositions. This could of course (although Curry does not mention it) mean a Tarskian expansion, among other things. Compared to extreme formalism, it might seem rather misleading from Curry to call his position formalism at all. But we should note that formalism expanded even slightly makes it far more appealing than the extreme position.

[84] One could use any mathematical textbook as a reference. The Bourbaki-group has perhaps come closest to giving a completely formal presentation of mathematics, yet even their work is nothing like pure formalism. See Bourbaki 1970 for an example.

informal elements in mathematical thinking. Only at that point do we arrive at problems. If we accept extreme formalism, we suddenly expand the Gödelian result of "undecidable in **T**" into "*absolutely* undecidable". Given the semantical arguments, and supposing that such an all-encompassing formal system would be consistent, this is a dangerously big success for the substantialist of truth. From a substantial notion of truth we move on to truths that could never be proved. That is a consequence we can hardly accept. For a non-revisionist substantialist, formal proof is still the way we establish true sentences of mathematics. Only the axioms and rules of proof are held to be true without proof (because *something* must be) – we cannot easily accept finding out other arithmetical truths that could never be proved. That would not be mathematical anymore, and it would contradict with the moderate versions of formalism. Single-system formalism not only destroys extreme formalism – it destroys sensible substantialism, as well.

Clearly it is not such a single-system formalism that we want. Nor should it be just formalism we want, single-system or not. Mathematics as a human phenomenon includes the formal systems, but it includes an informal side, as well – although this fact is often ignored in the philosophy of mathematics. Probably as a result of Hilbert's program and its esteemed formalist ideals, the philosophers of mathematics seem reluctant to recognize the role of informal thinking and presentation in mathematics.[85] However, I claim that it is of great importance when we try to understand mathematics as a phenomenon beyond the formal systems.

4.3 Tarskian truth and mathematics

In a way, we have been putting the cart before the horse by concentrating on formalism and its problems. All along we have

[85] It should be stressed once more that it was not Hilbert's ideal to remove the informal elements from mathematics. As can be seen from Hilbert 1925, his theory of formal proof was intended to be a clarification of the mathematical practice – which is something very different from extreme formalism.

been talking about truth in mathematics without having any philosophical theory of truth to base it on. Indeed, so far we have only specified that we are concerned with a Tarskian semantical notion of truth. Tarskian truth, however, is merely a condition of material adequacy: it gives us a condition that the definitions of truth must fulfil – it does not tell us what kind of property truth *is*.[86] Tarskian truth is a very allowing concept philosophically, and while it is usually thought that Tarski's is a *correspondence* theory of truth, the T-scheme can be used in many other ways, as well. It is important here to specify just what we mean by mathematical truth, and its relation to formal proof.

I said we have put the cart before the horse, but this actually seems to be the correct order when it comes to *mathematical* truth. By an implicit – and after Hilbert, explicit – agreement, formal proof *is* the way we find out mathematical truths, apart from those of the axioms and rules of proof. Whatever we mean by mathematical truth, there is an almost universal agreement on the methods of establishing true mathematical sentences. Mathematicians state that "Fermat's Last Theorem is true", and although there remains a lot to explain about that statement philosophically, nothing questionable is seen in the correctness of the statement itself – at least not in the non-revisionist circles. Most of us agree that mathematicians establish which sentences are true, and the philosophical question of mathematical truth is somehow distinct from this; aimed at explaining truth, not at deciding true sentences. For the most part, then, we are happy to agree on the set of true mathematical sentences, even though we may not agree on what truth actually is. This might seem paradoxical, but I do not think that it amounts to anything more than having a healthy respect toward our subject matter: the study of mathematics. It is only when the mathematical formalism becomes insufficient that

[86] A comprehensive general discussion on Tarski's definition of truth goes beyond the scope of this work. One of the most famous critics of the use of the T-scheme in truth is Putnam (1975). Jan Woleński (2001) has defended Tarskian truth against Putnam convincingly. See also Kirkham 1992, pp. 141-210 for a good introduction to the general truth-theoretic discussion concerning Tarski.

we need to concentrate on the more fundamental philosophical questions of truth – like those of reference and objective truth-values. But as we have seen, formalism is indeed in trouble, and we need to elaborate on our theory of truth in order to save mathematics from arbitrariness.

My contention is that we cannot approach mathematical truth simply as one case of truth in general. Truth as far as observational activity is concerned is bound to have different characteristics from the truth of mathematics and other deductive disciplines. One problem with deflationist theories of truth like that of Horwich (1998), and also in part Field's, is that they are more general accounts of truth that are only later applied to mathematics. But mathematical truth must get a study of its own, for reasons that should be obvious by the end of this work. The way one forms and justifies beliefs about physics is fundamentally different from the way we acquire mathematical ones. Mathematical theorems are not tested and corroborated empirically in the same way as the ones in other sciences are. All in all, mathematics both in its subject matter and as a human endeavour has many characteristics different from the empirical sciences. To respect that, we need an independent analysis of mathematical truth – just like we need an independent epistemology and ontology of mathematics.[87] This is not to claim that mathematical truth *is* indeed ultimately philosophically different from physical truth; it only means that we cannot *approach* the two subjects in a similar manner.

Coherence, for example, has been used as a criterion and even as a definition of general truth. When we consider mathematics, coherence means consistency, and as such it is so deeply entrenched into mathematics that we could not gain any new insight from introducing it as a criterion of mathematical truth, particularly in the context of this work. We know Hilbert's Consistency program, and remember that Gödel's incompleteness theorems only apply to consistent formal systems. We also

[87] Of course the main opponent of this point of view is Quine (1990, for example). In Quine's philosophy mathematical facts are simply one class of scientific facts, and they have the same ontological and epistemological status.

remember that we cannot establish the consistency of formal mathematical systems containing arithmetic. Considering all this, it seems that there is very little use in including consistency as a definition of truth, or a criterion of justification. As a definition it is impossible: knowing what we do about the consistency of formal systems, we could not establish *any* true formal theories of mathematics containing arithmetic. As a criterion of justification, it is redundant. Coherence would be an unwanted concept as a definition or a criterion of truth, but it would be that especially for the *deflationist*. Truth as proof is not the same as truth as consistency, although in the tradition following Cantor (1883) and Hilbert (see footnote 27 of this work) is often seen as such. The latter is obviously an impossible concept in formal systems containing arithmetic, while the first one is not.[88]

Pragmatic theories of truth seem to be summarized in Field's (1980, p. 15) contention that usefulness in physical theories should be the criterion for accepting and rejecting theories of mathematics. Although Field in his formulation did not see any need for the notion of truth in the first place, we can no doubt look at pragmatic truth as comparable to usefulness as a criterion of theory choice. In Chapter 4.1 we examined some of the problems of this approach, but there also other problems that follow from pragmatic truth. In turns, it can be either too strong, too weak or too obscure a requirement. Not all fields of mathematics have immediately apparent practical applications, yet we'd wish to retain the ability to speak about truth in them. In addition, for practical matters there seems to be little difference between *proving* something and showing it to be the case for up to large numbers. Goldbach's conjecture, as we remember, has been shown to hold for all natural numbers $n \leq 10^{18}$. Pragmatically, this can be considered to be

[88] It has to be remembered that consistency *is* a main feature when we talk about mathematical truth. We are concerned with consistent systems, and certainly wish that our mathematical theories were consistent. In practice, we implicitly assume that they in fact *are* consistent. But this should not be confused with consistency being suitable as a *definition* of mathematical truth.

almost as good as a proof for every *n*. Yet mathematically, it is no better than showing it to hold for every $n \leq 10$. Pragmatic considerations can no doubt be fruitful in areas concerning the applications of mathematics, but they seem to have very little potential when it comes to mathematical *truth* – for that they simply seem too un-*mathematical*.

All in all, from the competing general theories of truth, correspondence is the one theory we should concentrate on when it comes to mathematics. Correspondence can refer to many things, but at a first glance it admittedly seems to have an ominously Platonist flavour to it. The most natural way to understand correspondence is that true theorems of formal mathematics correspond to the state of affairs in some objective reality of mathematics, and that reality is easily interpreted to be Platonist. However, this does not need to be the case. If we use the starting point that mathematical truth is different from general truth, we see that correspondence can be understood as a more varied concept.

Tarski (1936, p. 153) thought that his semantic conception of truth gave an adequate account of truth for the correspondence theory. In a correspondence theory of truth a sentence is true if and only if it corresponds to reality. In Tarski's T-scheme this means the sentence of metalanguage corresponds to reality. Let us consider the classic example from before:

> "Snow is white" is true if and only if *snow is white*.

Now in the correspondence-interpretation of Tarski, the T-scheme takes the form:

> "Snow is white" is true if and only if *snow is white* corresponds to reality.

Of course Tarski's T-scheme is only a condition of material adequacy for truth, that is, all the true sentences of the object-language take the form of a T-instance in the metalanguage. It does not tell us anything about the nature of the correspondence, or how we can find out true sentences and justify believing in their

truth. In fact, it is far from being agreed upon in the truth-theoretic philosophy that Tarski's conception of truth is indeed the conception of correspondence theory of truth.[89] This is due to the T-scheme only giving the logical form of true sentences, that is, Tarski's being seemingly only a *logical* (or quasi-logical) theory of truth. From atomic true sentences we by induction over the structures of the formulas arrive at complex true sentences, but nowhere in the T-scheme do we learn anything about the truth conditions of the atomic sentences. Some philosophers consider this a weakness of Tarskian truth – and it is a common basis for discounting Tarskian truth as correspondence, since a description of correspondence would need something beyond logical constants.

That certainly may seem to be the case when we think of truth in general as a *language-to-world* correspondence: Tarskian truth itself does not impose truth conditions on the atomic sentences. However, when we consider mathematical truth, there are some other points that we must remember. It should first be noted that the logical nature of Tarskian truth can also be considered to be a strength: it gives us a framework in which to *establish* truth conditions and other problematic non-logical concepts. Obviously we would be surprised to learn that the truth conditions for mathematical sentences are the same as the ones for sentences of physics, yet Tarskian truth can be applied to both. This way the T-scheme can be a great tool for any philosopher concerned with truth – at least when we do not claim it to be more than a tool.

However, is it really the case that the T-scheme is totally empty of truth-conditions? It is clear that for a sentence like "Oslo is the capital of Norway" this is indeed so. Aside from the semantical content of the words in the sentence, there is obviously something that they correspond to that establishes the truth of the sentence. But in the special case of mathematical truth, it must be remembered that Tarski's is a truth definition also suitable for a *language-to-language* correspondence, and in particular a "formal system-to-formal system" (*formalist*) correspondence. In a formal system **T** the theoremhood of a sentence is naturally decided by

[89] See Kirkham 1992, pp. 170-173.

the axioms and rules of proof of **T**. When it comes to truth, this is established by the soundness principle $\forall x(\mathrm{Pr}(x) \rightarrow Tr(x))$, and it is given in the meta-system **T′** which, as we know, can be just **T** expanded with Tarskian truth. But obviously this *is* giving a truth condition for the sentences of **T**. In the system **T** theoremhood simply means being a theorem of **T**, while in the meta-system **T′** it is at the same time the truth condition. Hence we establish truth conditions for the sentences of **T** already by introducing Tarskian truth, and nothing more. In moving from PA to "PA + Tarskian truth" we claim that all the theorems of PA are *true*. As well as giving an adequacy condition, adding Tarskian truth to a formal system of mathematics at the same time clearly gives us truth conditions.[90]

That goes for the correspondence between formal systems, which is of course not what is usually meant by correspondence. But when it comes to formal systems of mathematics, Tarskian truth seems to be stronger than we thought: it is not free of establishing truth conditions for atomic sentences. For a correspondence theory this could be problematic, but it should not strike us as anything new. If we do not accept the Quinean (1951) rejection of the analytic-synthetic distinction, we immediately get a vast class of analytic sentences that are true without correspondence. As Jaakko Hintikka (2001, p. 6) has reminded us, we need to make a distinction between what is logically true, and what it is true due to the non-logical (and non-linguistic) facts of the world. This naturally also explains us why the T-scheme works as a truth condition in the formalist correspondence: the sentences of a formal system are true in the expanded system due to the very fact of them *being* sentences of that formal system.

If we take mathematics to consist *only* of formal systems, as the strict formalist does, we must commit to Tarskian truth also giving us the truth-conditions for sentences of mathematics. That is why we do not need any strong form of correspondence to go through the semantical argument. In fact, Tarskian truth added to formal systems of mathematics is enough to carry it out, just like Shapiro

[90] Here we must remember that with Tarskian truth we are concerned with interpreted languages.

and Ketland have argued. However, in that case we *do* need an endless hierarchy of formal systems, and that is something we should be wary of.[91] As we saw, with the mere addition of Tarskian truth, there is nothing to tie truth into anything outside formal systems, since the truth conditions are already given by the truth predicate. In this line of thinking one really must question the purpose of introducing truth predicates. After all, if we stick to the formalist correspondence, soundness principles will do the job as well as truth, as we know from Tennant's argument. If all correspondences are of the formalist type, the Tarskian hierarchy is quite clearly formalist itself. In this way, the semantical argument only has the dubious achievement of showing that, for the formalist, Tarskian truth requires an infinite hierarchy of formal systems and languages – which is something that we already knew from Tarski's undefinability result.

However, mathematical truth is not limited to the formalist correspondence. It cannot be – that implies extreme formalism, which in turn implies, I argue, that mathematics is arbitrary. If we hold all correspondences of Tarskian truth to be of the formalist type, then Tarskian truth is *just as arbitrary* as Tennant's soundness principle. If we hold one formal system to be true, what is there to prevent us from holding that *all* formal systems are true? Clearly at some point the Tarskian hierarchy of formal systems must be collapsed, and tied into (partly) non-formal elements – otherwise the formal systems remain arbitrary, as well as part of an infinite hierarchy.

Fortunately this is no problem, because Tarskian truth can be used in mathematics in a way that avoids arbitrariness. Indeed,

---

[91] Remembering Tarski's undefinability result for truth in formal (classical) languages, to give a truth definition for a system **T**, we must expand into a system **T′**. But to give a truth definition for **T′**, we must expand into a system **T″**. This is called the Tarskian hierarchy, and it will go on *ad infinitum* in case we commit ourselves solely to formal languages. By introducing a (partly) informal metalanguage we can restrict the hierarchy of languages down to two levels: a formal first-order object language and its partly informal metalanguage. More about this will follow.

when we look at mathematics, it is instantly obvious that not all formal systems are claimed to be true. It is only a carefully selected group that we study in mathematics – and we focus on them because we think they are *true*, whatever we mean by truth. By formulating and accepting the axioms and rules of proof we commit to that. This way, just like Oslo being the capital of Norway, the theoremhood of a sentence is actually a non-logical fact depending on our choice of axioms and rules of proof. For a formalist correspondence the mere addition of Tarskian truth suffices to give truth conditions, but the choice of axioms and rules of proof goes *beyond* the formalist correspondence. As I have argued earlier, we cannot think of formal mathematical systems existing independently. Hence also in mathematics Tarskian truth can still be used in the general manner: we can think of mathematical theorems being true or false due to non-logical conditions, and Tarskian truth as giving the adequacy condition for this.

The formalist correspondence cannot be all there is to mathematics, and we must include two other kinds of correspondence in the account of mathematical truth. The first one is the correspondence from formal to *pre-formal* languages. The second one is the *language-to-world* type of correspondence. The latter is of course by far more problematic philosophically. We would need to answer all the most difficult questions in the philosophy of mathematics in order to clarify that correspondence. Fortunately, for the purposes of this work, we only need to establish that there *exists* such a correspondence, however weak it may be. As long as that is the case, we can use the first type of correspondence for the arguments concerning truth and proof. This is not at all problematic – in fact, the opposite viewpoint will lead to extreme formalism with its arbitrariness. This will be the subject of Chapter 6.

As it turns out, the first type of correspondence – that between formal and pre-formal languages – is highly important philosophically, especially when it comes to the question of truth. The main problem of correspondence in mathematics is of course the seemingly Platonist flavour of it. However, with the help of pre-formal languages we can rid both Tarskian truth *and*

correspondence from such drastic ontological commitments. In Chapter 2.8 we tentatively formulated our intuition of "seeing" the truth of Gödel sentences as the following T-instance:

G(**T**) is true if and only if *G cannot be proved in* ***T***.

Since *G cannot be proved in* ***T*** is indeed the case, we concluded (tentatively) that *G*(**T**) is true. After that we have reached unambiguous ways of stating the matter, but all the time we have been talking about the same thing, the semantic content of *G*(**T**): that *G cannot be proved in* ***T***. That is obviously not a sentence of **T**, or even of the expanded system **T'** – it is a sentence of our pre-formal language of mathematics. Ultimately, even with the explicit semantical arguments, that pre-formal idea of truth of *G*(**T**) was what we were after. Gödel (1931, p. 151) noted it immediately in his proof, also already noting that we establish the truth in a metasystem.

Thus, the semantical arguments have all along been a way of formulating our pre-formal idea concerning the truth of Gödel sentences. In the next chapters I will claim that most of mathematics is ultimately about the very same thing: finding formal presentations for the pre-formal ideas. It is the correspondence between formal and pre-formal languages that gives us the basis for the semantical arguments – but that is also the correspondence that mathematics as a human endeavour rests on. When it comes to the question of truth, it is this correspondence between the formal and pre-formal mathematics that we must start the study from: philosophically, the (possible) language-to-world correspondence comes later on.

That latter correspondence is of course the central epistemological and ontological problem of mathematics, but we do not need to explain it in order to use Tarskian truth. Formal languages were designed to make our pre-formal mathematical ideas maximally unambiguous. Whether or not we postulate a theory about the connection between our pre-formal ideas and mathematical objects, the connection between formal and pre-formal ideas can be discussed. That is an important part of Tarskian truth, and correspondence, in mathematics.

How does all this relate to the general philosophical discussion on Tarskian truth? As was said before, it is a matter of debate whether the T-scheme is (in any conventional sense) a correspondence theory of truth at all.[92] This is connected to the debate whether the T-scheme is deflationary or not, which in turn mirrors the debate whether Tarski needs semantical concepts in his definition. These are all current topics of philosophy of truth, and providing a detailed presentation of them goes beyond the scope of this work. However, as far as mathematical truth is concerned, we can learn a lot from a point presented by Ilkka Niiniluoto (1994, p. 63) in the general truth-theoretic discussion. In his scientific realist account Niiniluoto claims that the T-sentences cannot be deflationary because, as a whole, they state something about the *relation* between a language and the world. This point of view had been contested earlier by Field (1972) who, in the usual disquotationalist way, claimed that the T-scheme merely gives us a list of true sentences. Once this list is enumerated, according to Field, we can eliminate truth from it, and see that the T-scheme stated nothing about the relation between the language and the world after all.

Even if Field's point were valid, it would presuppose that the expressive power of the T-scheme applied to an object system **S** is exhausted by the list of true sentences of **S**. But we are not only talking about the true sentences of an object system **S**. Let us look at the T-sentence:

"Snow is white" is true in **S** if and only if *snow is white*.

This is, of course, a sentence of the metasystem (in the metalanguage). The list of all sentences like this will certainly appear to be disquotational. However, in addition to such T-sentences we also have generalizations in Tarski's theory of truth. In most areas such generalizations will be finite and included in the list of T-sentences. At this point, however, we must remember that the subject matter here is truth in *mathematics*. We remember

[92] See also Woleński 2001, pp. 67-68.

Shapiro's sentence $\mathbf{T'} \vdash \forall x(\mathrm{Pr}(x) \rightarrow Tr(x))$, where **T'** is PA expanded with Tarski's theory of truth. Generalizations like this cannot be put down as finite lists, and here Tarski's theory shows its power. We should also remember that it is exactly such generalizations that Field believes to be the only role for the concept of truth. Hence, this is clearly something we *wish* from a theory of truth.[93] But by now we also know that the truth of sentences of **T** in **T'** is not deflationary to proof in **T,** because of the existence of Gödel sentences and the semantical arguments over them. In addition to being useful, which Field is ready to grant, truth is also substantial, unlike Field claims.

This is where we should apply Niiniluoto's point.[94] We have ended up with a substantial notion of truth, so it must indeed say something about the relation between language and the world – or in our case, between the formal mathematical theories and their reference, the pre-formal mathematics. After all the considerations so far, it should be easy to see what this relation is: it is accepting that **T** is *sound* – that the axioms and proof methods of **T** actually provide us with true sentences of mathematics. Clearly this relation is a semantical one, and it is not included in the T-instances of the sentences of **T**. There exist non-formal references for the theorems of formal systems, and accepting the proof methods and axioms is the semantical connection between those references and the formal systems. *That* is what expanding formal systems with Tarskian truth implies. It is not a case of formalist correspondence where we arbitrarily decree the soundness of formal systems. Instead, it is a statement that we make about certain formal systems being sound *because* they correspond to our pre-formal ideas of mathematics.[95]

---

[93] Although, as we saw, this particular generalization was rejected by Field. But his grounds for doing that were highly dubious.

[94] It must be noted that Niiniluoto is not talking specifically about truth in mathematics, so my use of his point here does not necessarily go together with Niiniluoto's intended approach.

[95] This works also as an argument against Jean-Yves Girard's famous criticism of Tarskian truth. Girard (1999, pp. 4-5) writes that:

Of course this does not mean that the belief in soundness of formal mathematical systems leads to believing in a semantical, substantial notion of truth. We remember that Tennant's soundness principle states the same thing: that our proof methods are valid and axioms are assertable. The problem with Tennant's approach is that he uses the soundness principle as something distinct from other aspects of mathematics. In his account, for arbitrary proof methods and axioms, we simply decree their soundness. Without any theory around it, that assumption cannot be defended, since it cannot distinguish between the soundness of our desired formal systems and the soundness of *all* formal systems. Applied to our desired formal systems, the assumption of soundness is of course the most natural one we could make. In fact, could a working mathematician believe any differently? However, as a philosophical answer Tennant's "detached" soundness principle comes down to two options. First, it is possible that our arbitrary proof methods and axioms have hit the bull's eye, and they are indeed the ones giving us the correct sentences. Or second, it could be the case that various proof methods and axioms are sound, and ours happen to be among them.

It should be evident by now that we cannot accept either one of these positions. The second one fails simply because different proof methods will prove different sentences and be contradictory. If mathematics had no application whatsoever, this second point might not be a problem for the extreme formalist. But when we consider all the applications of mathematics, it is impossible to think that contradicting statements could be as useful. We essentially use one theory of arithmetic, and one theory of calculus, and this is the case for a good reason.

> …the notion of *truth à la Tarski* avoids complete triviality by the use of magical expression 'meta': we presuppose the existence of a meta-world, in which logical operations already make sense.

Indeed this is a problem if we stick exclusively to formal languages. But when considering the formal languages of mathematics with regard to their pre-formal 'meta'-counterparts, the problem no longer exists. We *do* have a language prior to the formal ones, and in that language the logical operations make sense.

The first option obviously makes mathematics, and its applications, akin to religion or magic. Both approaches – indeed, any account of extreme formalism – remind one of stories of peasants going to town, seeing water taps for the first time, and buying a bag full of taps to take home to their villages. Formal mathematics works, but like a water tap, it works for a reason not visible from the end product. We cannot forget the background and simply use the formal part in the philosophy of mathematics; without the background we could have never developed the formal part in the first place. It is not the case that belief in soundness implies substantial truth and refutes extreme formalism. But I argue that it *is* the case that we can plausibly believe in soundness only if we reject extreme formalism and *arbitrary* soundness principles.

In this way, mathematics without any references to something outside the formal systems is impossible. The problem with Tarskian truth has been that philosophers tend to connect it with correspondence, and correspondence in philosophy of mathematics sounds ominously Platonist. But as Niiniluoto pointed out, Tarski's T-scheme says *something* about the relation between formal mathematics and its reference. In deeper analysis, this turns out to be nothing more than the belief in the soundness of our proof methods and axioms. We use certain formal systems because we believe that they give true sentences, *not* because we believe that they correspond to a Platonist world of mathematical ideas. In addition to many other desired results, this also explains why we can disagree about the nature of mathematical objects while agreeing on which theorems are *true*.

In this practical sense, our conception of mathematical truth comes before reference: we have a good idea which mathematical sentences are true, even though we might not be entirely clear on the meaning of truth.[96] The reference of mathematical theorems

[96] This is not to suggest that truth comes *constitutively* before reference. That *neo-Fregean* line of thinking will be the subject of Chapter 7. The point here is that one does not need to agree on the philosophical questions concerning truth in order to agree on the set of true mathematical sentences.

can be thought to be any of the various positions that philosophers have introduced, yet we can still agree on the use of truth. Platonism, naturalism, empiricism, structuralism and other referential philosophies are all compatible with Tarskian truth. Whatever the "mathematical reality" formal systems refer to is thought to be, if we trust mathematics, we must believe that our proof methods and axioms are the way of finding out truths about it.

How can we be so sure that Tarskian truth is compatible with every major philosophical account of mathematics other than extreme formalism? What I propose is that the first level of "reality" we are talking about here is in fact the realm of our pre-formal mathematical thinking, and pre-formal thinking is compatible with all the aforementioned philosophies. I also make a stronger claim: all philosophical theories of mathematics *must* be able explain pre-formal mathematical thinking as a phenomenon. Whether this second claim holds or not, the first one certainly will – and both of them are directly opposed to extreme formalism. Most importantly for the purpose of this work, substantial Tarskian truth itself does not require a Platonist framework, and the same goes for the semantic arguments based on it. All that is required is an account of pre-formal mathematical thinking.

In such an account, Tarskian truth can be used as the theory of truth between formal and pre-formal mathematical thinking – the meanings of the formal sentences being their pre-formal counterparts. In this way, truth is for the pre-formal mathematics what proof is for the formal systems. When we see the truth of Gödel sentences, we see it pre-formally. When in the semantical arguments we expand the formal systems to include Tarskian truth, we are only expanding them to include something that was already in our mathematical thinking. That is why the semantical argument is sound. Ketland claimed that truth beats soundness principles as an expansion because it is more natural. That is true. But I claim that this is the case because it is in fact *not an expansion at all*, once we look at the full picture of our mathematical thinking.

However, it must be noted that this is only the first level. Pre-formal thinking exists, and – in Tarskian terms – it works as the metasystem for the object system of formal mathematics. Formal

mathematical languages are designed to be maximally unambiguous abstractions of our pre-formal mathematical thinking, and as such Tarskian truth fits the larger picture of mathematical thinking perfectly. In particular, it goes well with one crucial aspect of mathematical thinking: that formal mathematics also influences our pre-formal thinking. Pre-formal mathematics is not fixed and independent of the formal part; on the contrary, the formal results of mathematics often suggest revisions to our pre-formal thinking. Yet this does not mean that pre-formal thinking can become superfluous: whether such results lead to changes in axiomatizations is once again a question decidable only pre-formally.

Still, pre-formal mathematics itself needs an ontology, and a metasystem, behind it – or else we face the same questions of arbitrariness all over, this time with regard to pre-formal mathematics. One is bound to ask what the references of pre-formal concepts are, and that is the second level of Tarskian correspondence in mathematics. The familiar ontological and epistemological problems of philosophy of mathematics follow us there, and the full picture of correspondence in mathematics can be drawn as follows:

(1) Formal mathematical systems (defines proof)

(2) Pre-formal mathematics (defines truth of (1))

(3) The reference of pre-formal mathematics (defines truth of (2)).

As we see, a full philosophical theory of mathematics needs to explain the connections from (1) to (2) and (2) to (3). Still, the mere existence of pre-formal thinking has philosophical importance, and I do not think that this has ever been fully pursued. Pre-formal thinking enables us to use the semantical argument, and as such it gives us an explicit case of a difference between truth and proof. But what is most important, it does this without troubling ontological commitments, since the connection between (2) and (3) can be left open. Of course there are other ways of defining

Tarskian truth for formal mathematical systems, most notably expanding them with set theory or second-order logic. However, these approaches are often attacked from the formalist camp with accusations of ontological wastefulness. Set existence in particular has been for a long time the paradigm case of problems in the ontology of mathematics, and second-order logic has been argued to run into similar problems.[97] The problem with Shapiro's and Ketland's approaches was that they needed justification for using such expansions. Pre-formal mathematics, on the other hand, is not an expansion at all. We get all the set theory we need for Tarskian truth simply because it is already a crucial part of our mathematical thinking. To formal theories like PA set theory is an expansion. To pre-formal mathematics it is not. Set existence, however, is a whole other question, and with pre-formal thinking we are not saved from the possibility of fictionalism. But that question comes *after* pre-formal mathematics, and that is why we can concentrate in this work on the connection between (1) and (2), while being content to show that the level (3) *exists*, even if we cannot present a satisfactory philosophical theory concerning the nature of it.

It is a fallacy to think that we cannot discuss truth in mathematics if we do not know what mathematical objects ultimately are. After all, I am arguing here for Tarskian truth in mathematics, and Tarskian truth is (notoriously) ambivalent concerning verification principles and such. By itself, it does not give us anything in terms of finding out true sentences, or establishing them as true. Nor should it. It must be remembered that we are concerned with mathematical truth. However one

[97] George Boolos (1998, pp. 73-85) has argued against this to the effect that his plural reading of second-order variables in second-order logic escapes such troubling ontological commitments. His solution has been contested among others by Michael Resnik (1988) and Philippe de Rouilhan (2002) on the basis that such plural readings may not be as ontologically innocent as they look. We will return to second-order logic later on in Chapter 5.4, and we will see that regardless of the success of Boolos' suggestion, it will not provide us with anything resembling an unproblematic way to introduce truth for formal languages.

looks at it, and here the formalist and Platonist agree, it is the purpose of mathematics (formal proof, in the end) to establish true sentences. The difference comes in what we mean by "true", or indeed, if we mean anything at all by it. That is an important question in the philosophy of mathematics, but it is not the same question as the nature of mathematical objects. However, from all that we have established so far, we must mean *something* by truth. The Platonist, structuralist, empiricist and reconstructive nominalist all agree on the set of true sentences. Clearly there is something about mathematical truth that does not change based on our philosophical leanings.[98] The most natural way to start unravelling this is to figure out the relationship between formal and pre-formal mathematics. In the next chapters I will try to clarify this position.

4.4 Another approach to mathematical thinking

Perhaps it is understandable that the philosophy of mathematics, with the apparent *a priori*-nature of its subject matter, has remained the one area of philosophy of science where arguments based on the actual modes of practice of the science are rarely applied. Whatever practising mathematicians do is one thing, but there seems to be a widely accepted notion that there exists *mathematics* outside the work of *mathematicians*, and the former is the real subject matter of the philosophy of mathematics. This does not need to be Platonist – in fact, ironically, many formalists speak with equal aplomb about mathematics consisting of formal systems, without having much regard to the way these formal systems are actually created or presented. While the philosophy of physics is nowadays closely intertwined with the beliefs of the working physicists, there is a widely entertained implicit belief that a philosopher of mathematics has some kind of direct access to the subject matter of mathematics. In fact, surprisingly few

[98] Although, of course, there are also philosophical programs like intuitionism that *do* change the set of true sentences.

philosophers seem to give the actual practice of mathematics *any* role in their theories.

This is not to suggest that the philosophers of mathematics should be realists because the standard language of mathematics is realist-flavoured, or any such superficial connection. That is what Shapiro (1997, p. 38) calls "working realism". Most mathematicians work *as if* realism were correct, using phrases like "there exists a function", etc. Working realism is an interesting facet of the mathematical practice, but it should not be thought to have direct philosophical importance. However, mathematics does have a widely accepted methodology, which in addition to the formal part also gives us a more general paradigm of mathematical thinking, and mathematical practice. Whether we choose to appreciate this in philosophy or not, at the very least we should be aware of it.

It must always be remembered that mathematics, whether its subject matter is considered to be Platonist, formalist or anything else, is a human endeavour as well as any other science. From absolutely no knowledge of mathematics it has been developed into unimaginable heights, providing along the way important, perhaps indispensable, tools for the other sciences. Mathematicians have also succeeded in finding out ways to pass this knowledge on to non-mathematicians for them to use to their own needs. Thinking back from the unknown first steps of mathematical thinking to the current educational system from universities to primary schools, it should be obvious that mathematics as a human endeavour is enormously widely spread, as well as highly advanced and complex. Of course this has not been an accident. Mathematics, probably more than any discipline of inquiry, has had universally established and explicit methods which have accounted for its success. These methods go beyond such obvious features as the accepted rules of proof. They include the whole accepted paradigm of mathematical thinking, including the language, practices, education and the connections to other fields of science.

It seems that this is something most philosophers of mathematics do not recognize widely enough. Of course philosophy should not be confused with sociology and psychology, but to get a complete picture of mathematics

philosophically, we must look at mathematics as a human endeavour in all its facets. This is the approach I now want to take. So far we have only been studying formal theories, which of course form a crucial part in the complete picture of mathematical thinking. But as I see it, mathematics as a human endeavour must include two important disciplines beyond formal systems: the work of creative mathematicians, and the work of educators. The former is practised almost exclusively at universities, and it can be said to consist of making *new* mathematics. The latter is practised on all levels of education. Curiously, the former seems to be more often appreciated in the philosophical literature. It is not rare to see philosophers of mathematics use examples of mathematical discoveries (inventions, if one leans toward formalism) to make a point. It is an interesting subject, to be sure, as every theorem has at some point been a new discovery. However, it is also a notoriously difficult subject to study – the group of creative mathematicians is very small and mathematical discovery is obviously hard to achieve in anything resembling controlled settings. Instead of double-blind experiments we are limited to a handful of subjective post-discovery narratives. Even so, some material does exist from the quasi-psychological accounts of mathematicians, and we should definitely learn as much as we can from the studies we have.

As interesting as that is, the more pertinent subject for us is the teaching and learning – *understanding* – of mathematics. Mathematical knowledge could not be possible if we did not have means of passing it on to other people, which brings us to the subject of education. As philosophers of mathematics, we should be very interested in how mathematics is taught and learnt. In this work, however, I will not try to present or analyse psychological theories concerning the learning of mathematics. Those are no doubt highly interesting, but I will only need minimal psychological evidence for my argument. Fortunately this evidence is there for everybody to see, even if we totally dismiss all psychological theories. When we consider mathematics as a human endeavour, the first observation to make is that practically all human beings have learnt mathematics from *textbooks*, whether directly or indirectly. From the textbooks of mathematics we

should get strong enough evidence for one central aspect of the psychology of mathematics. All textbooks of mathematics have one important characteristic in common: *they are not completely formal*.

Of course the level of formalism varies, but it is easy to notice a distinct pattern. Young children's textbooks have a minimal level of formalism, while university textbooks can be mostly formal, the informal part being restricted to very little "narration". Still, even the university students do not learn entirely formal mathematics. They need examples, diagrams, pictures and natural language explanations – the mixture of which depends on the subject. Why can we be so sure of this? All textbooks *have* them. Could there be any better evidence? This does not concern just the textbooks, either: even the articles in mathematical journals, which must be considered to be the most advanced form of mathematical communication, are not completely formal. In fact, we should feel safe in assuming that there are no human beings who process mathematics as computers do, that is, completely formally.[99] I could be wrong – the possible counter-example sounds intriguing – but that would not change the general argument here.[100] It is certainly true that if not everybody, at least the vast majority of people need something besides the formal part in order to understand and use mathematics. This could be called the *informal* part, but I prefer to call it *pre*-formal mathematics to emphasize the order in which our mathematical understanding develops.

---

[99] I am not including here the possible *savant*-type exceptions. This has been suggested to me as a formal way of learning mathematics, as the savant may not have any reference to outer sources and still process mathematics at high levels. However, what they are doing is *calculation*. Mathematics is obviously quite a different thing, and while it remains a possibility, savant-type mathematicians have not been reported.

[100] Of course this concept of understanding must not be confused with what happens at the level of the brain. On a cellular level, our thinking could be ultimately like that of a computer.

4.5 Pre-formal mathematics

The concept of pre-formal mathematical thinking is essential to this work, and it needs elaboration. The details that we attribute to pre-formal thinking, however, should not be considered to be crucial for the arguments here. More important is the fact that the phenomenon of pre-formal thinking *exists*, and that it is bound to bear enough resemblance to the account here in its central facets. As it is presented in this work, pre-formal mathematics consists of two sides. First, there is the individual learning of mathematical concepts, to which we will return later. Second, even those who are familiar with mathematical formalism still use the pre-formal element in their thought process all the time, even when the results get a purely formal presentation. Constructing a mathematical proof is not only about mechanically grinding out the formalism; it also includes the crucial stage of discovering the connections and ideas that will be the basis for the formal presentation. One crucial part of this is the discovery of *new* mathematical theorems. Of course this only concerns a minuscule part of all the practising mathematicians, but that part is all the more interesting.

Because of the elusive and heterogenic subject matter, comprehensive psychological studies of mathematical discovery/invention are obviously too much to ask. The best we have are the regrettably few accounts of the subjective experiences of mathematicians. Although obsolete in its psychological terminology, the mathematician Jacques Hadamard's *The Psychology of Invention in the Mathematical Field* (1954) is probably still the most important work in this area. The bulk of that book is based on Henri Poincaré's account of mathematical invention, where such matters as the unconscious element, mental images and the aesthetic aspects of mathematical discovery are given an important role. In Hadamard's research he found out that most mathematicians shared similar experiences. The details of them are fascinating, but as such not central to this work. What *is* important is that Hadamard's book gives us clear evidence that the psychology of mathematical invention is not reducible to the neat formal accounts that are the end product of mathematical studies.

Mathematical thinking as a human phenomenon is a vastly more complex and broad field.

However, it must be remembered that Hadamard is concerned with the discovery of mathematical truths, which is only half of the picture. At least as important is the way that we *justify* believing in such supposed truths. That of course happens ultimately by proving them. Mathematical discovery/invention by itself could be thoroughly un-mathematical – which it of course is not – but as long as the discovered/invented theorems can be proven, the non-formal elements included in the discovery could be philosophically irrelevant. But from Hadamard's book we get a different picture. The psychology of mathematical invention is closely connected to the formal mathematics, and all our non-formal ways of processing mathematics make for an indispensable part of mathematics as a human phenomenon. I want to extend that conclusion to mathematical thinking in general, and not just the context of discovery.

Of course this approach as such is nothing drastic: even extreme formalists would not claim that mathematics does not include a non-formal element. What they do claim is that in the philosophical accounts of mathematics this element is essentially superfluous. However, in this chapter I will argue that this is not the case. The recognition of pre-formal mathematical thinking is essential to the philosophy of mathematics. In the model proposed in Chapter 4.3, mathematics consisted of three parts. Starting from the end product, the part (1) is formal mathematics. The part (2) is pre-formal mathematics, which is our actual mathematical *thinking*, how we process mathematics "in our heads". This part is essentially semantical, dealing with the *meanings* of the theorems of formal mathematics. That is why in the pre-formal part we use examples, diagrams and informal presentations – they give us a better *understanding* of the meanings of the formal concepts. The part (3) is the reference of pre-formal mathematics, that is, the subject matter of mathematics: what the theorems of mathematics *ultimately* refer to.

How are these parts of mathematical thinking connected to each other? *Proof* is obviously in the realm of formal mathematics, and it is designed to correspond to our pre-formal ideas of *truth*,

which in turn corresponds to the part (3), the final subject matter of mathematics. In this way, there is a connection through all the stages. Had Hilbert's program been established successfully, formal theories of mathematics could describe a direct correspondence between the parts (1) and (3). However, that would not have done anything to make the pre-formal thinking obsolete. In the practice of mathematics it would most likely have caused no changes. Certainly the completeness and consistency of formal systems would have been important results in the philosophy of mathematics: ultimately, they would have shown pre-formal thinking to be superfluous in the connection between formal mathematics and their references. But even so, it would not have changed the fact that human beings process mathematics semantically. Although the philosophical importance of pre-formal thinking may have been diminished, all three levels of mathematics would still have been needed to make a theory of philosophy of mathematics complete. Knowing what happened to Hilbert's formalist program, it is all the more important to recognize all three levels.

We will return to the problematic questions concerning the level (3) later on, but for now we are concerned with pre-formal mathematical thinking. It is of course an indisputable fact that the formal theories of mathematics did not just suddenly appear to human beings. We know that it took the work of some of the most brilliant minds in ancient Greece to find an unambiguous presentation for the mathematical knowledge of the time, which in turn was based on centuries, even millennia, of earlier study. Although this presentation was mostly written in a natural language, and would not be recognized as formal by a modern reader, it was still essentially formal mathematics. Ambiguous considerations based on observations were replaced by exact definitions, axioms and rules of proof. However, all this was based on something – it did not appear via an epiphany. Obviously no written account of the process exists, but we can safely assume that, for example, Euclid's formal concept of the "point" as an entity without dimensions was not the original concept of "point". Rather, it was an idealization that the mathematicians needed and developed. When we think of a direct route from one house to

another, we are essentially thinking of a line segment between two points. Of course houses are not points and routes are not lines – nothing physical is – but they correspond to the same *idea*.[101] This idea of a straight line between two objects is quite clearly pre-formal, just like the ideas of circles, natural numbers and probabilities are. We do not need to know anything about the formal mathematical presentations of these concepts to be able to have – and even successfully *use* – their pre-formal ideas. That is of course because formal mathematics was developed to be a maximally unambiguous study of such existing pre-formal concepts. Pre-formal concepts were not *replaced* by formal ones, they were *clarified* by them.[102]

What kinds of areas belong to pre-formal mathematics and can we hope to give a satisfying account of it? Certainly these are not easy questions to answer, and I do not pretend to give a comprehensive explanation here. It seems that almost anything concerning mathematics as a human endeavour can be considered to belong to pre-formal mathematics – aside from the formal part, of course. In this way, every physical object is potentially an object of pre-formal geometry, and every quantity is an object of pre-formal arithmetic, or some other area of mathematics. Pre-formal mathematics can be thought to include the unconscious element of mathematical invention, and it can be thought to include dividing a pile of apples into smaller piles. However, clearly not *everything* we do with such objects can be considered to be pre-formal mathematics: an activity only becomes mathematical once we are trying to find out general truths about the objects and the relations between them – the ultimate phase of this activity being the

[101] Here I do not use the word "idea" in any Platonist sense, but rather in the most general sense we use it outside metaphysics.

[102] Tarski speaks about *formalized* languages, which corresponds to my argument here. In mathematics (for the most part) we are concerned with meaningful, interpreted languages, not arbitrary formal rules of symbol manipulation – which is what formal languages ultimately are for the extreme formalist. We will return to this question later.

formalization of mathematics.[103] Even so, admittedly, these considerations make pre-formal mathematics a vast and somewhat vague field. But in lack of a better account, there should be nothing troubling about using the one given here. The point I want to make is that the domain of mathematical thinking is much larger than the mere formal part.[104] This is important when we consider the problems of reference and truth in mathematics. The exact nature and scope of pre-formal mathematics should not matter a great deal, as long as we are more or less along the right lines. I do not believe it can be plausibly argued that we are not.

The pre-formal element can be witnessed everywhere, but nowhere more visibly than in education. The examples here will be simplified and, again, in no way do I claim them to be accurate and complete descriptions of the learning process in mathematics. But they should be plausible enough to give us some philosophical perspective into mathematical thinking. How do we initially learn about, for example, triangles? The teacher draws a triangle on the blackboard and we start examining its properties. This way we learn that the sum of angles of a triangle is that of two right angles. But of course at this stage we never *really* deal with a mathematical triangle, only an imperfect drawing of one. We did not prove that the sum of the angles is that of two right angles, either – we probably just had a visual presentation that convinced us.

---

[103] For an example, an illuminating one is a passage on mathematical knot theory by Crowell and Fox (1963 p. 3), quoted by Shapiro (2000a, p. 35):

> Mathematics never proves anything about anything except mathematics, and a piece of rope is a physical object and not a mathematical one. So before worrying about proofs, we must have a mathematical definition of what a knot is. […] The definitions should define mathematical objects that approximate the physical objects under consideration as closely as possible.

In this quote the authors are quite clearly concerned with formalizing the pre-formal, in this case physical, concept of knots.

[104] For a reference in the psychological study of mathematics, one can consult Davis 1984, which emphasizes how people *think* about mathematics, how they process it through meanings. Also relevant is Tall (ed.) 1994, a collection of articles that focuses on advanced mathematical thinking and the role of various non-formal elements in it. For philosophical studies Lakatos 1978 is relevant when it comes to the classification of the different stages of mathematical thinking.

Moreover, this does not need to be visual. In Hadamard's example (1954, p. 62):

> …everybody understands that, intersecting two parallel lines by two other parallel ones, the segment thus determined are equal two by two; everybody knows that […] But as long as it is not consciously enunciated, none of its consequences […] can be deduced.

This purely verbal presentation seems to be perfectly valid. However, in both cases, in the purely formal sense, we did not acquire any mathematical knowledge – we did not *prove* anything. Still, it would not make sense to claim that we did not gain any knowledge. In the first example, we did learn a property of triangles that we did not know before. We just did not make the knowledge *formally rigorous* by proving it from axioms, which is what formal mathematics does. This gives us a characterization of the basic distinction between formal and pre-formal mathematical thinking: any mathematical thinking, and knowledge, *that is not formally rigorous is pre-formal*.[105] This does not mean that we are unable to gain mathematical knowledge pre-formally. The sum of the angles of a Euclidean triangle, for example, is a mathematical *truth* that most of us initially learn pre-formally. We do not justify it rigorously in axiomatic systems until much later, but we undoubtedly have knowledge of it all along. Moreover, it is knowledge unlike memorizing a fact like "Nicholas II was the last Tsar of Russia". Clearly we learn it by establishing general connections between concepts like triangle and angle, rather than relying only on an authority to give us correct information. Indeed, not surprisingly, the way these connections are established pre-

[105] Here we deal with the term "rigorous" somewhat loosely. It could be that the results of formal mathematics are not completely rigorous, either, due to problems like the unprovability of consistency. On the other hand, pre-formal mathematical thinking can also be rigorous, even though this may not be unambiguously established until it is formalized. Nevertheless, the distinction between formal and pre-formal rigor here should not be problematic, which I want to emphasize with the concept "formally rigorous", which means proof from a specified set of formal axioms according to formal rules of proof.

formally mirrors the way they are proved formally. This is the important point here: *formal mathematics is designed to prove just such mathematical truths*.

Let us think of a child learning mathematical concepts for the first time.[106] From picture books and toys she sees a round shape. From her parents or other older people she learns that it is called a circle. For years she will deal with circles, perhaps learning some of their properties, like that any diameter will be equal in length, and that with a compass one can draw a circle. She, however, has no idea how these properties are presented exactly, or how they in fact relate to each other. This is the primitive phase, how we begin mathematical thinking, and it consists of getting ideas of the mathematical concepts.[107] After this the child learns the exact ideas behind, and between, the concepts. The child learns that a circle is a line that forms from the set of all points at a fixed distance from a fixed point. Now she has an explanation for the fact that circles can be drawn with a compass, and why all the diameters are equal in length. This intermediate phase of mathematical thinking consists of forming explicit and general notions of the mathematical concepts; in other words, abstraction.[108]

The final phase will come when she learns to translate these exact ideas into a formal mathematical language. She learns, for example, that any given circle can be expressed as an equation, and with this information she learns to treat geometric objects with algebraic tools. Now the child (or perhaps an adult by now) has the means of abstract mathematical thinking, and the ability to express it with maximal precision, that is, formally.[109]

---

[106] This example corresponds in its central parts with the account in Davis 1984, at its simplest given in pp. 8-20.

[107] This is called the "proto-mathematical" phase by Philip Kitcher (1983) in his account of mathematical knowledge.

[108] See Russell 1912 (pp. 119-121) for his account of the same phase in mathematical thinking.

[109] Imre Lakatos has made a similar distinction concerning mathematical proof (my account is about the whole phenomenon of mathematical *thinking*). According to him, there are pre-formal, formal and post-formal proofs. Compared to my account, both the pre- and post-formal proofs

What in this example is pre-formal mathematical thinking? The primitive phase is unquestionably pre-formal. While the intermediate phase consists of finding exact definitions, it is still done in a natural language, and thus cannot be considered formal in the strict sense that an extreme formalist would accept. It could be argued that the step from the intermediate phase to the final one is mere translation from informal to formal languages, the exact definitions being already in place. This is true up to some point, but it has to be remembered that the concepts themselves are still defined in a natural language, and are thus not entirely unambiguous. It is hence only the final phase that we can call formal in the strict sense. It goes without saying that thinking of mathematics only as this final formal phase gives us a highly incomplete picture also of the *philosophy* of mathematics. In fact, it would make mathematics impossible to learn or practise. However – returning to the semantical argument – that is *exactly* what the extreme formalist does when refusing us the right to expand formal systems.

I suggest that this rough account holds for individuals learning mathematics, as well as – *mutatis mutandis* – the historical way of human beings developing mathematics.[110] In addition, while for the developed mathematicians the primitive phase is largely forgotten, the intermediate phase is intimately present at every aspect of mathematical thinking. Obviously it is by no means easy to create a detailed account of mathematics as a human phenomenon, nor is this the purpose here. Rather than providing a comprehensive account of mathematical thinking, my aim has been all along to show that formal mathematics by itself will not

---

would count as pre-formal mathematical thinking. Otherwise Lakatos' ideas are similar as far as the different stages of mathematical thinking are concerned. His ideas about the fallibility of mathematics should be seen as a whole other matter. For details, see Lakatos 1978, pp. 61-69.

[110] The historical development of mathematical thinking has been a very widely studied area. John Barrow (1992, Chapter 2) gives a good overview of the knowledge we have about the early development. The bibliography assembled by Barrow is particularly important. Boyer 1985 is a widely relevant work on the development of sophisticated mathematics.

suffice. Whether one agrees with the details or not, I hope the considerations here are enough to convince one of the need for expanding formal mathematics in order to have a complete picture of mathematical thinking.

At this point we should think of the philosophical importance of all this when it comes to the subject matter of this work: the question of truth and proof. We remember that in the extreme formalist account the pre-formal element would be superfluous, and mathematics completely reducible to the formal part. However, once we acknowledge the role of pre-formal thinking, we see that this is not the case. In the pre-formal phases of mathematical thinking we have the notion of mathematical *truth*, and in the formal phase we have the notion of *proof*.[111] We have also seen that the truth of Gödel sentences is established pre-formally. That is why the semantical argument is valid: the Tarskian expansion corresponds to something existing in our pre-formal mathematical thinking. In order to carry out the semantical argument we do not need anything more complicated than formal systems expanded with a Tarskian theory of truth, but of course in reality people have all kinds of semantical notions behind the mathematical concepts. Tarskian expansion is enough – that is the beauty of the semantical argument – but from pre-formal mathematics we see that it is not an expansion at all. As I see it, that should be enough to make the semantical arguments successful.

This objection to Field seems to be very powerful, since in the extreme formalist philosophy there cannot be any substantial role for a non-formal part of mathematics. Clearly pre-formal mathematical thinking has an important role when we consider

[111] Perhaps it would be better here to follow Tarski and talk about *formalized* interpreted languages of mathematics, to emphasize the connection between pre-formal and formal mathematics. However, since it is my contention that the formal languages we discuss in the basic areas of mathematics like arithmetic and geometry are always formalizations of pre-formal thinking, the difference between formal and formalized seems to vanish, and the distinction becomes unnecessary.

mathematics as a human phenomenon, and we cannot neglect it. But once we accept the existence of the pre-formal element, we can also establish the truth of Gödel sentences, so the role is also substantial. Moreover, it corresponds naturally to our intuition of the way we see the truth of Gödel sentences. This is no small matter. After all, we initially establish the truth of Gödel sentence through its *meaning*, not by considerations on expanding formal systems. That immediately gives us an apparent difference between the concepts of mathematical truth and proof – and it turns out to correspond to the difference between pre-formal and formal mathematics. In short: the existence of pre-formal thinking refutes extreme formalism, as is made explicit by the semantical arguments.

So why has pre-formal thinking been left to such a small role in the philosophy of mathematics? While the terminology here could be foreign to many philosophers, most philosophers of mathematics do in fact recognize the pre-formal element. What they are not bound to agree on is the nature and especially the importance of pre-formal thinking. For philosophers that are campaigning for a specific realist theory, the existence of pre-formal thinking does not carry particular philosophical importance. In the Platonist philosophy, for example, formal mathematical sentences are thought to ultimately refer to abstract ideas in an ontologically independent world. It is quite understandably not considered to be very important that this happens via pre-formal semantical thinking. The importance and difficulty of the ontological and epistemological questions far outweigh such matters. However, in this work I will not try to present a specific general theory of philosophy of mathematics. Instead, I want to use as minimal and as widely applicable theory of non-formalist philosophy as possible. In such a project the existence of pre-formal thinking is of great importance: it gives us a general framework that can be interpreted in various non-formalist ways.

Indeed, the account here should be compatible with *all* non-formalist philosophical theories of mathematics. The intuitionist Michael Dummett has written about the role of formal systems:

> …a formal system does not *replace* the intuitive proof as, frequently, a precise concept replaces a vague intuitive one; the formal system remains, as it were, answerable to the intuitive conception, and is of interest to us only in so far as it does not reveal undesirable features which the intuitive idea does not possess. An example would be Gödel's theorem, which shows that provability in a single formal system cannot do duty as a complete substitute for the intuitive idea of arithmetical truth. (Dummett 1978, p. 172. Italics in the original.)

Both Dummett's example and his ideas on the status of the formal systems should sound familiar by now. What is remarkable is that Dummett as an intuitionist is still in agreement with the non-intuitionist account given here. Formal systems originate on pre-formal thought (intuition for Dummett), and the pre-formal element is always carried along with the formal systems. Of course we do not need to agree with Dummett's idea on testing theorems by comparing them only to our intuition, but that is a feature of his intuitionist philosophy of mathematics. What I think we should all agree on is that any satisfactory account of philosophy of mathematics should include *some* explanation for the pre-formal element and its connection to formal mathematics.

4.6 Philosophical importance of pre-formal mathematics

Although we have glanced at the philosophical importance of pre-formal mathematics, it should now be fully disclosed what it has to do with questions like mathematical truth. I think that the philosophical importance of pre-formal thinking can be divided into three parts. First, as we have seen, it is only with the inclusion of pre-formal thinking that we can draw a full picture of mathematical thinking. We should find it very surprising if this did not have any philosophical significance at all. Second, it immediately gives us a way to expand from formal mathematics into a more allowing conceptual realm. We can speak of Tarskian truth and the hierarchies of languages, we can speak of reference, and we can speak of the origins and nature of mathematical knowledge. In other words, we get all the conceptual richness we

need in order to show the substantiality of truth. This has been a factor up to this point, and it will be one for the rest of this work. Pre-formal thinking by itself does not solve the questions of truth and reference, but it gives us the conceptual tools to do so.

Third, and importantly for *all* philosophical theories of mathematics, we get justification to speak about the philosophy of mathematics in the first place. One problem with the self-standing formal theories of the extreme formalist type is the connection between mathematics and philosophy. If the mathematical concepts do not refer to anything non-formal, how can we even be justified to discuss them in a non-formal context? What have our non-formal concepts such as "proof", "axioms" and "numbers" got to do with the formal theories where such concepts are supposedly central? If the formal systems are *all* there is to mathematics, then any such talk is not only superfluous, but also misleading. Moreover, we have the problem of theory choice in an even stronger form. Even if we grant the formalist his criteria of, say, consistency and conservativeness as the basis for theory choice, how can we apply them to the formal theories? Indeed, how can we have *any* criterion of theory choice if formal systems are all there is to mathematics? At any given point we would have the contemporary set of accepted formal systems in mathematics, and that is all there is to mathematics. Strictly speaking, no means of choosing between them could exist, since these would have to be justified in a non-formal manner.

This might sound like an unfair limitation to the formalist position, and in a loose way it is. But it goes on to show how limited the extreme formalist position actually is, if taken to its logical conclusion. If mathematics is about purely formal concepts, then what is the relation between them and their non-formal counterparts that we deal with in philosophy? If formal systems do not refer to anything, I do not see any other alternative than to deny the existence of any such relation. Strictly speaking, it cannot make sense for the radical formalist to speak of formal concepts in a non-formal context. The natural numbers, for example, are defined by the Peano Axioms, and it could not make sense to discuss them in a way that is not included in PA. In essence, such strict formalism would make all philosophy of mathematics

impossible. Since all formalists have been ready to publish work about mathematics in a non-formal presentation, we can be confident that such a radical formalist does not actually exist.

All that notwithstanding, of course we should be generous enough to grant formalists the ability to deal with mathematics non-formally, and treat the writings as something akin to analogies. However, this obviously gives us a whole new vocabulary for the philosophy of mathematics, and the relevance of pre-formal thinking can no longer be denied. That brings us back to the second point: all the conceptual tools that pre-formal thinking gives us. We have the ability to talk about meta- and object languages, reference between them, and hence also about semantical truth. If we have such an ability, it would not make sense not to use it. Why insist on talking about mathematics only as the formal systems, when we can naturally include pre-formal mathematics to cover the whole phenomenon as human beings practise it? Of course most formalists are likely to accept this, while insisting that any such talk is superfluous since the formal systems are still all there is to *real* mathematics. However, that is not the case, since by including the pre-formal part we get tools that have relevance also on the formal part. Perhaps the most important of these is the ability to discuss theory choices. We can speak about axioms and rules of proof, their accuracy and their references, in a non-formal context. We can have criteria for the choices between them. In short: we can speak about *truth*.

At this point we should examine what Tarski's own image of the status of his definition of truth was, and how well it conforms to the account of pre-formal thinking here. The answer is obvious when looking at a passage from Tarski's (1936, pp. 166-167, quoted in Woleński 2001, p. 72) famous article:

> It remains perhaps to add that we are *not interested here in* "formal" languages and sciences in one special sense of the word "formal", namely *sciences to which no meaning is attached*. For such sciences the problem here discussed has no relevance, it is not even meaningful. We shall always ascribe quite concrete and, for us, intelligible meanings to the signs which occur in the languages we shall consider.

> […] The sentences which are distinguished as axioms seem to us materially true, and in choosing rules of inference we are always guided by the principle that when such rules are applied to true sentences the sentences obtained by their use should also be true. (Italics mine).

This is more or less exactly what I have been arguing for. With mathematics we are always interested in interpreted languages – in *meanings*, and whatever formal tools we develop, our languages never lose that attachment to the meanings. For Tarski it was natural that formal languages have meanings, and by committing to the axioms and rules of proof we commit to the truth of them. As we have seen, that is also how we avoid the arbitrariness of formal systems. If formal systems are considered to be completely self-standing – that is, un-interpreted languages of pure symbols, empty of content – like Tarski noted, the question of truth becomes meaningless. That corresponds perfectly to the account of extreme formalism in this work.

What Tarski does not say is what exactly is meant by "meaning" here. One thing we do know, however, is that in Tarski's account meaning, and hence truth, was relativized with respect to languages. Woleński (2001, p. 72-73) points out that in the contemporary discussion it was suggested (by Maria Kokoszyńska) that such a theory of truth should be relativized even with respect to the meanings of expressions. In that way, one could argue that meanings in fact come before the truth conditions in the sense that we must first have a full grasp of the former before we can establish the latter. Here we run into a danger of the truth conditions getting too drastic a relativity – something I certainly want to avoid when it comes to mathematical truth. It is clear that a statement like "there exists a number n: $2 < n < 3$" gets different truth-values depending on what set the concept "number" refers to, but in mathematics we certainly do not want to end up with the position that one concept can mean different things within one language.

How big a problem is that? Certainly we do not want to lose any of the clarity and expressive power that formal systems have, and that is what any strong sense of relativity of the truth

conditions is bound to do. However, it must be remembered that while Tarski was concerned with formal languages, he was concerned with them as being *formalizations* of informal languages, not as empty formal languages in the extreme formalist sense. As Woleński (2001, p. 73) points out, the difference between relativizing truth to language and meaning vanishes if we only consider *interpreted* languages, which is indeed what we have been doing all along. When it comes to the philosophy of mathematics, we can focus on the interpretation of axioms and rules of proof, and get the meanings and truth conditions as the side product. In fact, this is what I have been suggesting all along. Truth is relative to language, and formal system, but there is nothing problematic in that. When it comes to mathematics, meaning and truth conditions come to play when we *choose* the axiomatizations, and the interpretations. After the choice is made, the truth conditions are fixed, and we escape the threat of strong relativity.

Perhaps it is too strong a word to say that mathematical axioms "force" themselves upon us as true, but it cannot be denied that however we arrive at them, it tends to include a strong conviction where meanings and truth conditions are intertwined. The meaning of formal mathematics, of course, consists of our pre-formal ideas, and it is carried along throughout the process of mathematical thinking. If there are conflicts between the truth conditions and our ideas of meaning, the axioms can be revised, as has been done numerous times. Certainly the relation between formal and pre-formal mathematics is a dynamic one, and the meanings of mathematical expressions can change. But they change via changes in the axiomatizations and definitions, not *within* the formal systems. Relativity in any problematic sense does not need to enter the picture, and we can retain all the exactness of formal systems while introducing Tarskian truth.

4.7 Priority of semantics over syntax

All of the above has been presented as a way to stress the priority of pre-formal thinking over formal systems in mathematical thinking. It has been my purpose to take a wider angle into the whole picture of mathematics as a human endeavour, and thus gain evidence for the priority. However, that priority can also be looked at from another angle, and I believe this is the right point to address that issue. As we remember, although Gödel used informal presentations in his project by using concepts like self-reference and truth, the actual proofs of incompleteness were completely syntactic. Clearly it was Gödel's purpose to find a syntactic presentation for his pre-formal, semantical, ideas. That by itself counts as evidence for the priority of pre-formal thinking over formal mathematics – and hence, for semantics over syntax.

Fortunately, we do not need to limit ourselves to such considerations. The same priority can be seen by studying the relationship between Tarski's undecidability theorem and Gödel's incompleteness theorems. Raymond Smullyan (1992, 2001) has made this approach famous in the literature. While Gödel's theorems are often considered to be the more important result when it comes to the philosophy of mathematics, Smullyan has stressed the value of Tarski's theorem. As Smullyan (2001, pp. 78-79) notes, Gödel's incompleteness theorems follow from Tarski's undecidability theorem, while the opposite is not the case. From Tarski (1936) we know that a formal language cannot contain its own truth-predicate. Let us assume that a provability predicate Pr_T of a formal system **T** is sound. Now if it were also the case that Pr_T is *complete*, it would quite clearly follow that Pr_T is in fact the *truth-predicate* of **T**, which contradicts with Tarski's undefinability theorem. Since **T** was assumed to sound, there would have to exist some true but unprovable sentence in **T**.[112] Thus Gödel's first incompleteness theorem can be immediately reached from a Tarskian semantical starting point.

---

[112] Here Gödel's theorem gives us the extra strength of presenting the self-referential formulation of the unprovable sentence, thus giving us an explicit sentence with which to go through the semantical arguments.

What we gather from Smullyan is another argument for the priority of semantics, and hence, pre-formal thinking. There are others, and the whole relationship between syntax and semantics presents many interesting questions in logic. I cannot go further into that area here, and I hope that this part of the thesis is already sufficiently established. Furthermore, even with all the logical arguments on semantics and syntax, the extreme formalist could always have the last refuge of denying everything other than simple syntax from the domain of mathematical study. Although the Tarskian approach to Gödel's incompleteness theorems may seem unproblematic to us, for someone that only accepts formal systems consisting of primitive recursive functions in mathematics that could be a whole other matter. That is why I have wanted to stress that in mathematics we are not concerned with formal systems empty of meaning. When we acknowledge the existence of pre-formal thinking in mathematics, the semantical approaches like Tarskian truth present themselves to us as very intuitive and convenient tools – but most importantly as something that we simply *must* acknowledge in a philosophical investigation of mathematics, because mathematics without them would be quite different from the discipline we currently know.

4.8 Truth, proof and reference

From all that has been considered so far in this work, it seems indisputable that truth and proof are different concepts in the classical two-valued mathematics. The truth of Gödel sentences gives us a strong argument for the case, but it is not the only argument. The value of the semantical argument lies in its explicitness, and as such it is probably the strongest one. But even if we rejected it, there would still be no reason to conclude that truth and proof are the same concept. Deep down, my final assessment of the semantical argument only used the assumption that there exists pre-formal mathematical thinking. This is already a very weak assumption, as I have tried to show in the last chapter. But Gödel's incompleteness theorems themselves lie on even weaker assumptions, ones that even the formalist absolutely *must*

accept. In classical two-valued logic there is of course the law of excluded middle, and if we equate truth with proof in formal systems, then the Gödel sentence of "our fullest mathematical theory **T**" clearly has a forbidden middle "truth-value". By the very axioms of logic, either $G(\mathbf{T})$ is true (provable) or it is false (disprovable). Since it is neither, we must either revise our logic, or otherwise conclude that truth cannot be just provability.

In the next chapter we will look at the first possibility. But if we stick to classical two-valued logic, incompleteness already shows us why truth and proof are different concepts: all the arguments presented so far in this work are more or less only corollary to it. They are an important corollary, though, as we gain insight on just what the difference is, and how truth and proof are connected to each other. The work here aims to *explain* the difference between truth and proof; that there *exists* such a difference in classical two-valued mathematics is already established by Gödel's incompleteness theorems.

There are other problems for the extreme formalist, as well. We remember that Hilbert's goal was to formalize mathematics completely. This already should give us a hint that the extreme formalist viewpoint will turn out to be rather limited. For something to be formalized, it obviously has to exist first. This should never be forgotten. Even if Hilbert's program had succeeded, it would not have shown that we do not process formal mathematical sentences through their pre-formal meanings, or that they do not have objective references. It would only have shown that mathematics could also be presented (even if not practised) completely without those meanings. Of course Hilbert himself recognized that mathematicians use meanings in their work[113]; only the latter extreme formalists have reached the problematic notion that formal mathematics could be the only mathematics we have. This needs to be stressed once more: formal mathematics is a tool created in order to introduce a maximally unambiguous

[113] As we remember, Hilbert was aiming to *save* the old mathematics of abstract inferences from intuitionism, not to create new mathematics to replace it. See Reid 1970, pp. 155-157 for more on the basis of Hilbert's program.

notation for mathematics. In the case of proof, it was always meant to be (and, with the small exception of the Gödel sentences, still is) the maximally unambiguous way of finding out *true* sentences, by proving them from true axioms.

Then what *is* truth, the deflationist will ask: what is it that makes some of our pre-formal sentences true and some false – and how can we gain knowledge of that? Many philosophers have argued that the semantic theory of truth does not need to explain this. It gives us an account on what the true sentences are like, but it does explain what the concept of truth itself means. Dummett (1976, pp. 51-54), for example, has argued for this idea. As far as the initial purpose of this work is concerned, I am ready to agree with that. In our account of mathematical truth, we have seen that the semantic theory of truth, understood as the relation between formal and pre-formal mathematics, gives us a satisfactory basis for mathematics as a human endeavour. We might not know what mathematical truth ultimately is, but we can know that a semantic notion of truth in mathematics is substantial – and that it should be accepted.

When we consider the philosophy of mathematics as a whole, however, this is not satisfactory. If we use pre-formal thinking with Tarskian truth to make an argument, we must account for the reference of both the formal and pre-formal mathematical sentences. Pre-formal mathematics cannot rest on nothing, or we run into all the same difficulties of arbitrariness as the extreme formalist does with the formal sentences. In short, to make the approach here acceptable, pre-formal mathematical sentences must be saved from arbitrariness. Ultimately, this must mean that we have criteria for asserting some pre-formal sentences over others. While my approach of calling those criteria *truth* saves us from arbitrariness, this begs the question what it is that *makes* mathematical sentences true. The answer must be something non-arbitrary: that (at least some) mathematical sentences have objective references. Showing what these references are, however, looks like a daunting task: such an account would have to answer all the highly problematic metaphysical and epistemological questions in the philosophy of mathematics. Essentially, that

would make any independent account of mathematical truth impossible.[114]

Fortunately, however, we do not need to explain the *nature* of the reference of mathematical sentences here. It will be enough for the purposes of this work to convince the reader that such a reference – a reason for the objectivity of some mathematical truths – *exists*. Of course our luck continues because there only is one noteworthy point of view in the philosophy of mathematics that denies *any* reference for mathematical sentences, and that is extreme formalism/nominalism. In Chapter 6 I will argue that many doctrines that are called nominalism in the literature are not nominalistic in the strict sense required here, and are in fact perfectly compatible with Tarskian truth. Only formalism of the extreme type would conflict with the arguments of this work – and by now we should have a good idea how problematic extreme formalism is. If we manage to show that extreme formalism runs into arbitrariness, we must conclude that some objective reference for mathematics is needed. This way, other than extreme formalism, all accounts of philosophy of mathematics are compatible with the approach here. We may not know which objectivist philosophy of mathematics is correct, but we can still perfectly well know that the correct one must be *some* objectivist philosophy rather than extreme formalism.

However, before we go deeper into nominalism we need to clarify one assumption that has been used throughout this work: that our theories of mathematics are in *classical first-order two-valued logic*. Many of the central results used in the arguments here, most importantly the undefinability of truth within formal languages, depend on us using classical first-order logic. Let us next examine some of the possibilities that other logics present us with.

---

[114] This is not to suggest a form of *quietism*, according to which meaningful metaphysical debate is impossible. But I do contend that we should put such debate aside as long as we can. To exaggerate a bit, almost any form non-metaphysical explanation is bound to be more satisfactory than a metaphysical one.

5. Truth and logic

5.1 Different logics

So far Tarskian conception of truth and the semantical arguments have provided a strong case against extreme formalism and deflationism in the philosophy of mathematics, but that is due to us having to expand beyond formal systems in order to include an adequate definition of truth. If we could find a way of defining truth in formal languages *within* those languages, the question of truth and proof could be dramatically changed. Of course it would not change anything with regard to semantical thinking in mathematics, but it *would* change the way we look at the possibilities of formal systems. The whole point of semantical arguments, and the subsequent conclusion of the difference between truth and proof, is based on formal systems not being able to contain their own truth predicates. An adequate formal definition of proof within the object language would obviously change all that.

The reason we have been so complacent in requiring expansions to formal systems is of course that Tarski (1936) *proved* that classical first-order formal languages could not contain their own truth predicates. First-order logic is complete, as Gödel had proved earlier. By adding a truth predicate to a first-order language we can formulate the liar's paradox[115], which is in conflict with the law of excluded middle. Hence, including a truth predicate implies the need for a metalanguage, just like it was used in the semantical argument. This seems unproblematic to us since the use of first-order classical languages is so deeply entrenched in the practice of mathematics, especially when it comes to the

[115] "This sentence is false". Obviously the first-order language in question must have enough expressive power to be able to express the liar sentence. Clearly we can establish neither the truth nor the falsity of the liar sentence, or at least one of its more sophisticated variants. There is of course a close connection between the liar's paradox and the Gödel sentences, the main difference being that the latter are in fact not paradoxical but true.

question of the most basic mathematical theories. But are we in any way justified in making such limitations when we consider the philosophical question of truth and proof? We do, after all, have knowledge of many other useful formal languages in mathematics. If a change of the language can affect the semantical argument, we should certainly be aware of that.

The basic position since Tarski has been that a hierarchy of languages (and formal systems) is needed for adequate definitions of truth. We can define an adequate truth predicate for a formal language *L*, but it must be done in a metalanguage *M* of *L*. Obviously *M* cannot contain its own truth predicate either, and therefore another metalanguage is needed. Such is the hierarchy of Tarskian truth, and if we limit ourselves to purely formal languages, it does not collapse at any point. If we accept classical first-order logic, and want to include a truth predicate, that kind of infinite hierarchy of formal languages is inescapably required due to Tarski's undefinability result. Even if we use the strategy of this work, that is, introducing an informal metalanguage to collapse the hierarchy, we are still stuck with a minimum of two languages: a formal first-order object language and its partly informal metalanguage. This is, however, a somewhat problematic conclusion as far as its wider philosophical implications are considered. After all, the concepts of metalanguage and object language are in no way basic in the philosophy of language. Indeed, is it not the case that we all have *one* language within which we define truth, and enumerate the true sentences?

Granted, Tarski's undefinability theorem concerns formal languages, and our natural languages are always going to be informal. Obviously there is nothing in Tarski's work that prevents us from defining truth within informal languages. But it must be remembered that before Tarski's result, the monolingual approach was accepted also in mathematics, and it undeniably carries a certain amount of attraction in it. For the strict formalist it can be absolutely crucial, since we have seen how problematic the Tarskian hierarchy is in extreme formalism. If we run into an infinite regression of languages, the formalist program seems untenable. When we include pre-formal mathematical thinking, we do not have the problem of infinite regression in the same way,

since our metalanguage can be (and of course actually is, at some point) informal. But even for us non-formalists the hierarchy of formal languages can sound like an unnecessary awkwardness. Is it really the case that there cannot be just *one* formal language of mathematics? As Jaakko Hintikka (2006, p. 708) points out, the notion of metalanguage has been "forced on" us, it is not something that the philosophers and logicians have desired. When we think about the practice of mathematics, there is all the more motivation to avoid hierarchies. Outside of logic, hierarchies of languages do not really enter the picture in mathematics. In fact, if it were not for truth predicates, the whole phenomenon of object languages and metalanguages would seem very much undesirable. Tarskian truth may mirror the relation between our formal and pre-formal mathematics, but *within* formal mathematics using a single language has obvious advantages over the infinite hierarchies.

The basis for such a *monolingual* formalist project must lie in a different choice of logic. Classical first-order two-valued logic will not do if we want to include adequate truth definitions within formal languages – and that is indeed what we must do if we want to defeat the Tarskian hierarchy. For the choice of an alternative logic, three potential solutions have been proposed. The first, and the most famous one, is due to Saul Kripke (1975) and his use of the many-valued Kleenean logic. The second option is Hintikka's (1996) and Gabriel Sandu's Independence Friendly (IF) logic, a first-order many-valued logic.[116] The third option is staying within the realm of classical logic, but expanding into *second*-order logic.

5.2 Hintikka's truth

Let us now mix the chronology a bit and start from the most recent development, started by Jaakko Hintikka in his book *Principles of*

[116] Hintikka does not use this classification, but at least for the purposes of this work IF is a many-valued logic, even if it must be distinguished from such many-valued concepts as truth-degrees.

Mathematics Revisited (1996).[117] Hintikka has argued that his Independence Friendly logic not only contains its own truth predicate, but this predicate is also adequate – all this in a first-order language. IF-logic has been much discussed recently, and it is not possible to go into all its general merits and problems here. What we are concerned with here is the question of truth and proof, in particular the method of defining a truth predicate for an IF-language *L* within *L*. Here Hintikka's approach is revolutionary: the definition of truth for IF is a *semantic* one, based on game-theoretic semantics (GTS).

In GTS the truth-value of a complex sentence *S* is defined by the "verifier" Eloise trying to show that the sentence is true and her opponent "falsifier" Abelard trying to show that the sentence is false. The way this happens is, in propositional logic, that Eloise picks branches of disjunctions and Abelard branches of conjunctions. In predicate logic, in addition to that, Abelard picks values for universally quantified variables, and Eloise values for existentially quantified variables. In the case of negations in the formulas, the roles are reversed. The game goes on until an atomic sentence *A* is reached. If the sentence *A* is true, then Eloise is said to have a winning strategy, and the original sentence *S* is true. If *A* is false, then Abelard has a winning strategy, and the sentence *S* is false.[118]

As Hintikka (ibid., pp. 25-26) points out, there is an apparent circularity in using the above GTS as a definition of truth since it contains reference to the truth and falsity of atomic sentences. However, it must be pointed out that Hintikka is concerned with semantic truth, that is, truth in *models*. GTS come into play *after* we have an interpretation of a first-order language, that is, when we move from an uninterpreted language into a model of it. That way, the truth-values of atomic sentences come with the interpretation. That is the first stage of truth in GTS. The other stage is how we arrive from the truth and falsity of the atomic sentences at the truth

---

[117] Hintikka's book is more of a philosophical introduction on the subject and not really suitable as a textbook for learning IF logic. The recommended, and only, textbook on the subject is Väänänen 2007.

[118] For more details, see Hintikka 1996, pp. 24-26.

and falsity of complex sentences. Those rules are defined by GTS, and one must notice the similarity to the Tarskian definition of truth.[119]

As we know, however, Tarskian truth in the classical first-order logic requires metalanguages, while Hintikka wants to avoid them. For that purpose, it helps that IF is a richer logic than the classical first-order one – not by much, but with one crucial difference. The extra strength is carried by IF being able to express the independence of quantifiers from each other, and from free variables.[120] While in first-order logic the existence of self-referential "liar sentences" causes paradoxes that conflict with the basic laws of logic, in IF languages we have more expressive power to deal with them. Since truth is defined as the winning strategies in semantical games, the undecidable liar sentences, with the help of independence, are thought of as infinite symmetrical loops where neither player can win. This is the third "truth-value" in IF logic. With it the user of IF can avoid the liar's paradox (see ibid., pp. 159-160), and thus Tarski's proof of the undefinability of truth does not apply. [121] In fact, only the truth conditions given in game-theoretic semantics apply, and they *are* the definition of truth.

Moreover, those conditions are presentable in the same first-order IF language. This Hintikka does with the help of Gödel-numbering and Church's thesis, the widely held belief according to which (to be precise, one corollary of it, see ibid., p. 114) all

---

[119] This hidden reference to truth can be considered a weakness in Hintikka's theory, but since I am out to defend Tarskian truth, I do not hold that to be the case. In this work we are concerned with interpreted languages, so there should not be any problem in this sense. The use of model theory, however, will be seen to be problematic for Hintikka's approach.

[120] This is equivalent with the ability to express *Henkin quantifiers*. Henkin quantifier over a sentence φ is usually written in literature either as $Hxyx'y'\varphi$ or $\begin{pmatrix} \forall x & \exists y \\ \forall x' & \exists y' \end{pmatrix}\varphi$ and it is defined as $\exists f \exists g \forall x \forall x' (x, f(x), x', g(x'))\varphi$. Importantly, the selection of y depends only on x, and the selection of y' depends only on x'.

[121] See, for example, Väänänen 2007, pp. 102-103.

mechanically decidable relations are representable in basic arithmetic. Since the rules of GTS are quite clearly mechanically decidable and IF can express basic arithmetic, according to Hintikka, an IF language can contain the definition of its own truth predicate. In other words, in a model of an IF language, the rules of GTS give us the extension of a truth predicate *Tr*. Since the truth conditions are expressible in IF languages, an IF language can contain its own adequate truth predicate. This is why Hintikka has claimed that he has "exorcised Tarski's curse"; that we do not need to commit to a hierarchy of languages in order to give definitions of truth for formal languages.

However, Philippe De Rouilhan and Serge Bozon (2003) have argued that Hintikka has not quite achieved everything he claims to have done. While they agree that we can define an adequate truth predicate *Tr* for an IF-language *L* within *L*, they also point out that we cannot escape the hierarchy of languages if we hope to *recognize and use Tr* as the truth predicate of *L*. Simply put, in an IF language *L* there is a predicate *Tr* which has the extension of the true sentences of *L*, but we can only *show* this to be the case in a language richer than *L*. The technicalities concerning IF can be found in De Rouilhan's and Bozon's article (ibid., pp. 689-698), but the crux of their criticism is that for Hintikka's claim to hold, it would need to be the case that a *completely monolingual* speaker of *L* would be able to carry out the whole process involved in defining truth for *L*. This would mean, in addition to providing an adequate definition of truth for *L*, also coming up with the idea in the first place, expressing the concept of adequacy, having the tools to express the quasi-logical equivalences in T-sentences and having enough model theory and arithmetic to carry out the whole process. Most importantly, even if we did have at our disposal an adequate definition of truth for *L*, we would have to be able to *show* it to be such – that all the T-sentences of *L* are the logical conclusion of the truth definition – and do this totally within *L*. In short, De Rouilhan and Bozon argue, Hintikka manages to exorcise Tarski's curse only if his whole project could be achieved within an IF-language.

From Hintikka's and Sandu's work we do know that there is an adequate GTS truth predicate within *L*, but we know this while

operating in a metalanguage. To start with, Hintikka's book is not written in the language of IF logic. Model-theory, in particular, is used throughout Hintikka's work, but it is not expressed in IF logic. Hintikka also uses Gödel-numbering, and hence arithmetic, widely in his project, yet – as Jan Woleński (2006, p. 665) points out – it is not at all clear that in IF languages one can do all that is needed for arithmetic. Mathematical induction, for example, is not equivalent to any rule of GTS. In addition, there is the missing law of excluded middle to deal with. Truth predicates are a special case, but do we know that with IF logic we can define arithmetic in an otherwise equivalent way, or at least equivalent enough to go through all that is needed for the Gödel-numbering? It must be remembered that Hintikka has not actually defined model theory or arithmetic within IF languages.

Finally, needless to say, first-order IF languages do not contain enough expressive power to carry out the sort of theorizing that Hintikka does in natural language. In this respect, De Rouilhan's and Bozon's criticism resembles my criticism of Field. It is all well to use axioms to define concepts in formal languages, but we should not forget that all this is initially done in our pre-formal languages. When we switch from pre-formal to formal concepts, we cannot fool ourselves into thinking that the former never existed. Moreover, we cannot forget that they are the reason why the formal concepts are possible in the first place. Similarly, Hintikka's game-theoretic definition of truth for IF languages is adequate, but we cannot look at this result independently from all the background needed to establish it. De Rouilhan and Bozon argue that a first-order IF language itself *is not able to express this*, or any of the desired results around it. That is why Tarski's hierarchy exists also in IF. For us to be able to say that the definition of truth for IF is adequate, we need to have a richer language. The main problem for Hintikka (as well as for others that want to escape the Tarskian hierarchy) is that *everything* needs to be done in a single language. De Rouilhan and Bozon call this the *monolingual speaker problem*, and it follows us wherever we go with projects like Hintikka's.

The problem of the monolingual speaker in its many facets is the most important difficulty with Hintikka's claim of exorcising

Tarski's curse. Aside from the more general considerations, De Rouilhan and Bozon (ibid., p. 697) also offer a technical argument, culminating in the conclusion that the model-theoretic concepts of logical truth, logical implication and logical equivalence are all definable in IF languages only in a very weak sense. Indeed, as they point out, here IF loses something from classical first-order languages where those concepts are fully definable. The difference comes from IF-languages having more expressive power than the classical first-order languages. As such, IF languages are inevitably incomplete.[122] In addition, IF languages do not have contradictory negation. Both of these are properties which classical first-order languages possess. The excluded middle does not hold in IF, and thus we have the "truth-value" of *undecidable* to deal with. Obviously truth and falsity in IF do not behave as they do with contradictory negation. The problem for Hintikka, De Rouilhan and Bozon argue (ibid., p. 698), is that with these weak notions of logical truth, logical implication and logical equivalence, the monolingual speaker of an IF language cannot express an adequate definition of the truth predicate within *L*. In defining and showing that his GTS truth is adequate, Hintikka (1996, pp. 24-25, for example) uses quite a bit of model theory. The monolingual IF speaker, however, could only use very weak versions of the central model-theoretic concepts. Hence, we need stronger model theory to show *Tr* to be the truth predicate of *L*, and Tarski's hierarchy is still with us.

Hintikka (2006, p. 710) claims that this is not a problem, but his arguments are not ultimately very convincing. First of all, he dismisses the talk of logical truth and such as modal notions, and not relevant to the question of Tarskian hierarchies. Hence, according to Hintikka, the technical result of De Rouilhan and Bozon concerning concepts like logical truth does not show the kind of incompleteness they thought they had. In addition, Hintikka suggests that in order to discuss concepts such as logical truth we can enrich the IF language with contradictory negation.

---

[122] Incompleteness follows right away from the ability to express Henkin quantifiers. See Hintikka & Sandu 1996, p. 177 for details.

The way to do this is by enumerating all the contradictory sentences (inconsistent formulas) and including an axiom that "all contradictory sentences are not true". This can be done since the set of inconsistent formulas in an IF language is recursively enumerable (see Hintikka & Sandu 1996, p. 177). Now in this fragment of an IF language we can discuss logical truth in a new context, Hintikka claims (2006, p. 711), where there is no need for a separate metalanguage.[123]

Be that as it may, I think Hintikka is too quick to dismiss De Rouilhan's and Bozon's results concerning model-theoretic concepts as modal notions and not relevant to IF. It is the same kind of undefinability that makes it impossible for us to use the truth predicate of an IF language *L* as such within *L*. Whether one agrees with that particular counterargument or not, the general argument against Hintikka's claims of exorcising Tarski's curse remains strong. Talk of modal notions aside, there seems to be a lot behind an IF language that is not done in that language – and certainly model theory, alongside arithmetic, seems like the most important of these. Hintikka (ibid., p. 707) has responded to De Rouilhan and Bozon by arguing that he used terms outside IF logic merely for the sake of intelligibility. The syntax of *L* can be (with the help of Gödel numbering) expressed in *L*, and hence the truth predicate for *L* can also be formulated. Whatever other languages (model-theoretic and second-order logic, for example) were used, it should not matter, as the same could have been done without them. The truth in a first-order IF language *L could* be formulated within *L*, and that is what matters, Hintikka argues. With Gödel

[123] One unsatisfactory part of this strategy is that the contradictory statements are now false. Why false rather than true? The situation no longer seems to correspond to our intuitions about liar sentences. The liar sentence, unlike the Gödel sentence, gives us no reason to think that it is true rather than false. This is the strength of assigning the truth-value "undecidable" to such fixed points of the language. The liar sentence will get the same truth-value as its negation, which sounds intuitively satisfactory. So with Hintikka's proposed strategy we gain some of the completeness of classical first-order languages, but with the price of losing a lot that is of value in IF languages. This is also the strategy suggested by Kripke (1975, p. 715) in order to avoid truth-value gaps.

numbering we can translate the other languages into IF, and so the problem disappears.

However, one important part of the monolingual problem is to show that all this can be *actually* done. Hintikka has been content to assume that this is the case since IF sentences are equivalent to $\sum^1_1$– formulas, which are second-order existential formulas of the type $\exists f_1 \ldots \exists f_n \varphi$, where φ is a first-order formula. In $\sum^1_1$– languages we can do all the required model theory and arithmetic, and hence – Hintikka claims – there is no need to go through the details in IF languages. But what is the basis of stating that for all IF sentences there are equivalent $\sum^1_1$– formulas? That indeed is the case (see Väänänen 2007, pp. 86-90), but to avoid the monolingual problem that equivalency would have to be shown *completely within an IF language*. There is an evident logical difficulty in defining IF as the basic logic, and then using non-IF model theory and arithmetic to show connections between IF and other languages.

Overall, Hintikka skips an important step. We cannot be satisfied with simply avoiding Tarski's undefinability result. This Hintikka has achieved. An IF language *L* can contain its own adequate truth predicate *Tr*; or at least there is nothing to make that impossible. But in order for us to use *Tr* as the truth predicate of *L*, we must be able to show that *Tr* is materially adequate; that all the T-instances of *L* are the logical conclusion of the truth predicate of *L*. And for this we need a metalanguage, as De Rouilhan and Bozon pointed out. While the truth predicate *Tr* for *L* could be formulated in *L*, we cannot show *Tr* to be the truth predicate in *L* alone. Hintikka's reply concerns formulation of the truth predicate, but the more pertinent question here is showing that the truth predicate *is* in fact the truth predicate. For this purpose, the hierarchy of language must be brought in through the back door.

However, while I agree with the criticism that we need a richer metalanguage to be able to recognize the definition of truth in *L* as such, it can be somewhat confusing to call it a hierarchy of the *Tarskian* type. The reason for this is that we arrive at the Tarskian hierarchy in a different way. Tarski proved that such a hierarchy is needed in classical first-order languages, and at the same time he

also showed *why* it is needed: because of the liar's paradox. The reason for needing a metalanguage in order to recognize that *Tr* is the truth predicate of an IF language *L* is different. While a classical first-order language cannot even contain an adequate truth predicate, an IF language can. The problem is that an IF language *L* itself does not have enough expressive force to carry out everything we need in order to show that *Tr* is the truth predicate of *L*. We cannot escape the need for a metalanguage, and hence the hierarchies of languages. But this difference in the basis for hierarchies is important for us to acknowledge. Ultimately, I think this is the greatest problem with Hintikka claiming that he has exorcised Tarski's curse. His result of avoiding hierarchy in the usual Tarskian way makes him confident he has avoided hierarchies in all ways.[124] But as we have seen, this is not the case.

5.3 Why IF logic?

Above are the main problems in the case we accept IF as our basic logic. For us to show that the truth predicate of IF is materially adequate – and for us to use it – we must use some metalanguage. Essentially, Hintikka's project demands first-order languages rich enough in arithmetic to carry out the Gödel-numberings, plus rich enough in model theory to carry out all the necessary theorizing. Granted, a lot of this is immediately translatable into IF languages. But not everything is, nor is translatability by itself enough, and hence the monolingual speaker problem remains unsolved. Simply put, to include and recognize a truth predicate in an IF language, we need to expand beyond that language. The situation with truth is not essentially different from classical first-order logic and the parts of set theory that the Tarskian expansion demands.[125]

---

[124] This is not to say that there is any factual difference between the types of hierarchies of languages needed. But the arguments for *arriving* at the hierarchies are different, which is the important point here.

[125] Hintikka (1996, p. 29) mentions other problems of Tarski-type truth, such as infinitely deep languages. In this work, however, I must limit myself to standard languages and standard models. In any case, in our

However, it is another question whether we are prepared to adopt IF logic in the first place.[126] If Hintikka were correct, in order to have an adequate definition of truth we would have two choices, provided that we stay within first-order logic. First, we can stick to the classical first-order logic. In that case we have to expand outside logic, most appealingly to set theory and Tarskian truth.[127] Second, we can use IF, which is still a first-order logic.[128] Now what are the main differences between classical logic expanded with set theory and IF logic?[129] In particular, what is the motivation for replacing classical logic with an IF one? Aside from the added expressive power to logic, I think that the key to Hintikka's thinking is that he is very much committed to the logicist tradition of Russell. He believes that mathematics should be expressible in the language of logic. If we use set theory, we are immediately outside the logicist program, and commit ourselves to higher-order entities such as sets and relations. Of course one main motivation for logicism is its economical character ontologically: if one only needs to assume the domain of logic, in particular a first-

---

pre-formal thinking we should have enough tools to distinguish between finitely and infinitely deep languages. Hintikka's other objections (1996, pp. 110-113) to Tarski-type truth are based on its difficulties as the truth of IF, which is not the subject matter here.

[126] I understand that this is the most important question for Hintikka, as well as most of his critics. For Hintikka they key point is that he considers IF to be the *correct* logic. For his critics the step from classical two-valued logic to a many-valued one is never one easily taken. I must restrain from taking part into that discussion here, and focus only on the problem of proof and truth.

[127] It must be remembered that we are talking about quite a little set theory here, basically enough to carry out the $\forall x(\mathrm{Pr}(x) \rightarrow Tr(x))$-type sentences in the metasystem of Tarskian truth.

[128] This has also been contested. See Feferman (2006) for an example.

[129] Of course my position is that because of the monolingual speaker problem IF logic must be expanded with set theory or other richer theory, as well, and Hintikka's truth fails already in that.

order one, we greatly reduce our conceptual (and ontological) commitments.[130]

Before we go deeper into that, we must distinguish between two kinds of logicism. While Whitehead and Russell were after a comprehensive program of deriving mathematics from logical truths, Hintikka's logicism is of a considerably weaker type, as becomes evident already from his style of exposition. Hintikka (as well as Whitehead and Russell, of course) is after *descriptive completeness* for logical languages: that the formulas of mathematics could be presented also as equivalent logical formulas.[131] This is naturally a factor behind the program of having a logic with more expressive power than the classical first-order one. We know that classical first-order logic will not suffice.

However, there is also another difference between logicism of the Russell type and that of Hintikka – one not in Hintikka's favour. Logic for Russell and Frege was the logic of old, that is, the classical two-valued predicate logic that is still the basis for most mathematics – almost all outside the field of mathematical logic. Essentially, it was also the logic for centuries (millennia, in fact) before Frege, and a logic that, in the first-order case, carries many desirable qualities, such as completeness. It was thus very natural to try to base mathematics wholly on logic – it was basing mathematics on something that was almost universally agreed upon.[132] But Hintikka's logic is different in this sense. It is a logic

---

[130] See Hintikka 1996, pp. 183-210 for his views on logicism and the potential of IF in such projects. In that chapter it becomes obvious that even if successful, Hintikka's IF logic, not being completely axiomatizable, is hardly suitable for extreme formalism and deflationist truth in Field's sense.

[131] In Agustín Rayo's (2005) division between different types of logicism, Hintikka's is a mixture of *language-logicism* and semantic *truth-logicism* while Russell's is probably best understood as a combination of language-logicism and *consequence-logicism*.

[132] Of course this did not succeed, and Russell had to switch into type theory, which is a higher-order logic. In addition, to present such central set-theoretic concepts as equicardinality one needs second-order logic (or first-order set theory).

developed with the new (descriptive) logicist program in mind. It is *not* the logic that is used in mathematics – the main difference obviously being the lack of the law of excluded middle. It is not complete, and it is not compositional (the truth-value of a complex sentence does not follow syntactically from the truth-values of its components[133]). While the logicist program of Russell can be thought to have an advantage over, say, a set theoretic foundation of mathematics, it is not clear that Hintikka's program retains this advantage. As far as the conceptual commitments are considered, IF may be more economical than set theory.[134] But as far as intelligibility and the practice of mathematics are considered, it looks much more problematic.

That prompts the question: why should we care only about the conceptual commitments? The traditional line of thinking is that conceptual commitments equal ontological commitments, following one form of Occam's razor. But as far as the philosophy of mathematics in general is considered, this is a highly Platonist line of thinking. Postulating an object such as a set does not need to imply that we believe in its objective existence. In the Platonist sense, set theory can seem less economical than IF logic and thus get chopped off by Occam. But in the sense of us *learning and practising* mathematics, set theory (added to classical first-order logic) seems more intelligible. If we reject Platonism, it is not clear that logicist-based mathematical systems are any more ontologically economical than set theoretic ones. In such a situation, the intelligibility of mathematical theories seems like an important criterion – albeit one that has been often completely neglected in the philosophy of mathematics.

---

[133] See Hintikka 1996, pp. 106-112. The lack of compositionality should not necessarily be considered a *weakness* of IF languages, but it is definitely not something that many mathematicians would be ready to accept in their basic logic.

[134] Even this is not at all clear when we consider the problem of establishing the truth predicate of IF languages as such. Feferman (2006) argues that this is equivalent to full second-order logic, which has considerable ontological commitments, as will be seen in the next chapter.

This is not to say that set theory is without any problems. The status of the axiom of choice (AC) is the most obvious one. Hintikka, for one, is extremely reluctant to allow set theory into model theory, and that way into the theory of truth. That was the whole motivation behind truth predicates for IF-logic: to have a first-order language with expressive power traditionally only available to second-order logic, and thus avoid sets and such mathematical concepts. But is Hintikka exaggerating the problems of set theory? The assumption of set existence seems to carry the main initial trouble for Hintikka (1996, p. 19). As set theory – like second-order logic and IF – lacks completeness, he seems to think that set existence forms insurmountable problems as the foundation of mathematics. For a non-Platonist philosopher, this might seem a bit odd. After all, where do we need the assumption that sets exist objectively? Could we not just assume that set theory is about *possible* collections of elements? In fact, this is the argument presented by philosophers like Charles Chihara, more of whom in Chapter 6.5. We will see that Chihara's line of thinking is not without its problems, but the main idea seems agreeable enough: set existence does not need to be taken literally. Depending on the axiomatization of set theory (mainly concerning whether or not we accept the axiom of choice), we get explicit rules concerning the possible collections of elements. In this way, the question is not "do sets exist?", but rather "if elements exist, can we collect them into sets?" Ontologically, this seems much less problematic. It is not formalism, either: it simply moves the ontological burden from the sets to the elements. If we choose such an ontologically weaker philosophy of mathematics, the motivation for logicism is much harder to see.

However, even if we reject this line of thinking, Hintikka's position seems to imply that we need *full* set theory with the axiom of choice in order to have a Tarskian definition of truth.[135] This is not the case. The axiom of choice is not used in Tarski's definition. What we need are essentially the set theoretic relations required to handle "the set of true sentences". These are nothing more than the

[135] See Hintikka 1996, p. 19 where the "checkered history of axiom of choice" is mentioned as an example against the intuitiveness of set theory.

usual relations of set membership, inclusion, intersection, etc. The problems with axiom of choice of course exist, that must not be forgotten. Results like the Banach-Tarski paradox are a big obstacle to choosing ZFC as our meta-theory.[136] On the other hand, the negation of axiom of choice produces at least as problematic results, and without the axiom of choice (or its negation) these kinds of problems could not be decided at all. I see problems of this type as the basis for Hintikka's distaste for set theory, and it must be said that there is something quite understandable in that. For all its intuitive power, set theory has some potentially troubling aspects to it. But they do not need to come into the definition of truth. What we are after is an adequate definition of truth for PA (or for first-order logic), and for that purpose we do not need the axiom of choice. The downside is that now our metalanguage is not uniform: it is first-order logic expanded with parts of set theory. To the logicist mind this of course does not look pretty. But if we are concerned with mathematical truth, it should not be overwhelmingly problematic.

Above are considerations of set theory and the axiom of choice mainly from the classical point of view. Interestingly, the status of the axiom of choice has some important consequences in IF logic. Thomas Forster (2006) has shown that the introduction of axiom of choice actually has an effect on the set of valid IF sentences, while it does not affect the set of valid classical first-order sentences. His article concerns deterministic and nondeterministic strategies in the semantic games. In Hintikka's (Skolem-) semantics for IF-logic the strategies are assumed to be deterministic. If we hold on to that assumption, there is no problem. But Forster points out that if we accept nondeterministic strategies, the axiom of choice will have an

[136] In set theoretic geometry it follows from the axiom of choice that in three-dimensional space we can take a solid ball, break it into non-overlapping pieces, and proceed to form two balls equal to the original ball. Although this is called the Banach-Tarski *paradox*, it is not really a paradox, but rather an extremely unintuitive result. As such, however, it does work like a paradox, intuitiveness being a strong argument for set theory.

impact on the set of valid IF sentences. This becomes evident from a simple IF sentence:

$$(*) \quad \forall x \exists y R(x,y) \rightarrow \begin{pmatrix} \forall x & \exists y \\ \forall x' & \exists y' \end{pmatrix} ((x = x' \rightarrow y = y') \wedge R(x,y) \wedge R(x',y')$$

Now for the existential player Eloise to win the conditional (that is, for the sentence to be true), she clearly must have a winning strategy for the consequent every time she has a winning strategy for the antecedent. At this point the limitation to deterministic strategies has an impact. If we stick to deterministic strategies, then Eloise wins the antecedent only if there is a choice function that determines a choice of y for all x. If there is such a choice function, then Eloise also wins the consequent. So (*) is valid. However, if we allow nondeterministic strategies, the situation changes. If the interpretation of R has a choice function in a model $\boldsymbol{M}$, then the consequent is true whenever the antecedent is. Now the interpretation of R can be true in $\boldsymbol{M}$ *without* a choice function if we allow nondeterministic strategies. But without a choice function for R Eloise loses the consequent, and (*) is not true. Thus, if we accept nondeterministic strategies in the semantic games, there are IF sentences that are valid if and only if we accept the axiom of choice.

So there are two choices we must make in IF logic: first, whether we limit ourselves to deterministic strategies and if we do, second, whether we accept the axiom of choice or not. These choices could not be of greater importance: after all, they actually alter the set of valid IF sentences. There is a big difference between the classical first-order languages and IF languages here. The set of valid classical first-order sentences does not change with the introduction of AC. The set of valid IF sentences does. While classical first-order logic makes no commitment to set existence, by depending on the axiom of choice IF logic is more sensitive to such questions of ontological commitments. Since we do not need AC in order to use Tarskian truth, this difference also has a relevance to the question of truth predicates. Here Hintikka has the burden of

accepting AC (or not) to be able to have a truth predicate. This clearly adds to the problems of Hintikka's truth.

In fact, the above conclusion holds whether we limit ourselves to deterministic strategies or not. Although AC has the dramatic effect we have just seen if we accept nondeterministic strategies, it should be noted that the whole *concept* of deterministic semantic games requires the axiom of choice. Hintikka's own approach (1996, p. 32; pp. 41-42) has been limited to deterministic strategies and he uses the axiom of choice in his game-theoretic semantics. Essentially, to account for the universal quantifiers in the T-scheme[137], Hintikka needs to include an infinite series of determinate choices for the existential player. But this is equivalent to the axiom of choice. Indeed, if we do not introduce the axiom of choice, the T-scheme and Hintikka's GTS are not equivalent. Ironically, it seems to turn out that it is *Hintikka*, and not the Tarskian theoretician of truth, who needs the axiom of choice.[138] Hintikka's (ibid., p. 42) response to this is that the evidence for GTS is at the same time evidence for the axiom of choice in IF. This seems agreeable enough: the intuitive notion of making choices in the semantic game *ad infinitum* is not problematic as long as they are deterministic. However, it sounds intuitively exactly as unproblematic as the axiom of choice in set theory. There does not seem to be any reason to accept one and reject the other. In any case, when it comes to the question of truth, if AC is a problem, it is that mostly for Hintikka and not for Tarski.

We have seen that Hintikka does not succeed in exorcising Tarski's curse of the hierarchy of languages. But even if he did, should we abandon Tarskian truth in favour of that of IF? As far as

---

[137] After all, what Hintikka is doing is formulating a definition of truth extensionally equivalent to Tarski's (when it comes to first-order languages). (see Hintikka 1996, p. 32)

[138] This holds only when it comes to truth. Of course the first-order logicist will need the axiom of choice when he moves to full set theory and beyond. In any approach like mine in this work, the axiom of choice is introduced via it being an important part of mathematical practice.

the status of truth[139] is concerned, Hintikka's program is interesting, but it is that mainly from a (one kind of) logicist point of view. If we do not cherish the premise that all mathematics must be expressible in a language of logic, the motivation for IF is not at all obvious. I do not see any serious problems in applying mathematical notions, such as those of set theory, when we are talking about mathematical truth. The great ease that we have with set theory and classical first-order languages should not be dismissed, either. These are the basic tools for almost all mathematicians. When the principles of mathematics are concerned, this is not without significance. Combined with the knowledge that Hintikka's truth needs metalanguages, as well, I believe that we can safely continue to prefer Tarskian truth as the theory of truth for mathematics.

5.4 Second-order logic

Returning to two-valued logics, what prospects does second-order logic have when it comes to the question of truth? Outside such obvious properties as completeness, one main reason to prefer first-order logic to a second-order one is its apparent ontologically economical nature. As Hintikka (1996, p. 7) has pointed out, first-order logic is basically the logic of a nominalist. One is only committed to the domain of elements, while relations, functions and sets are merely ways of describing possible collections of the elements. There are no sentences of first-order logic stating that "there exists a relation/function/set". Thus we are restricted to an ontologically economical way of referring to mathematical objects. That is, of course, if we do not expand the first-order languages to include, say, set theory – which we indeed must do in order to include a truth predicate, or to do any interesting mathematics. Classical first-order languages, in addition to not being able to

[139] It must be stressed that here I am only concerned about Hintikka's contribution to the question of truth. This is only one aspect in the whole question of IF logic – the most important one being IF having more expressive power than the classical first-order logic.

contain their own truth predicates, also lack the tools of describing such essential mathematical properties as equicardinality. For that, either second-order logic or set theory is needed. We already glanced above at the ontological status of set theory. Against a somewhat common opinion in the philosophy of mathematics, I am not at all convinced that we should by any means possible try to avoid set theory. In fact, I am quite ready to commit to the opposite: if there is no real reason to reject quantification over sets, we should not try to get rid of set theory. The reason for this is simple: even with its problems, set theory is highly intuitive, as well as highly expressive. The intuitiveness is an asset for set theory when compared to IF, but as far as second-order logic is concerned, set theory does not carry such an advantage. With second-order logic we can do mathematics just as with first-order set theory, and the required constructions resemble closely the intuitive steps we make in forming collections and relations.

While IF-logic is still a developing and somewhat under-studied subject, second-order logic has been discussed widely. As we know, first-order logic requires (at least parts of) set theory in order to provide a basis for arithmetic. The problematic status of sets as mathematical objects is thus not possible to avoid in classical first-order languages. Second-order logic features quantification over properties, relations and functions, which give us the tools to define all the necessary arithmetical concepts without the need for set theoretic concepts. In addition, second-order logic provides the tools to do set theory itself, and hence also all the required model-theory and proof theory. Tentatively put, with second-order logic we seem to have all the tools needed for mathematics.[140] Compared to Hintikka, we in particular have the liberty of using as much set theory as we want.

Of course second-order logic features its share of problems, as well – otherwise we would not be discussing any other options. One of the major difficulties is the incompleteness of second-order logic. This is a major disadvantage compared to first-order logic, but not compared to IF. If we wish to stay clear of set theory, our

[140] See Shapiro 2005 for a good basic outline of higher-order logics as the basis for mathematics.

choices here seem to be IF (or other many-valued logic) and second-order logic. What ontological, logical or practical reasons can we have for choosing between them? More to the topic at hand, what do these reasons have to do with truth, and how do they compare to the Tarskian construction?

As far as the mathematical practice goes, the question is simple: mathematicians will use set theory when it is applicable and second-order logic when higher-order quantification is needed. Neither is considered particularly problematic in most mathematical practice. For philosophical reasons the question is more relevant. The main motivation for a second-order foundation of mathematics is essentially logicist. Terms of set theory, like the membership relation $\in$, are very useful, and in practice indispensable. Thus, the motivation for the second-order foundationalist is not to abolish set theory, but rather to provide a logical basis for it and other mathematical theories. Second-order logic, the second-order logicist argument goes, has the tools to define set theory in purely logical terms.

This approach has a few problems, even though it initially seems very appealing. The first problem concerns the notion of logical consequence. Classical first-order logic is complete, which gives us the neat feature that provability, derivability and logical consequence are all equivalent concepts. Once we move via set theory into arithmetic, we lose completeness.[141] Because of the incompleteness, we can use semantic arguments against deflationism of truth, and the classical first-order deflationist is beaten by the semantical argument. But the deflationist can seek refuge from this non-conservativeness of truth in second-order logic. We remember that one crucial problem for Field's deflationism was its inability to use mathematical induction over sentences concerning the truth predicate. Second-order logic, however, has the tools to present the induction principle, in fact in a single sentence:

---

[141] It should be noted that, strictly speaking, completeness in first-order logic is not the same property as completeness in arithmetic. However, that does not affect the argument here, and we do not need to go into the details.

$\forall f((f(0) \wedge \forall x(f(x) \rightarrow f(s(x))) \rightarrow \forall x(f(x)))$, where s is the successor operation.

With the induction scheme we can avoid the problems of the deflationist defence presented in Chapter 3.3 against Field. Shapiro (1998, pp. 508-509) notes that once we add truth to the second-order system **S**, we can use induction on it and derive the consistency $Cons_S$ and the Gödel sentence $G(\mathbf{S})$ without any problems. Of course this is not surprising: it is thanks to the mathematical induction in arithmetic that we could use the semantic argument in the first place. Other ways of presenting the induction principle are bound to have the same consequence. Indeed, Tennant's soundness principle was enough for that purpose. However, this by itself does not change anything for the deflationist. $G(\mathbf{S})$ as a sentence of arithmetic is of course unprovable (by Gödel), and so we have simply moved the semantic argument to a second-order framework.

However, there is another option. The key phrase in Shapiro's point above is "once we add truth". Of course truth must mean provability to retain the deflationist argument. In short, it is just the soundness principle $\mathrm{Pr}_s(\overline{\varphi}) \rightarrow \varphi$, except that we use the (now empty) notion of truth to present it as $\mathrm{Pr}_s(\overline{\varphi}) \rightarrow True_s(\overline{\varphi})$. But we know that all true sentences of **S** are provable in a deflationist system, and this would have to include the Gödel sentence. This is the proof-theoretic version of truth, and it causes the semantic argument: it states that truth is not conservative over proof in arithmetic. When dealing with second-order logic, however, we have another option: model-theoretic truth. We skip some of the technical details of Shapiro's argument here (for them, see ibid., p. 509), but basically, it is possible to take a second-order system **Γ** that can express its own syntax, and add to it a truth predicate *Tr* to form a system that has as its consequences all the T-instances (as enumerated by *Tr*). Let us call this new system **Γ′**. Now we can extend all the models of **Γ** to be models of **Γ′**. From the T-sentences we get the extension of *Tr* in the language of **Γ**. Hence, **Γ′** is

semantically conservative over **Γ**. In other words, if φ (in the language of **Γ**) holds in all the models of **Γ'**, it holds in all the models of **Γ**. That is, if φ is a logical consequence of **Γ'**, it is a logical consequence of **Γ.** Adding truth to **Γ** does not give new sentences as logical consequences. For the deflationist this looks like a very important result. Indeed, if we accept all the steps here, it is actually a valid refutation of the substantial notion of truth.

Of course I would not have waited this far if I actually thought that deflationism works in second-order logic. The key here is that in the model-theoretic argument the important concept is that of logical consequence. This deflationist argument relies *entirely* on the use of logical consequence. However, logical consequence in second-order logic is nothing like the neat, complete, equivalent in classical first-order languages. It is an extremely strong concept, in fact a great deal stronger than that of arithmetical truth, which is what we are after here. Shapiro points out (ibid., p. 509):

> There are second-order categorical characterizations of just about every major mathematical structure, including the natural numbers, the real numbers, the first inaccessible rank of the set-theoretic hierarchy, and beyond, well into the hierarchy of large cardinals. Therefore, truth for each of those theories can be reduced to second-order logical consequence. Moreover, there is a second-order sentence that is a logical truth if and only if the continuum hypothesis holds and another second-order sentence that is a logical truth only if the continuum hypothesis fails.

This gives us a clear idea of what we are dealing with. We can indeed deflate arithmetical truth by second-order logical consequence, but by doing so we introduce practically all the problems of mathematical truth and provability that we can imagine. This is a Pyrrhic victory for the deflationist. The notion of logical consequence in second-order logic is too "deep and intractable", as even the proponents of second-order logic like Shapiro (ibid.) and Church agree. In other words, it is just too difficult to adopt here, because it is much stronger than the concept – arithmetical truth – we use it to define with. Truth predicate reduced to the second-order logical consequence would give us the

true sentences of arithmetic, but at the same time it would also solve the truth (or falsehood) of the Continuum hypothesis. Something simpler is definitely needed. While a system of first-order logic is essentially a deductive calculus – the axioms unambiguously determine the system – a system of second-order logic is not, because of the intractability of logical consequence in it. We can achieve the desired results with second-order logical consequence, but we do not really know what *else* we commit to at the same time.[142] All we know is that it is an awful lot, much more than we wish for when discussing arithmetical truth.

In addition, second-order logic runs into problems familiar to us from Hintikka. Tapani Hyttinen and Gabriel Sandu (2004, p. 420) point out that we cannot define the concept of logical consequence in second-order arithmetic within second-order logic. We cannot express things like "is a logical consequence of a second-order theory of arithmetic" in second-order logic. The familiar monolingual argument arises. Can a monolingual speaker of a second-order language do *everything* she needs to deflate truth from the system? At least in Shapiro's second-order logic, this is not even tried. He (2005, p. 771) uses rich set theory including the power set as the meta-theory for second-order logic.[143] Of course Shapiro generally admits the need for metalanguages, so his system is perhaps not representative of those committed to deflationism. Still, it seems hugely problematic, if not downright impossible, to use *just* second-order logic to define second-order logic. It is safe to say, at least, that it has never been done.

There is one final problem with second-order logic. Set theory, when rejected, is most often rejected because it carries the ontological burden of set existence. Especially the power set (the set of all subsets) seems to carry the maximal ontological burden that *all* sets exist. But does second-order logic avoid this problem? Famously, Quine (1986, p. 68) stated that second-order logic is "set theory in disguise", that all the existential problems of set theory

[142] Similar point is made in Jané 2005.

[143] See Shapiro 2000b for a book-length introduction into the subject of second-order logic as the basis for mathematics.

follow us into second-order logic. In second-order logic we can quantify over the basic domain, but also over the functions and relations of that domain. In what way do these second-order objects exist? Shapiro (1998) quotes Church (1956):

> [Second-order consequence] presupposes a certain absolute notion of ALL propositional functions.

Church then defends this notion because it is presupposed also in classical mathematics, especially in analysis. This is an important point, although I probably make a different conclusion about it from the ones that Church would have wished. Church's point is that if we accept second-order consequence, we do not really have any reason to restrict quantification to second-order entities. But when it comes to the ontological question, how can we distinguish between the existence of second-order entities and those of higher orders? It seems that either we decree that only functions over the basic domain exist, or we must admit that *all* functions exist, and make up for a whole hierarchy comparable to the problematic power set of set theory. However, the existence of all propositional functions seems ontologically no less problematic than the existence of sets, including the power set.

Of course sets themselves are second-order entities in second-order logic, which was the idea behind Quine's original point. In this way, there does not seem to be any ontological reason for choosing between set theory and second-order logic. For a final measure, we can claim that only the entities of the basic domain exist, and quantifications over them are not real in the same sense. But this leads us into new trouble with such concepts as arbitrary choice. Hintikka (1996, p. 193) has elaborated on this subject. If we are to use second-order entities mathematically, we must commit to arbitrary sets and functions, as well as arbitrary choices from the basic domain. There could be a way out of this by a notion of higher-order choice of some type. However, at this point one must

ask whether starting from set theory is ontologically any more problematic?[144]

Above we have followed Church's notion that second-order logic means *full* second-order logic. The basic problem in this approach, as we saw, is that we have very little reason not to expand the logic to other higher-order logics, for every *n*:th order. This way quantification over all entities suggests the existence of all functions. Furthermore, the logical consequence relation in such systems becomes very much intractable and extremely strong. I believe that this is correct, but let us still look at what prospects the deflationist has if she rejects Church's notion. Instead of full second-order logic, she can use a subfragment of it. There are various ways of using subfragments of second-order languages, and we can examine one commonly used candidate, the $\sum^1_1$– logic, as the example here.[145] It is a subfragment of second-order logic that is limited to $\sum^1_1$– formulas, which are second-order existential formulas of the type $\exists f_1 ... \exists f_n \varphi$, where φ is a first-order formula. Basically, it is a subfragment of second-order logic restricted to quantification over second-order entities. We escape higher-order quantification, and as a result the logical consequence-relation becomes weaker, and much less problematic. Deflationism and $\sum^1_1$– logic are discussed in detail in Hyttinen & Sandu 2004

---

[144] The choice between *IF logic* and second-order logic is not an easy one to make. One advantage of IF logic is the inclusion of the *interpolation property*, that is, whenever in a system of language *L* it holds that $\phi \rightarrow \psi$, there is some formula θ of *L* in the system for which it holds that $\phi \rightarrow \theta$ and $\theta \rightarrow \psi$. Second-order logic famously lacks this property, which is another example of the intractability of the concept of logical consequence in it. As far as the ontological commitments are concerned, IF logic initially looks more economical – but as we know, this is not necessarily the case when we want to include a truth predicate, and most importantly, show it to be one.

[145] The field of mathematical logic called *reverse mathematics* is concerned with the subsystems of second-order arithmetic. In reverse mathematics one tries to find the axiomatization *required* to prove theorems of mathematics. See Simpson (1999) and Friedman (1976) for more on the subject.

(Chapter 5). They show that $\sum^1_1-$logic has more expressive power than classical first-order logic, including ability to define its own truth predicate in the model theoretic sense (see ibid. 5.5). In the case of arithmetical truth, there exists a $\sum^1_1-$formula $\Phi(x)$ such that for all models *M* of PA and all $\sum^1_1-$sentences φ it holds that: $\Phi(\overline{\varphi})$ is the logical consequence of *M* if and only if φ is the logical consequence of *M*. Clearly $\Phi(x)$ is now the truth predicate of *PA*.

We can benefit here from the knowledge that every IF first-order formula is equivalent to some $\sum^1_1-$formula. Not surprisingly, the problems we face with truth predicates in IF logic will also present themselves in $\sum^1_1-$logic. For starters, just like IF-languages, $\sum^1_1-$languages cannot be closed under contradictory negation. The double-edged consequence of this is that Tarski's theorem of undefinability of truth does not hold, but also that we cannot recognize the truth predicate of $\sum^1_1-$languages as such within those languages. The result is familiar from other logics, as discussed in earlier chapters: we can define the truth predicate for $\sum^1_1-$logic within $\sum^1_1-$logic, but we need a metalanguage in order to *show* that it is the truth predicate. If we wished to escape that, we would need to do all the required model theory within the $\sum^1_1-$languages. This cannot be done (see ibid. Chapter 6). Furthermore, we would face the other facets of the monolingual speaker problem that were presented in the previous chapter. Second-order logic neither in its full nor fragmented form seems to give the deflationist the tools that she needs – or it does, but provides too much additional baggage.

5.5 Kripke's truth and the potential of many-valued logics

Let us now return to the argument that started much of this development. The most famous effort to define a truth predicate for a formal language *L* within *L* is Saul Kripke's "Outline of a theory of truth" in 1975. Obviously Kripke can be no different from

Hintikka in that such a program cannot be accomplished in classical first-order logic. Kripke's solution was to use Kleene's three-valued logic to define a truth predicate. To handle the truth-value gaps (that is, the paradoxes), Kleene's logic has a third-value, which can be called "undecidable". Kripke realized that such undecidable sentences are instances of the fixed-point theorem.[146] By enumerating these fixed-points inductively, Kripke could show that the truth predicate would follow the three-valued logic; the undecidable sentences being neither true nor false.[147]

Kripke showed that such a Kleenean language *L* can indeed contain its own truth predicate. This is the same thing Hintikka and Sandu have shown for IF. But Kripke anticipated the kind of criticism that De Rouilhan and Bozon presented against Hintikka, and consented that *L* cannot contain an adequate *definition* of the predicate. Kripke (ibid., p. 714) admits, for example, that he had to use set-theoretic language in the induction over the fixed points. In other words, *L* can contain its own truth predicate, but we have no way of establishing within *L* that it is indeed the truth predicate. To have a satisfactory truth predicate for *L*, one would obviously expect us to be able to give it an adequate definition. Now the interesting question (also pointed out by De Rouilhan and Bozon) is whether Kripke's and Hintikka's truths are different in this sense? Kripke conceded that Tarski's hierarchy is still with us after his truth predicate. Why would Hintikka's approach be any different?

The obvious difference would seem to be that Hintikka's logic has game-theoretic semantics to contend with the Henkin quantifiers, while Kripke's (Kleene's) does not. Sandu (1996) in the Appendix of Hintikka's *Principles of Mathematics Revisited* shows that a result equivalent to Kripke's can be formulated for IF with the game-theoretic semantics. Because IF languages have more expressive power than standard first-order languages, the result is even stronger than Kripke's, which should make a case for

---

[146] See Chapter 2.6 for Gödel's use of the fixed-point theorem in his proof.
[147] Or, alternatively, they are assigned the truth-value false, if we want the truth predicate to be completely defined (Kripke 1975, p. 715). We already saw the problems of this approach in Chapter 5.2.

Hintikka's logic. Indeed, De Rouilhan and Bozon (2003, pp. 688-689) show that Hintikka's and Sandu's truth predicate is adequate *intensionally* (in all models of IF) as well as extensionally (in the desired model), while Kripke's is adequate only in the latter sense (in a model of Kleenean logic).

Another reason to prefer IF languages to Kripke's approach can be seen from the *supervenience* principle of semantics, as proposed by Michael Kremer (1988). In this line of thinking the sentences that fall under the concept "truth" are determined by, and only by, the interpretation of the non-semantic terms and empirical facts. This seems very reasonable: when we say that "snow is white" is true, it is obviously due to the interpretations of the words "snow" and "white" added to the empirical fact of snow's colour. But in Kripke's system this will cause a problem when we think of the truth-teller's sentence:

(TT) = "(TT) is true"

In Kripke's Kleenean logic (TT) gets one of the truth-values *true*, *false or undecidable*. An obvious feature of such three-valued semantics must be that the truth-value V of any sentence φ is the same as that of an interpretation of its theory of truth $\mathrm{Tr}(\varphi)$. That is, for every model there exists an interpretation such that $V(\varphi) = V(\mathrm{Tr}(\varphi))$. However, for (TT) there are *three* such interpretations, one for each possible truth-value of (TT). The truth-value of (TT) is *not* determined only by the interpretation of non-semantical terms and empirical facts, and hence the supervenience of semantics is violated. This point was used by Anil Gupta and Noel Belnap (1993) to argue for their *Revisionist Theory of Truth* (RTT)[148], but it also marks a difference between Kleenean logic and IF. The troubling part with Kleenean logic is the interpretation that (TT) is undecidable, that is, neither true nor false. This goes against the intuition we have about (TT), which is that it either is true or false.

[148] In RTT one "revises" the hypothesis of truth-value once a contradiction is reached. If such a revision also implies a contradiction, the sentence is paradoxical. That way RTT gives a model of the way we recognize paradoxes.

In IF (as well as in RTT) this does not occur, since we have the game-theoretic definition of truth. While Kripke's Kleenean logic is committed to the three possible interpretations of (TT), RTT and IF have the power to establish that (TT) is true or false depending on whether (TT) is true or false, respectively. Game-theoretically this is elementary. So also in this sense IF has the edge over Kripke's approach.[149]

Still, does it make a better case in the sense that one can adequately define the truth predicate within IF to escape the Tarskian hierarchy? This does not seem to be the case, as was argued earlier. The truth predicate of IF can be intensionally adequate, which is not a small matter, but this adequacy cannot be expressed in IF. While IF seems to have some edge over the Kripke-Kleene approach, I must agree with De Rouilhan and Bozon that it has not exorcised Tarski's curse. As far as expressive power is concerned, it has important potential, but in the philosophical question of truth very little seems to change with the introduction of IF. The monolingual speaker problem would have to be solved and, in the end, in that respect IF does not fare much better than Kripke's approach.

How about other means of using different logics to define truth predicates? There is an interesting article by Ketland (2003) about the ability of many-valued logics to contain their own truth predicates. Ketland (ibid., p. 293) reminds us of the "Revenge Problem" (one is prompted to Haack 1978, pp. 147-148) of the many-valued approaches to semantic paradoxes. If we consider the strengthened liar sentence λ = "λ is not true" in a many-valued logic with the truth-values *t*rue, *f*alse and *u*ndecidable, λ will get the truth-value *u*. Now clearly if the truth-value of λ is *u*, it cannot be *t*. But when we think of what λ *says*, that is: λ is *not true*, we see that this is exactly the case. So in fact λ is true after all and the familiar contradiction arises. In many-valued logics there may not be the law of excluded middle, but sentences can hardly be true

[149] Of course in two-valued logic such a problem does not occur, due to contradictory negation.

and undecidable at the same time. Adding the third truth-value changed nothing.

The main result of Ketland's article (ibid., pp. 295-296) is really just a corollary to the revenge problem. Ketland points out that in addition to being able to include its own truth predicate, a language *L* would also need to be able to discriminate between the truth-values and the other elements of its domain. Essentially, *L* would have to be able to establish that *t, f* and *u* are truth-values, and not among the "normal" domain of the language *L*. If such a distinction is not made, one can get a Tarski-type undefinability result by forming a contradictory self-referential formula also in many-valued languages, and hence show that the notion "truth degree (value) of a formula of *L*" is not definable within *L*. The only potentially problematic assumption one needs to make is that the identity relation in the language is bivalent, that is, we can distinguish between the three truth-values. But this is hardly a drastic presupposition, because denying that would mean that there exist formulas that are, for example, both true and undecidable at the same time. This result of Ketland's is in line with our earlier conclusions. Many-valued languages may contain their own truth predicates, but they cannot contain everything needed to recognize and use them as such.

Finally, we should consider the overall treatment of paradoxical sentences in many-valued logics. The strategy almost everywhere seems to be giving paradoxical sentences a truth-value "undecidable" or such. In this sense, many-valued languages can avoid the paradoxes, but with the cost that the semantic content of the liar (fixed-point) sentences is no longer followed. Here RTT and IF logic fare better than Kripke's approach because from them we can see how the paradoxes are recognized. When it comes to liar sentences, this indeed seems like the ultimate success. But the matter is different when we consider Gödel sentences. While a liar sentence is indeed undecidable, we should want our theory of truth to recognize the Gödel sentences as *true*, like Tarskian truth does, not *undecidable* like many-valued logics do.

5.6 Collapsing the hierarchy with pre-formal languages

At this point we must ask how Tarskian truth is better than the approaches considered in this chapter. As we remember, the most important problem with a Tarskian truth predicate is its demand for a hierarchy of languages. To be more precise, the main problem is that the hierarchy does not *collapse*. In the optimal scenario, like the one Hintikka is after, we would have one formal language and one truth predicate in that language. Tarskian truth contradicts with that on both counts: we cannot have just one formal language if we want to include truth in it, and no language can contain its own truth predicate. Not only that, but there is no way of ending this regression: to have a truth predicate for a language L_n we must always postulate a metalanguage L_{n+1}, and this goes for all $n > 0$. There is no way to collapse this hierarchy: for no n can we define its own truth predicate, and for no $m < n$ will L_m define the truth of L_n. Even worse, within that hierarchy of languages, we cannot seem to have any valid method of ending the regression to introduce the "basic" metalanguage. Infinite regressions must be avoided, and Tarskian truth seems to imply an infinite regression of languages. This looks like a daunting problem for the notion of semantical truth.

However, as suggested earlier, that is only a problem from the formalist point of view. For the non-formalist, there is a simple solution available. Tarski's result of undefinability of truth concerns *formal* languages. Should we limit ourselves to formal languages, the hierarchy indeed cannot be collapsed, and we must commit to an infinite regression of languages if we hope to include a truth predicate. Of course at this point the formalist will rather dispense with the truth predicate. But why should we limit the domain of philosophy of mathematics to formal languages? I have argued earlier for the acknowledgement of pre-formal mathematical thinking, and the same arguments can be used here for the introduction of pre-formal languages into Tarskian truth. At any convenient point in the hierarchy, to define truth for L_n we can use a *pre-formal* language to do this. In this framework all the symbols and sentences of L_n and all the T-instances of it are included in the pre-formal metalanguage. In addition, the actual

definition of truth can be explicitly stated and the hierarchy immediately collapses. Moreover, we can discuss truth and do all the necessary theorizing and philosophising to go through the project. Although this approach might look a little too easy, there is nothing new or controversial here: this is in fact the way the truth predicate is *actually* defined in all accounts of Tarskian truth.

The big question about the suggestion above is whether we want the pre-formal language to include its own truth predicate, and indeed, whether it *can* contain it. The latter question is somewhat problematic, as the pre-formal languages must include their own semantics and therefore suffer from the liar's paradox. However, the concept of pre-formal language is wide enough to include considerations of the diagonal method on which all known paradoxes over truth are based, and it can definitely have enough expressive power to label such paradoxical sentences as undecidable. Indeed, pre-formal languages have all the expressive power we need in meta-mathematics and philosophy. This does not save them from paradoxes, but it gives us means to handle them in a manner that conforms to the way paradoxes are actually handled in the literature. However, one must ask whether we should even hope to include an adequate truth predicate for our pre-formal languages? In other words, should we hope to have a single truth predicate, instead of a language-dependent hierarchy? This will be addressed in the next chapter.

Before we go into that question, we must emphasize the connection to the earlier considerations over pre-formal mathematical thinking. My point, of course, is that we actually use pre-formal languages all the time, whether it is implicit or explicitly acknowledged. The whole question of finding a metalanguage for a pre-formal language should never arise, since all the non-formal theorizing we do in mathematics and the philosophy of mathematics is done in a pre-formal language of mathematics. Mathematical languages can be divided quite naturally into formal and pre-formal ones, and both account for an indispensable part of the mathematical practice and theories. It should be obvious that in order to include truth in the formal language that we use in an area of mathematics, we must do it in a pre-formal language. And to remember the philosophical and

mathematical context we are concerned with here, this is what all non-deflationist theories of mathematical truth do.[150]

5.7 Why logicism and single truth predicate?

Second-order logic, IF logic and Kripke-Kleene logic are all forms of logicism – in one of the senses of the term – attempts to provide a purely logical foundation, or mode of presentation, to mathematics. However, none of them succeed in being completely free of extra-logical concepts, mainly set theory. For a language to be the basis for mathematics, it would have to endure against De Rouilhan's and Bozon's monolingual speaker question. None of the alternatives considered here quite manage to do that, which of course prompts one question: *just how committed should we be to the logicist ideal?* Logicism has an important history since Frege, Russell and Whitehead, but this history contains one important tragedy: none of the logicist programs have ever succeeded. In a spirit similar to that of Hilbert's program, logicism would have been a maximally economical ontological foundation to mathematics. It failed, but even so, one must ask whether its success would have made much of a difference to mathematical practice? All the pre-formal concepts we use, and indeed *must* use, would almost certainly have remained untouched. In addition, most of formal mathematics would have continued as before. Mathematical practice does not seem to be that tightly connected with the theories concerning the foundations of mathematics. Hence, the logicist motivation is not as obvious as is often implicitly assumed. It is widely presupposed that logical concepts like the rules of valid inferences (that is, logical connectives and their truth tables), are somehow ontologically simpler and less problematic than the primitive mathematical ones, such as set membership, and moreover, semantical concepts like truth. However, the layman might not see much difference between the

[150] In fact, all the deflationist ones as well, but they should not be called theories of truth in the first place, as the truth predicate exists only as an empty translation of proof.

rules of valid inferences and the truth they are supposed to preserve. Here it is hard to see why the layman would be wrong.

In this chapter we have examined the alternative ways of defining a truth predicate in logic, and in mathematical theories. The whole problem of truth in mathematics seems to unravel nicely into a series of questions and answers. Do we need a truth predicate? A negative answer implies deflationism, and hence ultimately extreme formalism. If we answer positively, another question looms: do we want the truth predicate to apply to its own language? If yes, by Tarski's proof we must commit to truth-value gaps or many-valued logic in the case of first-order logic, or else we need second-order logic. If not, we commit to a Tarskian hierarchy. The former have been discussed in the previous chapters. But what if we commit to a Tarskian hierarchy, that is, what if our truth predicates are always tied to specific languages (and their metalanguages) instead of there being a general all-purpose predicate? How problematic is this intuitively?

I think that the potential difficulties of an infinite regression and the non-intuitiveness of a Tarskian hierarchy of languages are exaggerated. First of all, there is the problem of defining truth in logic, even if we do not limit ourselves to classical first-order languages. Hintikka needs arithmetic plus he uses model theory and pre-formal concepts to establish that his predicate is indeed the truth predicate for IF. Second-order logic as the basis is altogether more difficult and has limited appeal over set theory. But there is also the questionable motivation for such all-purpose truth predicates. Kripke (1975, pp. 694-695) mentions the argument that we do not generally implicitly assume the "levels (hierarchies) of languages" when we utter things concerning truth. If we utter "snow is white", we cannot specify (implicitly or explicitly) what language we are discussing here. Ultimately, for each of us, there is only *one* language and as such only one truth predicate for it. This is the predicate that theories of truth should be after, not some hierarchy of "truths in languages". As we saw, this was also Hintikka's motivation. We have seen the problems that this kind of approach implies when we considered single-system formalism in mathematics. But as far as the general pursuit of knowledge is concerned, at first this line of thinking seems intuitively appealing

– especially if we are committed to some version of the correspondence theory of truth. Truth as a kind of relation (or description of a relation) between our utterances and the world seems to imply that there can only be one truth predicate, just as there is ultimately only one language (for each of us) and only one world.

However, that counterargument uses a very naïve version of the correspondence theory of truth, and the realism that usually follows it. Surely after all the scientific and philosophical work of the last century we do not expect a single scientific theory to explain everything there is in the world? It seems to me that as far as science (I am primarily thinking about physics here) is concerned, a Tarskian model-theoretic truth is *exactly* what we are after. Take the physical sentence "light consists of particles". It is true in one model of physics, and false in another, according to which light consists of waves. Empirically, we have no reason to prefer one to another, and both are needed in the best theory of physical knowledge that we currently have. It seems that to account for this we need a notion of truth-in-model (and similarly, truth-in-language), and an all-purpose single-language truth predicate is not necessary. Furthermore, we need metalanguages to discuss these object systems, that is, the scientific models.[151] In practice, our natural languages of course fulfil this function. Empirical science, just like mathematics, does not seem to conform well to all-purpose languages and all-purpose axiomatizations. This ideal seems outdated in the philosophy of science. I do not see any reason why we should stick to it in the theory of truth.

The only reason I can see for advocating an all-purpose truth predicate is to have one for our *natural* language. After all, that is the one obvious case where we cannot rely on hierarchies and

[151] Here one must remember that the concept of a model in empirical science can be very different from the one we have in logic and mathematics. However, as seen in the question of particle-wave dualism, there are also enough similarities to give grounds for the argument here. The problem of particle-wave dualism itself goes deep into the philosophy of quantum mechanics, and as such cannot be handled in detail in this work. See Omnès 1999 for a good introduction on the subject.

model-theory. But is it realistic to demand an explicit truth predicate for our vague natural language? Indeed, is it even desirable that we should have one? We have our own existing conceptions of truth that we use and we seem to give little attention to their intricacies. What do we lose if we restrict our philosophical investigations on the subject of truth to subject areas like mathematics and science? But this kind of discussion is going to be too far off from the matter at hand. My point is that the (often implicitly required) need for general all-purpose truth predicates is vastly exaggerated in some parts of the literature. This line of thinking usually includes the notion that there must be something inherently flawed in Tarskian truth since it does not fit well into such projects. We should not accept that reasoning too easily. Neither mathematics nor physics – not to mention the less formal sciences – can be feasibly presented as a single theory. In fact, given the need for pre-formal languages, Tarskian truth fits perfectly well with the subjects we should expect it to apply to: the philosophy of mathematics and science. The purpose of this last chapter has been to emphasize this point: as well as no motive *inside* mathematics, no motive *outside* mathematics should make us view Tarskian truth as fundamentally flawed.

6. Why not nominalism?

6.1 Semantical arguments and the trouble with reference

After discussing the possible logics and theories of truth for mathematics, we must return to the contention that started all this: that there *is* no truth. That is of course the deflationist and extreme formalist position. Unless we accept extreme formalism, the semantical arguments give us an explicit difference between the concepts of truth and proof in classical two-valued mathematics.[152] At its simplest, this happens once we expand the formal system with Tarskian truth. As we have seen, pre-formal mathematical thinking fits well together with Tarski's theory of truth, and it gives us naturally the distinction between proof in formal systems and truth in pre-formal systems. As far as textbooks – and all other human works – of mathematics are considered, this is the implicit distinction used everywhere. No textbook of mathematics is purely formal, even if the formal presentation contains the core of them. From the literature on the subject, discussions with colleagues and personal experience, I have not encountered any evidence of human beings processing mathematical sentences completely formally. Moreover, it seems that even extreme formalists would accept this. The difference is that for them formal mathematics captures all there is to mathematical thinking, and the pre-formal part is only a heuristic tool.

We know that the semantical arguments need expansions to formal systems in order to be carried out. In this work I have argued that when we consider mathematics as a whole, instead of just single formal systems, Tarskian truth is in fact no expansion at all. It is already contained in our pre-formal thinking. However, it must be acknowledged that this is only half of the picture. Tarskian theory of truth is also a theory of *reference*, and that is one aspect bound to trouble the extreme formalist. As we have seen, the formalist position is that formal mathematical systems are completely self-standing. In that picture there is no room for

[152] As we saw, the situation is not much better for the deflationist in any of the other proposed logics.

references. Against this, I have argued that formal mathematical concepts refer to their pre-formal counterparts. The formalist response to this is easy to predict: what, then, are the references of pre-formal concepts? If they have no references, why could we not accept the *formal* concepts having no references? This is a perfectly valid question. While it is indisputable that people use pre-formal concepts, they could still be ultimately redundant. Thus pre-formal mathematics would simply be another way of looking at formal systems, the *real* domain of mathematics. It could be the case, the formalist is bound to argue, that we just cannot expand formal systems to include referential concepts like Tarskian truth; there is nothing *out there* that these expansions refer to. If that indeed were the case, Tarskian truth would have to be deflationary, and the semantical arguments erroneous.

The above is ultimately the key question in the whole subject of semantical arguments. Without any appeal to the references of pre-formal concepts the arguments would fail. We would have trouble distinguishing between truth and Tennant's arbitrary soundness principle, and hence deflationism would survive unscathed. The incompleteness of formal systems would simply imply incompleteness of truth, and the apparent truth of the Gödel sentences would turn out to be an illusion. We have seen the problems with that line of thinking, but if all theories of reference fail, there would not seem to be any other option. Mathematical reference, however, is a highly problematic question, perhaps the most problematic one in the entire philosophy of mathematics. When it comes to the semantical arguments, the most troubling part is that from Gödel to Roger Penrose, some of the most famous proponents of them have used Platonism as the answer. For the philosopher who does not subscribe to Plato's world of ideas, his philosophy of mathematics presents us with many serious problems. But even for a proponent of Plato's ontology there is one big difficulty: *Benacerraf's dilemma*, the problem of combining Platonist ontology and epistemology. Paul Benacerraf (1973) reminded philosophers of this important question: if mathematical objects exist in the Platonist sense, how can we ever get knowledge of them? His question is based on the causal theory of knowledge. If a subject *S* believes that a sentence *P* is true, there must be some

causal connection between *S* and the reference of *P*. In the Platonist philosophy of mathematics the reference of *P* is an abstract object. What can be the causal relation between an abstract, non-physical, object *P* and a physical subject *S*? Since Benacerraf's work, an account of mathematical knowledge must be able to answer this dilemma.

For a hard-line Platonist, this question seems to cause insurmountable trouble when it comes to the central question of theory choice. We cannot use the applications of mathematical theories in physical sciences – as well as all the direct applications – as an argument, because one can only get knowledge of the ideas by reason, not by observation. Basically, we would need to have a direct way of telling from an axiom (or a theorem) whether it is true or not, that is, does it correspond to a Platonist idea. Outside of some form of mysticism, there seems to be only one way of telling that: does it follow our *intuition*? In essence, this is the position that human beings have a special epistemic faculty for gaining mathematical knowledge. This point of view is endorsed by, among others, Gödel (1964b), Dummett (1978) and Penrose (1989, 1994) in more recent times. But even though quite popular in philosophy, mathematical intuition is a notoriously vague concept, and a prudent philosopher will be wise to steer clear of it. In this work I cannot accept the possibility of replacing extreme formalism with Platonism.

Knowing the problems of Platonism that Benacerraf's dilemma and other counterarguments reveal, all this can make the semantical arguments seem less attractive than they actually are. The main point to remember is that while without references the semantical arguments fail, with *any theory* of references they succeed. I will argue that there is only one theory in the philosophy of mathematics that does not include any reference at all to non-formal concepts, and that is extreme formalism. All other theories include some references for mathematical concepts, whether they are the Platonist objects of Gödel, *ante rem* structures of Shapiro (1997)[153], empirical concepts of Phillip Kitcher (1983) or the

[153] Realism over mathematical structures, like the natural number structure, as opposed to entities such as numbers.

naturalist ones of Quine (1995). As long as there is *something* outside formal systems that our mathematical concepts refer to, the semantical arguments are sound. This does not need to mean that mathematical concepts or sentences refer directly to some outer objects. In the minimal form it is sufficient that we have any outer reason at all to believe that some mathematical statements rather than their negations are true. In Quine's philosophy, for example, this can mean that a statement works better in physical applications. That is how wide a concept reference is. In this work I have argued that not believing in *any* reference means that we believe in the arbitrariness of mathematical theories. In this chapter I will clarify that argument by taking a closer look at the possibility of no reference for mathematical statements.

So to answer our question about the references of pre-formal mathematical concepts we can twist the problem around. What reason would we have to believe in extreme formalism, that is, the position that there are *no* references for formal mathematical sentences? If we end up rejecting it, we know that we are accepting some sort of theory of reference, and we know that the semantical argument can be carried out. The references of pre-formal concepts can be Platonist, structuralist, empiricist, naturalist or some other option, but as long as we are rejecting extreme formalism, we are at least implicitly accepting one of them. It is my purpose now to show that this is indeed what we must do, since extreme formalism fails as a plausible philosophy of mathematics. Whichever ontological or epistemological theory we accept is irrelevant for the arguments in this work – as long as it is not extreme formalism. Of course this is not completely satisfactory, since it leaves open many of the most important questions in the philosophy of mathematics. Most worryingly, it can leave the impression that non-formalism in some way suggests Platonism. Against this, in the final section of this chapter I will give my brief outline of a non-Platonist, non-formalist approach. It should not be considered as central background for the arguments in this work – rather, it works as a reminder that the epistemology and ontology of mathematics is not exhausted by the two extreme positions.

6.2 Meno's paradox and theory choice

Gödel's (1951, pp. 311-314) own contention was that the incompleteness of formal mathematics suggested a form of Platonism. According to him, mathematics cannot be our own creation because we cannot create properties that we cannot possibly know – that is, prove. This is of course highly dubious. Why not, one must ask? Formal mathematical systems can be highly complex and the axiomatizations can imply many sentences that we can neither prove nor disprove. It seems perfectly possible that some of those theorems are also undecidable in principle, due to the properties of formal systems. Indeed, if we have enough tools in our language to formulate self-referential sentences, this will happen.[154] There does not seem to be any problem in us being able to formulate undecidable sentences. Clearly we must look elsewhere if we hope to refute formalism.

It seems to me that the main motivation for extreme formalism can be illuminated by the ancient Meno's paradox. The main flaw of that paradox is also projected into formalism. Meno's paradox, concerning inquiry of knowledge, is the following:

(1) If we know what we are inquiring, inquiry is not needed.

(2) If we do not know what we are inquiring, inquiry is not possible.

(3) Because we either know or do not know what we are inquiring, inquiry is either not needed or not possible.

Now projected to mathematical truth and proof, the paradox gets the following form[155]:

---

[154] It should be noted that there are also undecidable sentences that are not self-referential, Continuum hypothesis being one example.

[155] My argument here is motivated by one of Hintikka's (1996, pp. 34-36) for different purposes.

(4) If we know what sentences we want to prove in mathematics, truth is not needed.

(5) If we do not know what sentences we want to prove, finding out mathematical truths is not possible for us.

(6) Therefore, mathematical truth is either not needed or not possible for us.

I think this corresponds to the extreme formalist inference, even though in a rather unusual form. When we have our axioms and rules of proof, why do we need truth? If the axioms and rules of proof are correct, then we can use them without any reference to truth. After all, proofs in mathematics do not mention truth – except informally – outside the realm of logic. If the axioms and rules of proof are not correct, then we are not finding out true sentences to begin with. In either case, truth is not something we need to include in mathematics. Or so the formalist argument seems to go.

Of course we should be quite sceptical over such paradoxes having direct relevance to a subject such as mathematical truth, and a modern formalist is not likely to use Meno's paradox explicitly in an argument. Nevertheless, I claim that the formalists are doing something very similar in their actual deflationist arguments. That is why we can learn a thing or two from Meno's paradox and its flaws. Obviously the main flaw in Meno's paradox is that we do not differentiate between the types of things we are *inquiring* and the types of things we are *finding out*. If we consider a specific moon of Saturn, say Titan, then obviously inquiring the existence of a moon with Titan's location and characteristics implies the paradox. But if we are looking for moons of Saturn in general, then we can find new knowledge and the paradox disappears. We can know that we look for moons of Saturn and gain knowledge of new moons, as is indeed done constantly in our times. But we do not know which *particular* moons we are looking for. Thus, "we know what we are inquiring" in (1) and "we do not know what we are inquiring" in (2) do not need to refer to the same object of inquiry, contrary to what is assumed in the paradox.

It is not the case that (3), that we either know or do not know the object of our inquiry. Hence, under analysis, the paradox is not really a paradox at all.

That much is more or less elementary, but it gains new power when projected to the version about mathematical truth. Of course proof is the mathematician's way of finding out true sentences. If we know which theorems we want to prove, truth is not needed. If we do not know which theorems we want to prove, truths are not possible to find out. Now is it the case that either we know or we do not know which theorems we want to prove? Obviously not, as was seen above in the original version of the paradox. Against (4), we do not know all the theorems that we want to prove in, say, arithmetic. We do not know if we want to prove Goldbach's conjecture, because we do not know whether it is true or false.[156] But against (5), we know that we want to prove theorems that satisfy our choice of arithmetic, that is, the Peano axioms and rules of proof. We choose the axiomatizations of arithmetic that fulfil our basic intuitions and experiences over natural numbers – in other words, what we conceive as the basic *truths* of arithmetic. The object of proof in (4) is different from the object of proof in (5), and hence neither (4) nor (5) is the case. We do not know *exactly* what we want from our formal systems, but we do know something – in fact quite a lot – by making theory choices and requiring theorems to be consistent with them.

One notices how we differentiate between truth and proof in the two cases. We have to – otherwise the paradox emerges. Indeed, when talking about proof and truth in the strict formalist way, we are always talking about the same concept. For the extreme formalist the paradox definitely arises, and as such truth could indeed be conceived as deflationary. If we want to prove provable sentences, quite clearly truth is a redundant concept. But can proving provable sentences be all we want? The answer should be obvious by now: *of course not*. All systems of mathematics prove sentences provable in them. Our mathematical knowledge,

---

[156] It must be remembered here that we are talking about a strict notion of mathematical knowledge. Certainly it can be plausibly said that we have a good *guess* about the truth of Goldbach's conjecture.

however, does not concern all systems of mathematics. It concerns a select few, and the choice between them and the other alternatives could not have been completely arbitrary. This is where extreme formalism is at its weakest, as was already seen in the discussion on Tennant and Ketland. The current chapter has been just another effort to illuminate the difficulty that the problem of theory choice presents for the extreme formalist. Clearly we are only interested in very few specifically constructed axiomatic systems, while the number of all possible axiomatic systems is infinite. It cannot be a simple matter of chance that we have stumbled upon the ones we use. This is the problem of theory choice and it is the most serious obstacle for extreme formalism. In fact, I claim that extreme formalism is refuted by it since the only remaining alternative for an extreme formalist is the arbitrariness of mathematics.

If mathematics is a fiction without any outer reference, how do we explain theory choice? Since it cannot be because of anything outside the formal systems, the reasons behind theory choice must be internal. To be sure, internal criteria are used in mathematics all the time. Simplicity, for example, is generally accepted as one such criterion. But all the internal criteria only form one side of theory choice. We must never forget what the lack of any outer criteria in theory choice implies. It implies arbitrariness, and arbitrariness means that we could have just as well thought that 2 + 2 = 5. None of the evidence, applications in science, practice, and intuition – common sense, if you like – would matter: we would have to be ready to accept that it could have been possible to put two and two apples together and proceed to take one and four apples from the pile. That is an extreme case, but arbitrariness is an extreme position. Mathematics has highly sophisticated applications in other sciences, but grouping apples into piles is also an application of mathematics. No internal criterion will prevent us from thinking that 2 + 2 = 5 = 1 + 4, but to most of us it seems that some outer one quite clearly does.[157] I contend that whatever we may want to call

[157] If one wants to argue that a system where 2 + 2 = 5 = 1 + 4 is inconsistent, he must remember that the inconsistency only follows from the Peano (or some other accepted) axiomatization where 2 is the

the criterion for that choice, we might as well call it truth – for the purposes of this work it does not make a difference what the exact conditions of theory choice are.

6.3 Benacerraf's dilemma and nominalism

I have mostly been using the term (extreme) formalism, as defined in Chapter 2.4, to refer to the philosophical doctrine that mathematical objects do not exist and mathematical sentences do not have objective truth-values. In the literature, the term nominalism is nowadays used more often, and essentially it is the same point of view. Nominalism by its very name is the position that mathematical objects do not exist – only their *names* do. This is exactly what extreme formalism is, applied to the specifically mathematical concept of formal systems. According to formalism only the formal systems exist, and since formal systems are human creations, this clearly implies nominalism.[158] So the choice of terminology is not important here. However, we will see that the word "nominalism" carries a danger of misinterpretation, since it also refers to nominalism in physics, which has some quite different philosophical characteristics. Hence, I would like to continue using extreme formalism as the preferred term. But as it does not make a philosophical difference, I will conform to the established terminology, and use the word nominalism from now on.

As John P. Burgess & Gideon Rosen (1997) have written in their comprehensive book on the subject, the modern history of

---

successor of 1, 5 the successor of 4, and they do not have other successors. In addition, most importantly, one would need to accept that all natural numbers have successors. But from all the possible axiomatizations, ones that satisfy these conditions form only a small fraction. Furthermore, due to Gödel, consistency alone cannot be a criterion for theory choice since we can never know it from formal systems of arithmetic. Finally, how can consistency be thought of as an internal criterion in the first place if it is something we *informally* require from formal systems?

158 It also implies *fictionalism*, a term used mostly for Field's nominalism/formalism.

nominalism in mathematics can be traced back to two philosophers. Nelson Goodman and W.V.O. Quine presented in 1947 their nominalist program, the message of which was simple: they renounced abstract mathematical entities such as properties, relations and classes. Quine later gave up nominalism and Goodman's arguments for it were shown to be very weak[159], but the basic form of nominalism has ever since followed Goodman's and Quine's trail. The arguments have evolved and developed a wide heterogeneity, but basically nominalism is still the viewpoint that abstract mathematical entities do not exist – the exact formulations of the arguments do not change that.[160]

Ironically, perhaps the most important contribution to nominalism was made by a non-nominalist. That is of course Benacerraf's dilemma. Benacerraf's article is a basic one for nominalism, but it is only that in the negative sense of being anti-Platonist. Keeping in mind that Benacerraf is not a nominalist himself, it is hardly surprising that he did not offer any nominalist alternative. Unfortunately, presenting an alternative would have been highly relevant, because Benacerraf's dilemma has the worrying potential of repudiating *all* philosophical interpretations of mathematical knowledge. The most important question left open is the one that the realist is bound to present first: if not an abstract object, then what sort of object is the reference of a sentence P that satisfies the causal connection? It is too often implicitly assumed that Benacerraf's dilemma makes a case for nominalism. It *does* make a case against Platonism, but in what

---

[159] One famous argument by Goodman (1956) was against the existence of classes (sets). In it he argued that the sets $\{\{a\},\{a,b\}\}$ and $\{\{b\},\{a,b\}\}$ cannot be different from each other since they consist of the same elements a and b. Thus sets cannot exist, because according to set theory the two sets are different. Of course, like Burgess & Rosen (pp. 27-28) point out, this is not unlike saying that a statue cannot be different from the lump of clay it is made of. After all, they consist of the same molecules.

[160] The term "nominalism" is sometimes also used to refer to the viewpoint that *sets* in particular do not exist. Here I will follow the definition given in Chapter 2.4, which conforms to the ones given by Field (1980, p. 1) and Burgess & Rosen (1997, pp. 13-25).

way does that imply success for nominalism? The enemies of one's enemy are not necessarily one's friends, and this is certainly the case with nominalism. It is not easy to see how a nominalist can account for the causal theory of knowledge in mathematics any more than a Platonist could.

If not abstract, then mathematical objects – if they exist – would have to be concrete for the nominalist.[161] How can we get mathematical knowledge of concrete objects? The only option seems to be by sensory perception, whether directly or indirectly. But in that case mathematics would be empiricist in the sense that Philip Kitcher (1983) has proposed. In Kitcher's words, mathematics is "an idealized science of operations which we can perform on objects in our environment" (ibid., p. 12). This way, mathematics would not differ essentially from any empirical science. We get mathematical knowledge by observation, and the process is more or less similar to the one in other sciences. Kitcher's theory has been shown to be very problematic, but that is a minor point here.[162] The important point is that it is not *nominalist* in the strict sense we are concerned with in this work. Mathematics would clearly have a reference, and it would make sense to speak of the truth of mathematical sentences. In addition, if we get mathematical knowledge empirically, we have all the problems of general philosophy of science with us.

---

[161] I will not go into the problematic details of the abstract/concrete dichotomy in here. I believe that the nominalist has more serious problems, ones that are unrelated to the exact nature of the dichotomy.

[162] Perhaps the most famous proponent of empiricism in mathematics was John Stuart Mill (1874), who claimed that mathematics is empirical in a very direct way (see Ayer 1946, pp. 74-75 for a good criticism of Mill's philosophy). Kitcher's philosophy is different in the sense that mathematics is a *construction* based on empirical evidence. The important point in both Mill's and Kitcher's philosophy is that mathematics loses its *a priori*-nature. Kitcher's more sophisticated account is an appealing one when we apply it to the *origins* of mathematical thinking, but when it comes to developed theories like complex analysis, it is hard to retain the empirical connection. See Hale 1987, pp. 123-148 for further problems in Kitcher's philosophy. The other famous proponent of empiricism in mathematics is of course Quine, for whom *everything* is empirical.

Here it is easy to confuse scientific nominalism in general and *mathematical* nominalism. At least as far as the subject of this work is concerned, mathematical nominalism can only be understood as Field has described it: *fictionalism*. There is clearly a difference between the scientific nominalist denying the existence of abstract objects in, say, physics, and the mathematical nominalist denying the existence of *all* objects of mathematics. While scientific nominalism is hardly a uniform position, one is not likely to find many modern nominalists who hold the outer world to be a fiction. The mathematical nominalist, however, must contend that mathematical entities do not exist *at all* if he wants to remain a nominalist in the strict sense required for deflationism over mathematical truth. If he did not, mathematics would clearly have a reference: our physical world. Perhaps we should not even begin to speculate what these physical mathematical objects could be, but evidently our mathematical statements would have to refer to something in those objects. As far as questions like truth, proof and the semantical arguments are concerned, empiricism would be no different from Platonism. Mathematics would clearly have a reference, and when our statements correspond to the state of affairs in this reference, it would make perfect sense to speak of them as true, and not only provable. This is something the strict nominalist obviously cannot accept.

Evidently, mathematical objects cannot be concrete for such a strict nominalist. By the definition of nominalism they cannot be abstract. Only the third option remains: they cannot be *anything*, that is, they are fiction. That is what Benacerraf's dilemma seems to imply when understood in a nominalist fashion. All theories of mathematics are simply conventions without any reference to anything other than conventions. But this is to say that all mathematics is, in one word, arbitrary. By now we should be familiar enough with the problems of this position.

However, I do not think that nominalism is repudiated by this argument. I do think that the argument from *causal theory of knowledge* is flawed, but it is hardly the most appealing argument that the nominalist can make. After all, why should we accept a straight causal theory of knowledge, let alone for *all* types of knowledge? The background for Benacerraf's problem is the causal

theory of knowledge by Alvin Goldman (1967). There is no need to go into all the details of that theory here, but it should be pointed out that the requirement of a causal connection between the subject and the object has not turned out to be as unproblematic as it may have once seemed. For a modern philosopher of science it is perfectly acceptable to speak of electrons in physics, yet we can never establish a direct causal connection between electrons and our beliefs about them. For mathematical knowledge, one would not know how to start applying Goldman's theory. Basically, by not requiring a strict causal theory of knowledge for all knowledge, I contend that Benacerraf's dilemma does not need to be the behemoth it is often thought to be.[163]

This is not to say it is not important: clearly it does capture one difficulty in gaining knowledge by mathematical intuition, or other suggested ways of getting purely abstract knowledge. But we do not need to think that mathematical objects are abstract without any *connection* to the concrete. Mathematical knowledge can be conceived as recognizing patterns in the physical world, and abstractness can still be included in the structural reference of these patterns.[164] We can think of the arithmetical structures as being abstract, and still believe that we get knowledge of them by studying collections of apples. There is no contradiction in that, and it seems to satisfy our basic experiences of learning mathematics. The nature of the connection between the abstract mathematical patterns and the concrete in such an account is of course one of the most fundamental questions in the philosophy of mathematics. However, it is simply too much to require that in order to be an anti-nominalist one needs to have a definite answer to that question. After all, whatever the proposed answer is, it can hardly be less plausible than the only seemingly *completely* nominalist explanation for mathematics: arbitrary conventions.

---

[163] For more on the problems of the causal theory of knowledge in mathematics, see Burgess & Rosen 1997, pp. 35-39 and Maddy 1984.

[164] See Shapiro 1997, pp. 109-116 for a philosophical outline of pattern recognition. Also Resnik (1981, 1982) writes about patterns as the key to the epistemology of mathematics. See Davis 1984 for an example of pattern recognition in the psychological studies of mathematics.

Before we go into other nominalistic strategies, we need to address two appealing lines of thought. The first one is what Burgess and Rosen (pp. 97-122) call the *geometric* strategy to nominalism. According to this philosophy, all mathematical objects are ultimately geometric objects. Geometry, in turn, should take a wholly synthetic approach developed by the likes of Hilbert (1900) and Tarski (1959). What this means is that the domain of geometry is *points*: points, and only points, are the entities of geometry. All the other geometrical objects, and ultimately all other mathematical objects, should be constructed synthetically as relations between points.

The advantages of the geometric strategy are instantly recognizable. First of all, the important connection between physics and mathematics is evident from the beginning: both deal within the domain of points of space, or space-time. Secondly, nominalism would be based on something: it would not make mathematics arbitrary. Mathematics would have as its foundation an explicit theory of existing entities, that of points in space-time. All the other entities, like lines, triangles, natural numbers and real numbers would be constructed synthetically from them. This greatly facilitates the explanation of the ontological status of mathematical entities. After all, geometry as traditionally understood was the study of the actual physical space, and even the developments of non-Euclidean geometries have not totally changed this premise. In fact, the connection between physics and geometry is the same for Newtonian and Einsteinian physics: only the geometrical theory that is applied is different. In this way, mathematics is intertwined with physical theories, and the choice of the best theory of mathematical geometry would depend on the choice of the best theory of physics. This would make the ontological questions of mathematics similar to the ontological questions of physics. The most reasonable answer to the existence of mathematical objects would be, somewhat simplified: whatever physics holds it to be. Consequently, Benacerraf's problem would not exist – unless we consider it to exist in physics, as well.

One problem concerning this geometric strategy is that it can be developed only when the existence of a continuum is presupposed. If space-time is presupposed to be discrete, we cannot use the

synthetic geometrical strategy to include real numbers.[165] The continuum assumption may not be a problem for Newtonian physics or general relativity, but it could be highly problematic for quantum physics. If we cannot divide quantities into smaller measures than Planck units, we would not need the assumption of continuum in physics. It would seem that, with the geometric strategy of nominalism, we would not be warranted in making the assumption in geometry, either.

The problem of the continuum assumption exists in the other direction, as well. In Field's (1980, p. 31) nominalist program it follows from his axioms that the points of space-time exist, and that there are as many points as there are real numbers.[166] This way Field's program presupposes that there exist uncountably many points of space-time. His own contention is that this is much less problematic than postulating even a single abstract object. However, postulating a continuum (even the power set of the continuum, actually, see Shapiro 2000a, p. 232) of concrete objects is making a very definite ontological statement – one that may very well contradict with our best current theory of microphysics. We do not know whether our universe is even infinite, but we certainly do not know whether it forms a continuum. It does not seem at all clear that postulating abstract objects would be any more problematic.

Even if the problems above were solved (and they perfectly well could be, as it is still a somewhat under-studied subject), in this work the geometrical strategy of nominalism is not a factor. That is because, like empiricism, it is not really nominalism at all in the sense relevant to the question of truth and reference. Strict nominalist truth is the deflationist theory that our axioms and rules of proof, *and only them*, determine which sentences are true. In this

---

[165] Of course we can *formulate* them from the rational numbers as Cauchy-sequences or Dedekind cuts, but one has to wonder whether it is acceptable to formulate such basic elements of universe in the geometric strategy. Essentially, we would be assuming a discrete universe to formulate the universe of continuum.

[166] Or to be more precise, as many points as the *realist* mathematicians claim there are real numbers.

philosophy the axioms and rules of proof have no references. They could not have – otherwise it would make sense to speak about the objective truth of them, which contradicts with deflationism. But in the geometrical strategy, even if most mathematical objects would not exist, the points of space-time would. Ultimately, the points of space-time would be the reference of mathematical theories. With this approach, it would make sense to present such questions as which geometry is the *true* one – do the points of space-time, for example, form a continuum? Hence truth and reference would play a role in the theory. In the usual sense in the philosophy of science this would be nominalistic, but not in the strict mathematical sense we are concerned with. The better term for this position is *physicalism*, as it makes mathematical concepts a subset of physical concepts. The possible confusion in taxonomy is due to nominalism in physics being the view that abstract notions do not exist, but not the stronger fictionalist doctrine that *no* objects exist. That is also why mathematical nominalism would be more appropriately called extreme formalism, as it carries the more sophisticated notion of formal systems, but not the connection to physical nominalism.

Another related supposedly nominalistic theory has been suggested by Haskell Curry (1954, pp. 205-206). He proposed the concept of *acceptability* as a criterion for choosing between formal systems. Acceptability according to Curry can mean intuitiveness and consistency, but above all the *usefulness* of a theory.[167] This is an empirical criterion designed especially for physics, which is why Curry calls his position "empirical formalism". In short, if a theory is useful in physics, we have a reason to choose it over its competitors. Furthermore, we have a reason to believe that the theory is true. Yet we cannot call this approach nominalism any more than the empirical or geometric strategies. Clearly mathematical theories now have references, and again they are essentially the same one that the physical ones have. If empirical usefulness is a criterion in mathematics, does that make mathematics ultimately an empirical science? If so, what is the

[167] This corresponds to Field's ideas seen in Chapter 4.1 of this work.

ontological status of mathematical objects? How do we arrive at our axioms? Curry (ibid., p. 203) calls them conventions, but that only postpones the question. Conventions can arise in a number of different ways.[168] These kinds of questions will inevitably be raised. Granted, we have come a long way from the mathematical realism of Plato – but Curry's philosophy is still not nominalism in the strict sense, for the usefulness of mathematical theories cannot be explained by them being just conventions. Objectivity and truth once again creep in through the back door.

In conclusion, there are two distinct positions we are concerned with here: one that mathematical objects are abstract and one that mathematical objects have references. In the first sense the above strategies have been nominalist, but in the second sense they have not. The actual classification into nominalism is by no means uniform in the literature, and the one used here is certainly somewhat unconventional. In this work, only the extreme form of nominalism is called nominalism. However, there is a reason behind this, and it has to do with the subject matter here. If empirical data is in any way the basis for mathematics, then it clearly makes sense to speak of mathematical truth, even if only in the sense that it is one branch of scientific truth. Certainly that is not Platonism, but it is not strict nominalism either, as the formal systems would now have a reference: our empirical data. It would make sense to speak about mathematical truth outside of our formal systems, and that would be enough to make the semantical arguments valid, and truth a substantial concept. We are concerned here with the type of nominalism that does not include *any* reference to truth. In only that kind of extreme nominalism will the semantical arguments fail, and truth and proof would be the

---

[168] The extreme formalist, and the truly nominalist, position being that axioms are *only* conventions, and there is nothing behind them. This is the position I attribute to Field, but it is certainly held at least by Ludwig Wittgenstein (1976, 1983). Essentially, in Wittgenstein's conventionalist philosophy, the status of the axioms and the rules of inference is like the status of language. But as such, conventionalism of course faces the problem of theory choice.

same concept. Above all else, it is my purpose to show that this sort of extreme nominalism does not make sense.[169]

6.4 Field's nominalism revisited

All the considerations above bring us to my preferred choice of nominalism throughout this work: that of Hartry Field's. Field's position is, as we remember, deflationist and the utterance "truth" can only perform some light linguistic duties. Other than that, Field's nominalism is best described in his own words:

> In defending nominalism […] I am denying that numbers, functions, sets or any similar entities exist (Field 1980, p. 1).

We should not worry about the possible extensions of the term "similar entities", as ultimately it can only mean other entities that do not exist. Such entities are generally called abstract, but as we know, the abstract/concrete distinction is not an easy one to formulate. In any case, Field's position seems to be simply that none of the entities postulated in mathematics exist.

In a way Field's nominalism is a variation of the geometric strategy. That is how Burgess and Rosen see him and it comes across in Field's (1980) own writings, as well. He holds that the points of space-time exist, and that they are concrete instead of abstract. In this way he is a proponent of the geometric strategy. However, I cannot agree with this completely. It is true that his

---

[169] I also think that, when it comes to mathematics, this is the best way to do the philosophical taxonomy, and not just suitable for the purposes of this work. The basic meta-level argument of this work is that we can approach the question of mathematical truth prior to answering questions of mathematical epistemology and ontology. This way, the first question we should answer in the philosophy of mathematics is that of objectivity. Nominalism as a term may be tempting to use since it opposes Platonism, and this can be seen as one motivation for calling various philosophical programs nominalism. However, as I have argued, the nominalist strategies that ultimately lead to objectivism would be better labeled as something else.

strategy starts from (Newtonian) physics and the development of nominalistic mathematics for it, which might make it look like a part of the geometric strategy. But this does not go well with what was considered in the previous chapter and what we know about Field's attitude toward truth in mathematics. After all, Field is a deflationist over truth and a fictionalist concerning mathematical entities. Only under the interpretation that the physical points of space are also fiction does Field's fictionalist philosophy fit fully within the geometrical strategy. As was stated above, there are two distinct positions here: one that mathematical objects are abstract, and the other that mathematical objects have references. If the account given here is correct, Field must be against both if he is to remain a deflationist over truth, which is his most important contribution as far as the subject of this work is concerned. This is why I consider Field's work straight fictionalism, and not a proper part of the geometric strategy.

The possible difference in interpretations notwithstanding, Burgess and Rosen (pp. 41-49; pp. 191-196) have provided a very useful analysis on Field's (1989) nominalism. Loosely following them, Field's position can be dissected as follows. Field's main criticism is targeted against the "reliability thesis of the anti-nominalist", that is, the position that when mathematicians believe something about the entities of mathematics, then that belief is true. This is something that Field does not accept. Instead, what he wants is an explanation for the reliability thesis. What reasons do we have to believe in it? According to the causal theory of knowledge, the beliefs of mathematicians must somehow causally follow from the entities of mathematics. But the latter are abstract which, following Benacerraf, makes justifying the reliability thesis impossible. Mathematicians in general believe, for example, in set theory. Moreover, they think that the axioms of set theory are true. Now what is the connection between the two? If the axioms of set theory are true, why are we justified in believing them? What in those abstract axioms can cause the mathematicians' belief in them?

The question is an interesting one. After all, it could very well have been that we *do not* believe in them. The axiomatizations, whether they refer to anything or not, are human creations. Had

history unfolded somewhat differently, we could believe in different axiomatizations. Indeed, not all mathematicians want to use the same axiomatization of set theory even now. Unlike the Platonists and intuitionists argue, true mathematical theories do not seem to force themselves upon us, which is also what Benacerraf's dilemma points out. Mathematics may be something else on the side, but it is also a creative process, a human endeavour. We cannot think that we have *necessarily* come up with the current set of beliefs in mathematics and on these grounds the reliability thesis can indeed be questioned.

However, we are asking the wrong question here. The real question does not concern us arriving at the mathematical beliefs – it concerns us coming to widely *accept* some beliefs. Mathematical discovery (or invention) is a complex phenomenon that is not really essential to our task at hand. For all we care, we could even grant that mathematical creativity is simple arbitrary guessing. But can we ever even begin to believe that mathematical *justification* is only arbitrary guessing? The justification may or may not follow the causal theory of knowledge. That question is open to interpretation, although Benacerraf's dilemma may have enough power to suggest the latter option. However, it is the justification – not the discovery – that we should be concerned with. From all the competing axioms, why do we choose to believe in some rather than others? This is once again the question of theory choice, and it is just about the most important basic question we have in the philosophy of mathematics.

Now that we have moved to justification, the deflationist position of Field shows itself in all its implausibility. Taken to its logical conclusion, it must imply that we cannot offer any outer justification for our mathematical theories. If we did, we would have some theory of reference for mathematical entities. After all, even the simplest instrumentalist justification "this theory works in practice" seems to tell us a whole lot about the connection between mathematical theories and the physical world. If nothing else, it would mirror our thinking about scientific theories in general. Of course we could use different terminology for evaluating theories, for instance words like "acceptability" or "usefulness" (like Field

1980, p. 15) but the situation would not be essentially different from calling some theories true and others false. Acceptability and usefulness by themselves, without any explanations behind them, look like empty concepts. Philosophically, it cannot be satisfactory to say that we prefer theory **T** to theory **S** because it is more useful in physics. The real question is *why* **T** is more useful than **S**, and why this reason is essentially different from what we call truth?

The scientific status of mathematical theories is a matter of debate both in the philosophy of mathematics and the philosophy of science, but in any such account we could hardly remain nominalists in mathematics and realists (or some other alternative) in physics. That is one lesson we have learnt from Quine and the discussion following him: for a strict empiricist (or nominalist) the abstract entities in mathematics are not fundamentally different from the abstract entities in physics. If we wish to retain abstract entities in physics, we cannot easily justify dispensing with those of mathematics. In case we stick to concrete entities in physics, we lose concepts like electrons, force, weight and temperature. I do not want to advocate a Quinean view of science here, but if one rejects abstract concepts, it is hard to differentiate between electrons and the mathematical functions needed to describe them. In fact, depending on the interpretation, we can lose everything: how clear are the references of *any* of our terms, mathematical or non-mathematical, scientific or non-scientific? Facing this, one must wonder how far the fear of abstract mathematical entities can take us.

It must be noted that while Quine's philosophy can be useful in clarifying the relationship between the abstract concepts in mathematics and physics, there is another way in which the Quinean theory gives Field's objection too much power. It was Quine's indispensability argument that Field targeted his program against, and that is unsurprisingly the framework in which it works the best. If we only consider the best theories of science (physics, in particular), there remains the possibility that they do not include mathematics in any substantial way. That will most probably never happen in practice, but Field's work with Newtonian physics is certainly not without merit. There is no question that modern theories of physics are thoroughly

mathematical in practice and in notation. But that is a different problem from the one that Field addresses, and the philosophical question should be kept distinct from the practical one. Before Field's project, Newtonian physics was widely thought to be unfitting for a nominalist interpretation. Deep down, mathematics may not be as indispensable to other sciences as we often think.

However, if we reserve the role for mathematics *only* as a part of theories of physical sciences, we do not get the full picture of the practical uses of mathematics. Certainly mathematical applications in other sciences form a highly important field, but there are also practical applications of mathematics that are not theories of any other science. Grouping apples into piles is not based on any theory of physics, but rather on what may be called *directly applied arithmetic*. Of course we would expect our "best scientific theory" to explain it, but that is going way too far into obscure dreams of a single-theory science. As for now, we should be very much justified in requiring a theory to explain why the grouping of apples conforms to the equation 2 + 2 = 4, and not 2 + 2 = 5. Do we want to go into the best scientific theories for such explanations and try to show them to be nominalistic, which is a pipe dream even with the current scientific theories? Or do we accept that the basic arithmetic is not a fiction; that a theory of *mathematics* can also explain the world?

Furthermore, even if we dismiss all such direct applications of mathematics, Field's program does not have all the power it initially appears to have. It cannot be enough for Field to show that the established mathematical theories of science can be translated into non-mathematical ones. That would only show that the mathematical *notations* are not necessary for science. In order to make a truly convincing case, Field would need to show that we could *arrive* at equally useful scientific theories without mathematical tools. Here Field's sample case could not be worse: the historical fact that it was Newton's invention of calculus that

made the theory of mechanics possible for him speaks volumes to the exact opposite direction.[170]

This is the framework in which Field's nominalism should be examined. The whole premise of fictionalism is extremely problematic, far beyond the problems considered in this work. As I have argued here, the mere implausibility of our time-tested feats of mathematics being simply arbitrary fiction should be enough to refute Field's deflationism. But nothing has been said so far about the *inherent* problems in Field's theory. Shapiro (1997, pp. 219-228) has made a number of important remarks on them. Two in particular stand out. The first one is on Field's philosophical explanation for his "modal fictionalism" (meaning, roughly, that mathematics is a fiction that *can* be created for use in science). The second one concerns Field's main technical project, his nominalist work of showing the conservativity of Newtonian mechanics over mathematics.

The first remark is of utmost importance: Shapiro points out that everything in the fictionalist account can be translated into realist terms. If we are expecting our new modal account to provide a solution to the ontological and epistemological problems, surely it cannot be directly translatable into the realist terminology. As we will see, this is a powerful objection that will follow us wherever we go with nominalism. In the most naïve case we could replace "there exists *A*" with "it is possible that *A* exists". In more sophisticated versions it can be replaced by "we can construct *A*" or "if there exists a mathematical system **S**, there exists *A*". But all these are simply translations of the first, seemingly realist, version. All the instances of *A* perform

[170] In addition, one should remember the considerations in Chapter 4.1 of this work. It seems particularly important to me that Field's approach looks thoroughly mathematical, even if numbers are not used in it. Of course this is not to even fully concede that Field has managed to show mathematics to be conservative over scientific theories. Even his treatment of Newtonian mechanics is not without its problems (see Shapiro 1983 for the problem of Gödel sentences in Field's system), and that is concerned with 300 year-old mathematics developed particularly for a theory of physics. Many of the mathematical tools of current science, especially the statistical and probabilistic ones, seem much less likely to be conservative.

equivalently in mathematical theories. It does not sound at all convincing that all the metaphysical and epistemological problems in the philosophy of mathematics are easy enough to solve by mere translation.[171]

It seems quite clear to me that the ontological commitments of phrases "there exists *A*" and "it is possible that *A* exists" are in principle one and the same. For *A* to exist, there has to be a possibility for it to exist. It is hard to see how the other versions of nominalism escape this problem any better, without any theory of reference. If a modal account is to get anywhere beyond translation, it would need to explain *why* it is possible to construct *A*, and do this in a way essentially different from realism. The situation is similar for the other versions of nominalism. If the nominalist strategy is to think in purely formalistic terms, "is possible" would mean simply that "*A* does not cause a contradiction". But *why* does *A* not cause a contradiction? It must be because we have selected our axioms and rules of proof in a certain way. Here we come again to all the problems of theory choice, reference and arbitrariness that we have discussed earlier, not to mention the new problem of unprovability of consistency in arithmetical systems, due to Gödel.

The second problem that Shapiro (ibid., p. 227) points out is that Field uses mathematical theories, namely set theory, to prove his point. In other words, he uses mathematics to show that mathematics is a fiction. The logical difficulty here is evident, and if developed far enough, Field's theory would almost certainly run into the same kind of trouble as Kripke's and Hintikka's definitions of truth. Many other problems would also arise. In

---

171 The one advantage I see with such translations is that, strictly speaking, a realist phrase "there exists A" is *false* in a fictionalist account, since mathematical objects do not exist. While this is certainly Field's (1980) view, I work here under the interpretation that Field's point of view is better understood as a case of extreme formalism, where mathematical sentences are considered to be meaningless rather than false. In the case of semantical arguments, for example, all Field's considerations over the consistency of PA and other such matters would seem rather irrelevant if PA is considered to be false.

order to be sure that the nominalist interpretation is truly non-mathematical, we would need to define what mathematics is. But can we define mathematics in any meaningful way without the use of mathematics? We remember that Field's approach was to present science without *numbers*, but science completely without mathematics seems like a much more difficult task – while just as relevant.

In addition, since Field uses mathematics in his account, one would wonder how we could use parts of fiction to prove an essentially ontological point. Perhaps this latter problem is one that Field could avoid, but we should know by now how difficult all these kinds of approaches are. Ultimately, Field would have to solve the monolingual speaker problem we discussed in Chapter 5. But this should not matter a great deal at this point, as it is certainly not the most difficult problem in Field's theory. That role is reserved for the question of reference and theory choice, and their inevitable only truly deflationist counterpoint, arbitrariness.

6.5 Modal reconstructivism

Field's nominalism – at least under the above interpretation of it – can be called *destructive* nominalism, as it is aimed to abolish mathematical entities from philosophy, but not to provide a surrogate solution. The other type is *reconstructive* nominalism, according to which we must construct mathematics in other ways to avoid all the ontological and epistemological problems. As I see it, the basic motivation for reconstructive nominalism comes from all mathematical sentences like "there exists *A*" being strictly-speaking *false* if there did not exist any mathematical objects. Mathematics might not be true in any substantial sense, but it cannot be accepted that a great part of our accepted mathematical sentences are false, either.

The most appealing form of reconstructivism is the one proposed by Hilary Putnam and developed by Charles Chihara (1973 and 1990), among others. Putnam (1967, pp. 297-301) presented the idea of translating our "Platonist" language of mathematics into the concepts of modal logic. The basic idea is to

replace phrases like "there exists x" and "for all x" with modal terms like "it is possible to select x" and "necessarily x", usually written $\Diamond x$ instead of $\exists x$ and $\Box x$ in place of $\forall x$. Similarly, functions and other mathematical objects are not thought to exist, but to be *possible* constructions. Both ways of doing mathematics, Putnam argued, are equivalent in the same way that the wave and particle interpretations of the electron are equivalent. This way the language of mathematics would be free of realist allusions to the actual existence of mathematical objects, and yet nothing from the "old" mathematics would be lost.

What is the motivation behind such reconstructive programs, if they do not aim to change anything of substance in mathematics? The main reason seems to be a linguistic one, but with ontological results. Most mathematicians practise their trade *as if* realism were correct: they have no problem using phrases like "there exists", even if many of them are not basing their use of language on any philosophical theory. This is Shapiro's "working realism", and it gives the practice of mathematics a seemingly realist, even a Platonist flavour. According to the proponents of reconstructivism, much of the popularity of realism in the philosophy of mathematics follows from using such realist phrases in practice. With a change in language realism would supposedly lose much of its appeal, and the philosophical questions concerning mathematics could be examined without the conceptual bias toward realism.

Chihara's (2005, pp. 499-500) position is perhaps the best-known modern version of modal reconstructivism, and it is similar to Putnam's. Instead of existential quantifiers, we should use modal quantifiers of the type "it is possible to construct". This project Chihara calls "Constructibility Theory". His motivation (ibid., pp. 507-508) is to defend the use of mathematics in science without reference to truth, that is, to show that we can justify the inferences of mathematics on a nominalist basis. In this work I have used the anti-nominalist argument that mathematics cannot be an arbitrary fiction since it plays such a crucial role in applications, both scientific and direct ones. Chihara wants to show that the same mathematics can be constructed in a nominalist fashion, and the applications of mathematics in no way depend on

mathematics being *true*. Basically, in the terminology here, this is equal to trying to find a nominalist answer to the problem of theory choice. If the Constructibility Theory can achieve everything that realist mathematics does, Chihara argues, the applications of mathematics cannot be used as an argument for realism – or truth – in mathematics.

Another similar position is the modal structuralism of Geoffrey Hellman (1989). Structuralism is the philosophical view that rather than mathematical objects, mathematical *structures* exist. This approach has a lot of advantages. Instead of asking what the natural numbers are, and running into problems of definition, the structuralist can say that natural numbers are only places in the natural number structure. As long as a number serves the same purpose in that structure, it does not matter how we define it in, say, set theory. Now the ontological question follows us to structuralism. Do these structures exist? Shapiro (1997) is a proponent of a realist, *ante rem*, structuralism. But according to Hellman, we can avoid the ontological difficulties of realism in modal structuralism. Instead of saying that the natural number structure exists, we should say "it is possible that the natural number structure exists" or "it is possible for us to construct the natural number structure". Hellman's (2005, p. 553) formulation is a bit different, and technically more sophisticated, but this is the main idea.

Whatever appeal these philosophies have, there is of course the familiar and disturbing matter in the whole modal reconstructivist project, whether it is Chihara's or Hellman's: the new basis for mathematics is essentially just a *translation* of the realist one. Shapiro (1997, p. 228) has pointed out the implausibility of a mere change into a vocabulary of diamonds and boxes solving the problems in the ontology of mathematics. Quite clearly, Chihara's "it is possible to construct" is a simple translation of "there exists". But if the languages are equivalent, perhaps it was not the realist language that was to be blamed for the problems after all? It is hard not to agree with Shapiro here. When we take the reconstructivist language at its face value, of course we must agree that it is *possible* to construct all the mathematical objects. That is just what mathematicians have *actually* done: mathematics did not

simply appear to human beings. All the notations are human inventions, and this is the case even if we hold there to exist a realist basis for mathematics. We have endless ways of formulating new vocabularies to construct mathematical objects and the philosophical problems cannot simply come down to a choice between equivalent vocabularies.

The real question is why we have constructed the mathematical objects the way we have? Of course it is always possible to translate theories that *already exist*. The reconstructivist program is not about creating mathematical theories; it is about translating the existing ones, and for translation it is irrelevant whether mathematical objects exist. No matter what the ontological status of realism may be, the theories would be translatable into reconstructivist terms. The existence of mathematical objects and the objectivity of truth-values for mathematical sentences have nothing to do with them being translatable into a non-realist terminology. Indeed, we could have absolutely certain knowledge about the existence of mathematical objects, and we could *still* make the translation into any number of non-realist languages. But we should not be looking for a translation: whether mathematical entities exist or not may be irrelevant terminologically – but it cannot be that *ontologically*.

One must remember that the case I have made against nominalism never relied on us using realist language in mathematics. Just like no proper argument for realism can be based on us using phrases like "there exists x", no argument *against* realism can be founded on such phrases being translatable into a non-realist terminology. We could do the proposed translation into modal terms, but nothing would change. Indeed, had mathematics originally developed in a modal language, we would not think of making a case for realism by showing that all the modal terms are translatable into realist ones. This is not a matter of words and symbols, and translatability has got nothing to do with the philosophical problems involved. The real question the nominalist faces is that of theory choice: *why* we have the mathematical theories we have. Could mathematics have *developed* in a truly nominalist fashion? It does not matter that we use words like "exist" instead of "is possible to construct" – that can always

be changed. But why do we think that *modus ponens* is a valid rule of inference, the induction axiom of PA is valid, or – indeed – that 2 + 2 = 4? So far this has caused insurmountable problems for the nominalist since without any outer reference our choice of mathematics seems to be completely arbitrary. The choice of language has got little or nothing to do with it.

It seems that the only way to make the nominalist argument powerful is to lose this spirit of translation and show that the nominalist can achieve something non-realist with his program. If he is going to end up with the same theories of mathematics via translation, one has to question the role of *nominalism* in the argument. Indeed, if the exact same argument can be used for translations like "there exists *a*", "it is possible to construct *a*" and "it is possible that *a* exists", one is bound to wonder whether it is really modal constructivism and nominalism we are discussing or just the possibility of translation of quantifiers. In this way the non-translational programs, like intuitionism, are more interesting, and their philosophical content much stronger.

It must be pointed out once more that the realist language of mathematics does not make a case for anti-nominalism, and if this has ever been used as an argument, in this work I want to steer clear of it. Rather, the used language is *irrelevant* to the problems of truth and existence of mathematical entities. This cannot be stressed enough, as I suspect that this misunderstanding is the main motivation for programs such as Chihara's. They want to show that the realist language is so deeply entrenched in the mathematical practice that it is inevitable that we start to think realistically in the philosophy of mathematics. This is an important point, to be sure, and not without a kernel of truth. It is not hard to predict that had the language of mathematics developed to be modal, the modal philosophies of mathematics would be more popular. The language *does* matter in that sense. But once we go beyond language into the real essence of mathematics, we must realize that the choice of language is irrelevant, as long as the options are equivalent in terms of translation. Only after that do the real problems of the philosophy of mathematics present themselves – and for the strict nominalist those problems are very

difficult indeed. The translatability of languages is hardly relevant if arbitrariness of theory choice must be accepted.

Considering all that, for the reconstructivist theories there is always one fundamental problem: that of motivation. Burgess and Rosen (1997, p. 60) raise this question. Other than for purely philosophical motivations, why should we want to reconstruct our mathematical theories? If there is something wrong with them, then why do we want to arrive at the exact same theories from a different background? If mathematical entities interpreted as abstract cause us problems, do we have any hope of solving those problems by interpreting those entities to be concrete or modal? It seems like the only motivation is that we have to deal with the intrinsic *philosophical* problems concerning the abstract entities. Abstract entities (or the postulation of them) *work* in mathematics – most philosophers and mathematicians agree on that. According to the Quine-Putnam indispensability argument, we should use abstract entities because without them mathematics and science would not work. However, even when that argument is accepted, the way one deals with it can differ greatly. Quine's position was that such abstract mathematical entities exist. The reconstructivist position is that we should find another interpretation for the abstract entities. This is bound to have a strong flavour of translation in it, but at least it retains all the power of the original mathematical theories. That is why from the nominalist theories the reconstructivist ones are by far the most acceptable ones: they do not *change* anything of value in mathematics. At the same time, that is also why the motivation for them is so hard to see.

Finally, there is one crucial problem that concerns most of the supposedly nominalist endeavours: they are not as ontologically unproblematic as they claim to be. As we know, the ontological commitments of realist mathematics are the main motivation for developing nominalist alternatives. Usually with a crude form of Occam's razor it is inferred that we do not need to postulate abstract entities if their names are enough. But of course the names by themselves are *not* enough, as we have seen. All the nominalists need translations for quantifiers, in the very least. One possible exception to this is Field, depending on the interpretation. If we consider him a strict fictionalist, then we do not need translations.

"There exists *A*" is simply a (somehow) useful fiction about a fictional entity *A*. If we follow the geometric interpretation, we need a translation, something along the lines of "from the points of space-time we can construct *A*." Now exactly how ontologically unproblematic is this? After all, as we have seen, we need to assume a continuum of points of space-time in order to save all mathematics in the geometric nominalist strategy. Not only is this a strong ontological assumption, it is also one that is potentially in conflict with our best knowledge of quantum mechanics. The ontological simplicity seems to be greatly exaggerated.

How about the modal strategies? The basic phrase "it is possible" of course implies a whole universe of possibilities, some of which are actualized while others are not. What is the ontological status of the unactualized possibilities? By the modal strategy, they have to exist – after all, we are making a direct reference to them. If all the possibilities were actualized, we would not need a modal strategy in the first place. But the unactualized possibilities cannot be concrete, by definition. What we must have is a whole universe of abstract unactualized possibilities. The ontological thinness turns out to be achieved by smoke and mirrors. Granted, the ontological burden of the nominalist reconstructions may not be as heavy as in full-fledged Platonist theories. Still, it is not at all the case that these reconstructive strategies are ontologically unproblematic. Only fictionalism is, and it has much graver problems of its own.

6.6 The power of objectivity: Penrose's question

Benacerraf's dilemma is the basic problem for a Platonist philosopher of mathematics. It is commonplace in the modern philosophy of mathematics that any account in the Platonist direction needs to answer this question. As a result – among philosophers – the Platonist philosophy of mathematics has perhaps never been less popular. The epistemological problem just seems insurmountable without appeal to a special epistemic faculty of mathematical intuition. Not many philosophers are prepared to appeal to this anymore.

Meanwhile, the nominalists have been let off much easier. Certainly, strict nominalism in the fictionalist sense has next to no epistemological problems. If mathematics is completely our own creation, surely there is no difficulty in getting knowledge of it. However, that kind of nominalism has two at least as grave problems as Benacerraf's dilemma, both of them based on the arbitrariness of theory choice. The first one is the question of physical applications. Almost all sciences use mathematical tools, but perhaps nowhere are they more indispensable than in physics. Ever since Galileo famously stated that the book of nature is written in the language of mathematics, physics has been a thoroughly mathematical science. The Quine-Putnam indispensability argument carries a lot of weight: it seems that physics cannot be properly practised without mathematics. The main objection to this, of course, has been Field's science without numbers. Much has been said about Field's project, but one fact about it is beyond dispute: nobody practises Fieldian physics. Even if Field's nominalistic physics were in principle possible, in practice mathematics has as big a role in physics as it has ever had. For the fictionalist, this role must be explained. In Hao Wang's (1974, p. 239) words:

> …the close connection to the physical world is an essential feature which separates mathematics from mere games with symbols. Mathematics coincides with all that is the exact in science.

It is all too easy to claim mathematics to be fiction and then explain this connection by praising it as "useful fiction" – indeed, noting that mathematics is useful is rather redundant.[172] The connection between mathematics and our knowledge of the physical world is

[172] At the first glance, terms like "useful fiction" may seem like an ontologically unproblematic way of rescuing the applications of mathematics. However, on a closer look one realizes that *anything* can be dubbed useful fiction in the same manner: we can be fictionalists over the general theory of relativity and explain its success as it being useful. But of course this is not an explanation at all, and no serious philosopher of physics would use it as an argument. I fail to see how the situation is different when it comes to the philosophy of mathematics.

something a philosophical theory of mathematics must *explain*, not just casually state. Hence Field's nominalist program, even if successful, could not solve the problem of physical applications of mathematics for us. If mathematics is simply a fiction, why do certain theories work much better in physics than others? It is a question that has to be answered, and the fictionalist answer of arbitrariness cannot possibly satisfy us. However, this is not a problem only for the fictionalist. The applications in physics are also a problem for the Platonist. In a variation of Benacerraf's dilemma, we can ask how the physical world can work according to (in part) abstract non-physical laws. The problem of physical applications seems to rule out both of the extreme positions in the philosophy of mathematics.

The second problem of nominalism, however, is something that Platonism answers. It is also something that is remarkably often simply ignored. Keeping in mind how notoriously unreliable and inconsistent we as human beings are, how can it be that in mathematics we find such robustness, clarity and consistency? Roger Penrose (2004, pp. 12-13), for example, has explicitly presented this problem. As he is one of the few contemporary philosophers who can be called a Platonist, I will call the problem *Penrose's question*.[173] The question is more pertinent than may seem at first. One obvious answer could be that mathematical standards are all just man-made norms. But this by itself is unsatisfactory when we examine the kind of norms that we have. Man-made norms exist, and we know quite a bit about their nature, whether they are ethical, political, rules of games or some other type. For most people familiar with mathematics, the robustness of mathematics is quite obviously something different. No matter how one looks at it, practising mathematics just does not appear to

[173] Of course Penrose has not been the first to ask this question, but Benacerraf was hardly the first one to recognize the epistemological problems of Platonism.

be comparable to solving a Sudoku puzzle.[174] Both have explicitly laid out rules, both can be subject to the same kind of rigorous study, but unlike Sudoku, mathematics seems so clearly to be something other than an arbitrary game we have created.[175]

Forgetting for a while all the ontological and epistemological problems, Platonism seems to have a good answer to that robustness. Indeed, this is probably the reason why most mathematicians are working realists. For most mathematicians, the subject matter of mathematics is something that they research and discover, not something that they create. The varied philosophical arguments notwithstanding, for many mathematicians it certainly *seems* to be the case that mathematics is objective.[176] This is no small matter. Working realism is a large part of the convention of mathematics and the nominalist cannot dismiss the way we have arrived at our mathematical knowledge. In a philosophical account of mathematics, one cannot just throw away the ladder after climbing it. After all, we are looking for an explanation of mathematical knowledge. How can mathematics be just a subjective endeavour when there is so much in its practice and results that points to objectivity? Whether this objectiveness is an illusion or not is another question; but the problem of apparent objectivity is definitely something that the nominalist must

---

[174] One entertaining counterargument against fictionalism is that mathematics is not make-believe like "playing Cowboys and Indians". While altogether more amusing, this analogy has the problem of including a clear reference to something (formerly) existing, as well as being overall a bit unfair to the fictionalist. The philosophers of mathematics should be grateful for the current popularity of Sudoku in providing a perfect analogy to the extreme formalist mathematics.

[175] Of course the study of Sudoku patterns can be mathematical, as well. The point here is that not *all* mathematics is like that. Moreover, in the case of mathematical studies of Sudoku solutions, mathematics clearly has an outer object, although man-made and in its finiteness obviously different from most of mathematics.

[176] In the lack of research on the actual philosophical beliefs of working mathematicians, I cannot make the stronger claim that *most* mathematicians think this way. As a conjecture that sounds highly plausible.

answer. Any answer resembling arbitrariness is thoroughly unsatisfactory.

Field (2001, pp. 315-331) has tried to explain this objectivity with the concept of *logical* objectivity. His contention is that what we mean by the objectivity in mathematics is actually only the objectivity of rules of proof, that is, *logic*.

> …logic, hence mathematical *proof*, is fully objective. And because proof is so important in mathematics, this concedes most of what we may have had in mind in calling mathematics objective. It ought to be obvious that if mathematics is objective only in this sense, then the link between mathematical objectivity and mathematical objects […] is wholly illusory: you don't need to make mathematics actually be about anything for it to be possible to objectively assess the logical relations between mathematical premises and mathematical conclusions. (ibid, p. 317, Italics in the original).

So according to Field, the objectivity that seems to force itself upon us when dealing in mathematics is only that of the accepted rules of proof, which can be thought to be fully objective. Obviously this is a very weak form of objectivity, only enough for the conclusion that mathematicians universally accept the formal derivation procedures.

However, objectivity defined in this way does not really tie mathematics into anything. It is a good question whether this can be called objectivity at all, or just a form of conventionalism. There are two questions in particular that are immediately raised. First, Field speaks about the rules of proof, but what is the status of axioms? Second, if we are truly anti-realist in mathematics, why should we accept that the rules of proof are objective? The first question is a very difficult one, because contrary to my conclusion about his position, sometimes Field does not actually seem to accept arbitrariness. In fact, he (ibid., p. 322) accepts that "considerations of utility play a role in our selecting some mathematical axioms over others". My immediate reply is one that Field anticipates: are we not bringing truth in through the back door by appealing to utility? Field claims that this counterargument fails since the concept of utility is relative to the

purpose at hand. To give an example, he mentions that the different axiomatizations of set theory can be useful in different situations. However, this goes profoundly against mathematical practice, for there is no denying the remarkable consensus on the choice of axioms in practically all areas of mathematics. If considerations of utility come into the choice of axioms, practice tells us that this choice has *almost always* been the same. In fact, Field's example seems to make the exact opposite point from the one he wanted: *if* there were great diversity in the choice of axioms, Field might have a point. But as it happens, this could not be farther from the truth. Certainly there is not full consensus, but there is more of a consensus than in almost any other human activity. Either the appeal to utility must be resisted, or we must admit something objective into the choice of axioms. In the first case we end up with arbitrariness, in the second one *truth*.

The second question is not any easier for Field. Why can rules of proof be considered objective if axioms cannot? While we are at conventionalism, it would seem possible to follow Wittgenstein (1983) into thinking that *nothing* in mathematics is objective, not even the process of drawing conclusions with formal derivation procedures. Field (2001, p. 316) dismisses such radical anti-objectivism, but can he remain consistent in denying the objectivity of axioms? Of course Wittgenstein's position sounds absurd to most of us, but it *is* the logical conclusion of the radical conventionalist line of thinking. In fact, when we remember everything we know about the disagreement over rules of proof – intuitionism, many-valued logics, etc. – the objectivity of rules of proof is not at all obvious. The objectivity of certain accepted rules of course holds, but that is a trivial matter. The objectivity of logic as such is a strong statement, and it does not go well with Field's fictionalist program.

When it comes to the subject matter of this work, by all the considerations above Field's position is simply untenable. He wants to rid mathematics of objects but retain objectivity. He wants to include utility as a criterion of theory choice, yet he refuses any role for truth and reference. The motivation for this project is obviously to minimize the ontological commitments, which is an understandable goal. But one can only go so far with this strategy,

or one must go all the way. Either mathematics is arbitrary, or the objectivity must be tied to something. As we have seen, the latter option can refer to a number of positions – but all of them give grounds for truth and reference in the philosophy of mathematics.

6.7 The power of nominalism and potential ways out

All the variations of nominalism in mathematics seem to run into one of two difficult problems. Either they make mathematics arbitrary or they are translations of the old "realist" mathematics. The motivation for such translations is questionable and, as we have seen, the ontological problems are not much easier to deal with. Moreover, it is not at all clear that such strategies are really nominalist, at least not in the strict sense required here. Of course we should not expect anything more from a solution based essentially on translation. The reconstructive nominalist strategies do not end up with much appeal, as far as the subject of this work is considered. However, the first problem is obviously the more serious one. As I have argued, nominalism of the fictionalist, the only truly deflationist type is repudiated for good because of arbitrariness.

Yet it must be asked what makes nominalism such a common (although, it has to be remembered, still mostly marginal) view in the philosophy of mathematics? I can see the motivation arriving mainly from the appeal of nominalism *elsewhere* in philosophy. Originally, nominalism was targeted against Plato's idealism in ontology. It demolished the ontologically difficult world of ideas, and moved the focus to the empirical world. From a modern point of view this was obviously a step forward. Instead of "chairness", we like to think that the individual physical chairs exist, and what connects those individuals is similarity in the use and design, which exist due to people. In the case of chairs, few would still claim that there exists an abstract idea of a chair, completely independently of us. When it comes to natural objects, that development was achieved by science. In place of the problematic concepts like "catness", we can move into the domain of genomes and have an unambiguous and exhaustive nominalist replacement

for the earlier universal concepts. This way, when it comes to these kinds of classical examples of the universal-nominalist debate, the philosophical problems have been largely resolved.[177]

Chairs, cats and the like, however, are not the only types of entities we deal with. Indeed, in a strict sense, it is highly anthropocentric to claim that even chairs exist. Chairs are constructed of molecules, which are constructed of atoms, which are constructed of protons, neutrons and electrons, the former two of which are constructed of quarks. All of these are ways of explaining phenomena in physics, and all of these can be modelled with the yet more fundamental micro-level laws of quantum mechanics. Which of these objects exist in the strict ontological sense? It is not my purpose here to go into all the problems that nominalism presents to us, but I do want to emphasize how nominalism even in physics is ultimately nothing like the neat theory that "physical objects exist". Physics does not tell us unequivocally what exists in the nominalist sense. Seemingly theoretical concepts like "force" exist essentially in the same sense as seemingly empirical ones like "atoms": they are all parts of theories about explaining measurable phenomena. In addition, of course the postulated entities and their relations are not all there is to physics. The other part is formed by the mathematical theories that play a crucial part in the explanations. Perhaps we do not need to go as far as Quinean holism, but it certainly does not seem easy to completely distinguish between the theoretical and empirical content in physics, at least as far as their ontological status is concerned. Mathematical theories are a part of physics, and the distinction between mathematics and physics in those theories should not be thought to be a straight-forward one, either.

On the other hand, philosophically we seem to have a clear idea how such a distinction is made. Mathematical objects are *abstract*; they are non-temporal, non-spatial and causally inactive. There are other types of non-temporal, non-spatial and causally inactive objects, but none seem to be like the mathematical ones in every facet. Works of art (a novel, for example, as opposed to all the

[177] Or – to be safe – at least they do not look like the kind of fundamental problems they once were.

printed copies of it) are non-spatial, but they are clearly temporal. If we reject extreme formalism, we cannot intelligibly ask when 2 + 2 started to be 4, while from each work of art we can clearly ask when it was created. Species of flora and fauna are causally inactive (only their representatives are active), but evolution and extinctions make them temporal. Even ethical concepts seem to have a temporal context: according to most people they only arrived with the development of human of beings. Whether that is true or not, certainly they cannot be thought to exist before life evolved sufficiently. However, here we are quickly closing the gap to mathematical objects. While mathematical objects can seem to be non-temporal, obviously we can only inquire about them if we exist and have a propensity toward mathematical thinking. If there is something in human beings that makes mathematical explanations natural and successful to us, could we ever be able to distinguish this from an objective feature of the outer world? In both cases we would have objective mathematical truth and the mathematical entities would seem to be abstract. Yet in this anthropocentric view the ontological status of mathematical entities is much less problematic. They do not exist objectively in the outer world, but they *do* exist objectively in the way human beings acquire and process information of the world.

In this sense, the empiricist approach developed by Kitcher has a lot of potential, but only if we apply it to our most basic mathematical knowledge, such as simple arithmetic and geometry. Simply looking at the education of young children, it seems clear that empirical knowledge forms a central part of primitive mathematical thinking. Certainly, from looking at some sophisticated fields of mathematical inquiry, such as topology, one is tempted to see it all as fiction. Up to some point, that could indeed be the case.[178] But sophisticated mathematical theories are not self-standing: they have developed over time on top of more basic ones. Complex numbers, for example, were formulated to

[178] For example, in the empiricist account, *infinity* is one concept that could turn out to be fiction, strictly speaking, if it could be established that the universe is finite. This does not concern only the more sophisticated theories of mathematics, but basic arithmetic as well.

resolve one particular limitation of real numbers[179], just like real numbers were formulated in order to account for irrational numbers. That way, it could be argued, every step forward in mathematics brings us further away from the original pre-formal – and possibly empirical – concepts. As a result, mathematics may start to look more and more like fiction.

However, it must always be remembered that (almost) all new mathematics is based on older theories. There is a continuing development from the primitive first steps of mathematical thinking into modern theories and their highly formal presentation, and in this way none of even the most sophisticated mathematical systems are completely self-standing. Obviously a complete account of mathematical reference would need to specify the ontological status of all such theories, but we cannot go into such pursuits here. For the purposes of this work it is sufficient to note that even the slightest general human propensity toward mathematical thinking can be considered to provide the basic mathematics with an objective reference. This could mean, among other things, categorizing observations into simple logical structures, geometrical constructions or quantities. Suddenly the ontological demands are vastly lighter, while – returning to the direct subject of this work – the same account of Tarskian truth can still be applied.

I will shortly present a tentative outline of such a project, but it should not be thought to form the basis for any of the main arguments presented in this work. There are arguments for and against this kind of objective anthropocentricity. On the one hand, the fact that different cultures, the Mayas for example, developed mathematical theories independently of our culture seems like evidence for the theory. On the other hand, there are cultures where mathematical thinking did not develop. In this sense mathematics does not seem like the kind of universal human ability that language, for example, is. These are important questions, and empirical research can give us a lot of information on them – even though ultimately the question is bound to remain largely philosophical. The idea here is merely presented as a

[179] That of negative real numbers not having square roots.

reminder that not all anti-nominalist theories need to be realist in any Platonic sense. It has been noted (by Chihara 2005, p. 512 for example) that anti-nominalism without any positive development of alternatives is hardly satisfactory. There is some truth in that, although the implicit idea often seems to be that anti-nominalism suggests Platonism. However, if nominalism fails, we *can* make philosophical conclusions even without presenting explicit alternatives. It is not completely satisfactory to leave open the question of which alternative to anti-nominalism we should pursue, but I hope that the approach taken here is enough for the theses in this work. After all, the subject matter here is Tarskian truth, which is a very widely applicable theory. It only requires that we can speak of some reference for mathematics. We can be Platonists, empiricists or structuralists in mathematics – and as we have seen, also modal reconstructivists and nominalists of the geometric strategy – and still be proponents of Tarskian truth.

It has not been the purpose of this chapter to attack nominalism with other theories of mathematical knowledge. My aim has been to show that strict nominalism fails on its own: whatever we want from our philosophy of mathematics, it cannot lead to the conclusion that our preferred axioms and rules of proof are merely completely arbitrary conventions. We must always remember what kind of axioms and rules of proof we are talking about. From reading the nominalist literature one almost gets the impression that our established mathematical axioms are somehow elusive and contingent and we could have chosen their negations just as well. To see the evident fallacy in this, one can return – for example – to the Peano axioms of arithmetic and try to convince himself how we could accept the negation of one of them.

To such claims of self-evidence of axioms one is usually answered with non-Euclidean geometries, where Euclid's parallel axiom is abandoned. This counterargument is difficult to understand. After all, it was *not* the case that mathematicians did not know that there can be non-Euclidean geometries. By drawing lines on a balloon everybody realizes they have encountered one. Euclid was not shown to be *wrong*: the parallel-axiom still works perfectly well in plane geometry. What was realized was that there are also other geometries of interest. In a wonderful triumph of

physics and mathematics one of these geometries has been shown to be the best one to explain our universe on a macro level. Mathematics developed into new directions but Euclid's geometry works as well as ever for planes and the Euclidean three-dimensional spaces, which are still used as the physical model in most applications.

Wherever we look, mathematics is on a much stronger basis, and much more self-evident than some of the nominalists like to admit. The problems we have in the fundaments of mathematics are over-magnified into something completely fantastic. Talk of a crisis in mathematics is not uncommon when dealing with concepts like incompleteness. Yet all such "crises" have been found *within* mathematics, and have in fact worked to clarify the formal mathematical theories. In addition, mathematical research has continued as before, and mathematical applications in science have proved to be as useful as ever. That usefulness is of course the one thing that the nominalists admit, but it is also something they cannot explain – at least if they want to remain nominalists in the strict sense.

6.8 Ontology of mathematics: an alternative outline

What, then, are mathematical objects? The following is no more than a brief outline, meant to work as a tentative example of a non-Platonist, non-nominalist alternative. It must not be thought that the main arguments of this work depend in any way on the proposal here. However, it is an outline that is immediately applicable to the arguments here with minimal ontological burden. As such, it should work to remind us just how wide a field objectivism in mathematics is. But any objectivist philosophy will fit with the account of mathematical truth given here, and my proposal is just an example.

After that clarification, let us begin. If mathematical objects are anything, it seems likely that some form of structuralism is the answer. As we remember, in structuralism numbers and other mathematical objects are thought to be places in structures, rather than existing independently. Structuralism, originally proposed by

Michael Resnik (1981 and 1982) and later developed by Shapiro (1997), provides a solution to the problem of defining concepts like natural numbers, tackled first by Frege. One famous example of the problems involved with such definitions is the so-called "Julius Caesar problem". What reason do we have of saying that a natural number is the same as something and different from something else? For instance, how can we say that "Julius Caesar = 2" does not hold? Another example is the two ways of formulating natural numbers in set theory. Von Neumann presented the formulation that zero is the empty set ϕ, one is the set formed by the set of zero $\{\phi\}$, two is the set formed by the sets of zero and one $\{\phi,\{\phi\}\}$, three is the set formed by zero, one and two $\{\phi,\{\phi\},\{\phi,\{\phi\}\}\}$, four is the set $\{\phi,\{\phi\},\{\phi,\{\phi\}\},\{\phi,\{\phi\},\{\phi,\{\phi\}\}\}\}$ and so on. In Zermelo's account, we get the successor to each natural number by forming a set out of it. In this account, two is the set $\{\{\phi\}\}$, three is the set $\{\{\{\phi\}\}\}$ and four is the set $\{\{\{\{\phi\}\}\}\}$. The obvious difference between the two accounts is that although both definitions are arithmetically equivalent, in Von Neumann's account $2 \in 4$, while in Zermelo's account $2 \notin 4$. The structuralist can avoid this difficulty because both formulations of 2 have equivalent roles in the natural number structure, and only the place in the structure matters. Similarly, if Julius Caesar somehow performed the same role in the natural number structure, it would be equal to 2. [180]

The Julius Caesar problem, plus Zermelo's and von Neumann's conflicting but arithmetically equivalent definitions of natural numbers are very strong arguments for structuralism. But there are even simpler ones, like the fact that we denote the natural number between 4 and 6 with all kinds of different names, including "5", "101", "V", "five" and "cinque". All these can be parts of equivalent (seemingly) self-standing mathematical theories, and outside of practical questions of simplicity and clarity of notation there is nothing to tell them apart. However, in different structures the names notate different objects. We cannot replace the decimal name "5" with the binary name "101" without changing the names

[180] For more on the Julius Caesar problem, as well as von Neumann and Zermelo, see Shapiro 1997, pp. 78-81.

of all the other numbers in the structure. Hence it is the place denoted in the name-structure, not the name itself that matters.[181]

In fact, this is the case especially if we lean toward nominalist mathematics, where names are not thought to have outer references. With Tarskian semantics we obviously have an easy way to deal with this. In a metalanguage, like we are doing on this page, we can discuss the different notations of natural numbers. But how can we in the nominalist mathematics say that the numerals "5" and "V" denote the same object? Of course we have to think of the object as fictitious and think of all the different names of "5" as a circle of translation where nothing refers to anything non-fictitious. Without a structuralist account this would be disastrous. In a strict nominalist account, where only *names* exist, binary mathematics would be different from decimal mathematics, even though they are completely equivalent in arithmetic content. So it must be the case that it is not really a name, but a "place in a name-structure" that the nominalist thinks numbers to be. That is also the case, *mutatis mutandis*, when it comes to Platonism. It is absurd to think that our current notation of numbers somehow refers to the existing natural numbers while other equivalent ones do not. Also in the realist accounts, structuralism seems to be the way to go. Indeed, the Zermelo-Von Neumann example has the maximal force when applied to Platonism.

So we should focus on the structuralist accounts. In which way do these structures exist and how do we get information of them? Here everything we have learnt about Platonism and nominalism can be used. Structures cannot be arbitrary fiction, and we wish for something ontologically more economical than existence in a Platonist sense. With this background, forms of naturalism[182] could

---

[181] This point of course holds for all languages, not just those of mathematics.

[182] Naturalism in mathematics is fundamentally the viewpoint that philosophy can never contradict with mathematics. See Maddy 1997 for one kind of naturalist project. Quine is obviously the other famous proponent of naturalism, although his theory does not treat mathematics separately from other pursuits of knowledge.

be very tempting, but they have a tendency of sweeping the ontological problems under the rug by rendering them a part of mathematics. No doubt the practice of mathematics should give us a great deal of information on what exists, but unlike the strict naturalists, I also see a role for philosophy in this. Even if we think that mathematical entities exist, we do not need to think that *all* of them do. In Cantor's (1883, p. 896) famous account of mathematics, every introduction of new concepts is justified as long as they are consistent and defined from earlier concepts. This freedom of mathematical thought, when given a naturalist interpretation, means that everything defined in such a manner exists. However, this is not something I am ready to agree with. When comparing natural numbers to complex numbers, I cannot help but notice a potential difference in their ontological, epistemological and pragmatic status. Natural numbers are used almost everywhere by everybody. When we notice a difference in sets with sizes of two and three, we are using a (perhaps primitive) notion of natural numbers. Animals can do this. Complex numbers, on the other hand, were specifically developed to enable calculations including the square roots of negative numbers. One is not likely to have such a direct comprehension over what a complex number is, or what complex numbers refer to in the world.

I do not mean to suggest that such a difference is necessarily ontological. There is a direct line of definition from sets to complex numbers and it is not easy to see at which point we move from existing objects to the realm of fiction, or indeed whether we ever do. However, in practice many mathematicians tend to hold natural numbers to be more fundamental than complex ones, and hence I do not find the naturalist strategy of lumping everything together to the same ontological category satisfying. I believe that in mathematics we have the potential both to explain existing concepts and create fictitious ones – at least there is nothing *inherently* problematic in that, and the philosophy of mathematics should allow that possibility. An analogy in physics could be the macro world being ultimately a useful fiction while the micro world is the truly existing one – supposing that they can be combined completely. Not every theory of mathematics – even if true – needs to refer to existing objects; there is a lot in

mathematics that cannot be looked at as anything other than conventions.[183] But that should in no way suggest that *everything* in mathematics is a convention, either. As I see it, there is no such thing as the "whole of mathematics" that should be neatly categorized under ontological and epistemological conditions.

In such manner we should concentrate our efforts on the most primitive mathematical concepts, like the natural numbers and the objects of geometry. Here Kitcher's empiricism could be a good starting point. However, Kitcher goes astray when he thinks of mathematics as another empirical science. For instance, mathematical theories are hardly empirically corroborated, at least not in the sense that physical ones are. Showing a statement to hold for every $n < 1\,000\,000$ is mathematically no better than showing it to hold for every $n < 1000$. Nevertheless, Kitcher does have merit in emphasizing the empirical part of mathematics when it comes to our very basic familiarity with mathematical concepts. In the learning of mathematics this is absolutely essential, yet it is often almost completely neglected in philosophy. Kitcher's notion of mathematics as "an idealized science of operations which we can perform on objects in our environment" sounds very fitting when applied to the learning of basic mathematics. Initially we learn arithmetic – at least in part – empirically by grouping objects and counting. This is of course not truly mathematical in the sense that proving theorems is, but – as I have argued – it nevertheless gives us pre-formal mathematical information. Empirical information does play a role in mathematics, and Kitcher's account can be useful in explaining this. However, this does not need to mean empirical in the sense of passive observation. Mathematical knowledge can develop via trial and error in fitting certain patterns into outer objects. This is empirical, as well, but it can also be distinguished from purely (or at least *more* purely)

[183] One illuminating example is pointed out by Ian Stewart (2006, pp. 162-163): that multiplying a negative number with another negative number gives us a positive number. Certainly there are good algebraic reasons for us to hold that convention, but as Stewart points out, no matter how obvious it sounds and how much sense it makes, it is still difficult to see it as anything else than a convention.

observational activity. Moreover, we need to remember the difference between the origins on mathematical thinking and the sophisticated theories we have arrived at. The former can perfectly well be empirical while the latter retain all that makes mathematics different from empirical sciences.

Although I do not agree with his overall philosophy of mathematics, Ludwig Wittgenstein has the strength of recognizing the origins of mathematics as something distinct from our developed theories of it. Basically, I am ready to agree with passages like the following two:

> All the calculi in mathematics have been invented to suit experience and then made independent of experience. (Wittgenstein 1976, p. 43).

> [25 * 25 = 625] was first introduced because of experience. But now we have made it independent of experience; it is a rule of expression for talking about our experiences. (ibid., p. 44).

It is quite natural to make the hypothesis that the origins of mathematical thinking were something similar. However, Wittgenstein (ibid., p. 22) moves from that into conventionalism, emphasizing how mathematics understood in such manner is not about discovery, but invention. This is not very convincing because the concept of experience is understood in a strangely insignificant fashion. Certainly in a weak sense mathematics *is* invention, but in the same sense physics is invention. Human beings create the theories, and the presented notations can always be understood as inventions. However, if arithmetical theorems help to explain our experiences, just like Newtonian mechanics explain planetary motions, how can we say we are not *discovering* something about the world, or at least about the way we perceive the world? We can think of the truths of mathematics as inventions and conventions – which they both are, in a way – but how are we to explain the fact that they continue to explain our experiences, and not only when it comes to simple arithmetical theorems?

Wittgenstein's approach has appeal when we consider the origins of mathematical thinking, but after that it runs into the problem of arbitrariness. Although mathematical theories do start

to live a life independent of their empirical origins, philosophically we cannot dismiss the fact that these origins exist. In fact, they could very well be the key to understanding the nature of mathematical reference. Although we can learn from both Wittgenstein and Kitcher, we must not follow their pursuits too far. If we acknowledge the role of experience in the development of mathematical thinking, the epistemological and ontological problems are much more manageable than in a Platonist project. In addition, the direct applications of arithmetic and geometry are instantly explained with such a theory. However, mathematics quite clearly has both a subject matter and methods essentially different from those of empirical sciences, and I do not think that any plausible study of the philosophy of mathematics can remain along Wittgenstein's or Kitcher's lines for longer than is needed for the origins.

Now there remains the problem of connecting this "semi-Kitcherian" account with structuralism. How does our operating with piles of apples give us information about an objective natural number structure? For explanation there can be two options: either the world is organized according to such a structure, or human beings have an objective tendency to observe and explain the world according to such a structure. The first *Galileo-Newton* viewpoint is perhaps the route traditionally more often taken. The second one, a form of epistemological naturalism, is a somewhat less studied position. In such a naturalist account, classifying experiences with basic mathematical structures would be comparable to seeing colours: an objective feature that most human beings have in experiencing and describing the world.

Of course many philosophers would say that the two positions could not be distinguished from each other. In such Kantian philosophy we cannot discuss the world outside our categories of observation. When it comes to mathematics, that could very well be true, but the proposed account of mathematical ontology does not depend on it. Both options account for the objectivity of mathematics, and the choice between them would require solving the most basic ontological problem of the philosophy of mathematics – but also the problem of intersubjectivity of observations in general philosophy. The ontological problem is

indeed the most difficult one in an account like the one proposed here, and because of that the naturalist account looks more appealing. Whichever option we pursue, with semi-Kitcherian empiricism some of the epistemological problems – Benacerraf's dilemma, in particular – are answered much more easily than in Platonism. In addition, we do not need to postulate a world of abstract mathematical ideas. However, most importantly, we can still retain objectivity in mathematics.

In conclusion, a semi-Kitcherian account of the origins of mathematics combined with a structuralist and a naturalist account sounds like a very promising alternative to develop. It is not ontologically radical, at least no more than us having objective tendencies for observations is. Yet it gives us the means to talk about objective truth in mathematics. Perhaps, when we develop our ability to explain the brain, we will someday find evidence for such a viewpoint. If there is something in us that makes us think mathematically, it could be detectable in the structure of our brain. However, aside from making the hypothesis that we should look for such structures, that is a question for empirical science to explain. What I have wanted to suggest here is that such a theory sounds perfectly acceptable as a non-Platonist, non-nominalist alternative in the ontology and epistemology of mathematics. The choices are not limited to arbitrary nominalism and epistemologically impossible Platonism.

7. Truth and reference

7.1 Counterfactuals

One interesting argument for Tarskian truth is based on its ability to deal with counterfactuals. Let us return to the basic example of a T-instance:

> "Snow is white" is true if and only if *snow is white*.

Now according to the deflationist, this instance of the T-scheme is just disquotational and the mentioning of truth is redundant: everything in the T-instance is covered by stating "snow is white". But what if we had learnt to use "snow is white" in a very different way, for instance, that it would have the truth condition of *grass is red*? Clearly the T-scheme now gives us the tools to say that "Snow is white" would be false, since "grass is red" is false. The disquotationalist, however, is not equipped with such tools, since all he can state remains to be "snow is white". Strictly speaking, the disquotationalist can only claim that had it been the case that "Snow is white" has the same Tarskian truth conditions as "Grass is red", he would be using the phrase "Snow is white" differently. However, this goes beyond a linguistic or quasi-logical concept of truth. Here Tarskian truth seems to carry more expressive power than its disquotationalist interpretation and it could be suggested as an argument for Tarskian truth.

Naturally, Field (2001, p. 133) objects, stating that:

> In considering counterfactual circumstances under which we used "Snow is white" in certain very different ways, it is reasonable to translate it in such a way that its disquotational truth conditions relative to the translation are that grass is red.

Field thinks that this is the "cash value" of the counterfactual and there is no problem for the disquotationalist. As far as this example goes, I must agree with him. To me it seems obvious that our expressions concerning the outer world are determined by observations on the conditions in the world and we use phrases

like "Snow is white" in a way that suits those observations. When it comes to the general theory of truth, counterfactuals are a matter of much debate, but I do not think that counterfactuals as such are enough to refute disquotationalism. Generally, there is a way to get around them, although it might require a bit more tinkering than is desirable. We might have to give up neat (quasi-)logical theories in order to accommodate counterfactuals, and along the way lose some of the immediate expressive power of the T-sentences. However, if one is an ardent believer in disquotationalism, it certainly sounds acceptable enough to claim that had it been the case that "snow is white" has the same truth conditions as "grass is red", the disquotationalist would not be stating the former (as well as the latter) sentence. Even a moderate belief in our ability to gain knowledge of the world – and use language accordingly – results in that.

In fact, when it comes to empirical sciences, we actually focus more on *justification* than truth. Roughly speaking, science tells us what the world is like, and had the world been different, science would have developed differently. There is nothing controversial about that, and aside from certain philosophical meta-level considerations, the deflationist can continue his pursuit largely unharmed. It is hard to see how a theory of counterfactuals would change the practice and results of empirical sciences. If our use of language does not conform with the empirical findings or new observations demand new concepts, we can rest assured that scientists will make the necessary adjustments. Aside from some proponents of Kuhnian philosophies, most of us should be ready to accept that in empirical science problems are primarily decided by observations – and with regard to observations counterfactuals do not seem to have the same power.[184] Of course Tarskian truth gives us a philosophically convenient way of dealing with

[184] This is not to belittle the effect that hypotheses and theoretical background have on empirical data. My contention is simply that empirical sciences are primarily *empirical*, and the theoretical element alone does not determine empirical findings in any such radical way as some philosophers following Kuhn (1962) have suggested.

counterfactuals, but to make a case for a substantial notion of truth, something more is needed.

To that effect, here we once again find that *mathematical* truth has characteristics of its own and counterfactuals prove to be a much more serious problem for the deflationist. Let us think of the Banach-Tarski paradox (BT) in axiomatic set theory.[185] The relevant T-scheme is:

(I) "BT" is true if and only if BT.

Of course in the deflationist account of mathematics, truth means provability and the T-scheme can be replaced by the mere utterance of BT. Now if we accept the axiom of choice (AC), then BT is provable. But if we do not accept the axiom of choice, BT is not provable. Clearly the provability and hence, for the deflationist, the truth of BT depends on the axiomatizations of set theory that we use. Now AC has been proven to be undecidable from the other axioms of Zermelo-Fraenkel (ZF) set theory, which leaves both options open. In other words, stating that BT is true includes stating that AC is true, and hence the truth conditions of BT are equivalent with the truth conditions of AC. This way, the relevant T-scheme can also be stated as:

(II) "BT" is true if and only if AC.

Now what does (II) mean for the deflationist? Quite clearly it must only say that BT is provable exactly when AC is provable. If we were deflationists and accepted AC, we would think that BT is true. If we did not accept AC, we would think that BT is false. This is of course all just basic mathematical knowledge, and mirrors the way the connection between the Banach-Tarski paradox and the Axiom of Choice is presented everywhere.

Yet, however basic and fundamental mathematical knowledge it may seem to be, it is something that the deflationist cannot deal with. For him, there is no such thing as "accepting AC". We must remember that for a deflationist like Field, as far the subject of this

[185] See footnote 136.

work is considered, formal systems are everything there is to mathematics. Either we have ZF or ZFC, and consequently think that BT is false or true, respectively – but the choice between the two goes beyond formal systems. Most mathematicians currently accept AC and hence think that BT is true. However, the fact is that AC is undecidable from the ZF axioms and if we did not accept AC, BT would be false. Now let us think of the counterfactual case that AC turned out to have the truth conditions of a disprovable sentence of ZF. AC is undecidable in ZF, but outside of deflationism, that does not mean it could not be *false*.[186] In that case BT would obviously be false. With Tarskian truth we can easily state such things and express all the mathematical knowledge we have about the subject. But what can the deflationist do? He can only assert BT or ¬BT, depending on his choice of axiomatization – which is, as we remember, a choice that he cannot express formally. Changes in axiomatizations can never reach the deflationist of mathematical truth. Even if we somehow got every reason to believe in the falseness of the axiom of choice, the formalist committed to ZFC could be none the wiser and would still continue to assert BT.

Although that is a hypothetical example, the general case is not hypothetical. Changes in axiomatizations happen and the truth-values of sentences change accordingly. Take the following sentence of geometry:

(III) "The sum of the angles of a triangle is always 180°".

Obviously the truth-value of (III) depends on whether we accept the parallel axiom or not. As it happens, our best current theory of macro-level physics states that (III) is false, even though it was thought to be a necessary universal truth for thousands of years. Do we not wish that a theory of truth could deal with such interesting – in fact, crucial – parts of mathematical knowledge? Yet in the formalist account of mathematical truth we cannot

[186] Of course the situation is not any easier under the deflationist interpretation, AC being undecidable in ZF (if ZF is consistent), and thus contradicting the law of excluded middle.

discuss truth outside provability. For the formalist mathematician who does not accept the parallel axiom, (III) was true – after all, it was a theorem of the *only* accepted geometrical system of the time – and then became false. The worrying part is that when a revision in the axioms or rules of proof means that previously true sentences can become false, the truth of mathematical sentences turns out to be a function of time.[187]

For the non-formalist all this is very easy to explain. As we gain more mathematical knowledge, we correct our mistakes and admit that what we thought to be true in fact turned out to be false. The truth-value of the sentence itself never changed. The problem does not need to concern changes in axiomatizations, either. When we think of mathematical problems like Fermat's Last Theorem (FLT), this is exactly what happened, the only difference being that in the strict formalist account the sentence was not considered to have a determinate truth-value at all before Wiles' proof. In the deflationist account FLT's truth-value changed as a function of time.[188] In the Tarskian account FLT was true all along, but we just could not prove it. When it comes to mathematics, with its apparent image of a non-temporal subject matter, there should be no question which theory of truth satisfies our intuition better.[189] The deflationist clearly lacks some of the tools that Tarskian truth has; in fact, it could be argued that deflationism never seemed more problematic. Could we ever accept an account of mathematical truth according to which the truth-value of a simple sentence of arithmetic changes as a function of time?

This mirrors the discussion on semantic realism in the literature, and I believe that the current approach can be very

---

[187] Although strictly speaking, the formalist does not even have the tools to make changes in axiomatizations, since the choice of axioms obviously goes beyond formal mathematics.

[188] Or it acquired a determinate truth-value as a function of time, depending on the underlying logic for the deflationist.

[189] "Non-temporal" does not need to be understood in the strict abstract manner here. It is strange enough that a basic statement of arithmetic could change its truth-value mid-decade while the underlying axiomatizations remain intact.

fruitful in evaluating that debate. The opponents of semantic realism, such as Dummett (1977), hold that since we have no guaranteed method of determining the truth-value of sentences like Goldbach's Conjecture, we cannot think that they have determinate truth-values. Dummett's intuitionist approach against semantic realism is to challenge the notion that we can have knowledge of the truth of a sentence outside our ability to recognize that truth by providing a proof for it. To put it somewhat simplistically: since we do not know whether Goldbach's Conjecture is true or false, we have no justification to call it either determinately true or determinately false. This has of course wide-reaching consequences, the most important of which are probably the rejection of the bivalence of truth and the overall destructive nature that the approach has on mathematical realism. Both of these consequences go so deep into the basic premises of both mathematics and philosophy that they have the potential of making the arguments of this work seem like footnotes. However, whether one accepts the bivalence of truth or not, it is hard to justify the concept that simple arithmetical truths change as a function of time. Yet, once we reject semantic realism, this seems to be the immediate consequence of sentences like FLT. In this sense, combined with extreme formalism, Dummett's approach has an ominous flavour of question-begging: one has to buy the idea of FLT changing its truth-value as a function of time in order to accept the idea that sentences of mathematics do not have determinate truth-values. For the approach of this work, the apparent absurdity of mathematical sentences changing their truth-values over time should be problematic enough – there is no need to go deeper into intuitionist logic and other such issues.[190]

[190] Or how does "FLT became true in 1994, before which it was not true" sound in a textbook of arithmetic? This might seem facetious, but if one thinks that it is absurd, he will also have a lot of trouble accepting the extreme formalist rejection of semantic realism. Of course the arguments against semantic realism are not refuted by this – but I do propose that the argument here must be answered by the semantic anti-realist. Full-blown departure into an intuitionist logic, for example, is not a solution that many modern philosophers and mathematicians are ready to take.

How are these considerations of mathematical truth different from the general theory of truth and the counterfactual case of "snow is white" developing a different meaning and truth conditions? The most important difference between mathematical truth and that of empirical sciences lies in the status of justification, as well as that of the status of truth. Whereas physical sciences focus on providing justification for statements about the world, the justification of formal theories in mathematics is that of axioms and rules of proof. Once they are fixed, the justification of theorems is *ipso facto* inevitable. Whereas a deflationist physicist can find evidence contradicting the theory, this is simply not possible for a strict formalist mathematician.[191] As such the physicist's theories are fallible and counterfactuals can be resolved. In fact, the case is similar to errors in the theory. If the world were different, we would get different evidence and have different theories – just like erroneous theories are contradicted by new evidence. Deflationism in empirical sciences does not commit us to infallible theories, and as such it conforms to the practice of science.

Contrary to this, mathematical theories are not fallible for the strict formalist. Axioms and rules of proof fix the formal system exhaustively and, if the system is consistent, theorems can never contradict them. If there were an error in the axioms or rules of proof, we could never find it out. *That* is why counterfactuals are such a huge problem for deflationist truth in mathematics. Outside inconsistency, the deflationist is immune to errors in rules of proof and axioms. If the axioms refer to something, then for the deflationist mathematician this reference is decided once and for all by the choice of axioms. In short, the strict formalist is stuck with whatever mathematical theory he once accepts, and if it turned out to be false he could never find this out.

This difference becomes obvious when we consider the status of truth in the theories of empirical sciences and mathematics. Although the aim of empirical sciences may be truth, it is generally accepted that the best we can ever hope for is verisimilitude or probable knowledge. Even with the most basic laws of physics,

---

[191] I am not including such obvious counter-examples as inconsistency here.

physicists still acknowledge the possibility that they turn out to be false. History has been a good teacher in this matter. This can also be seen in the justification of the sentences of empirical sciences. The more evidence we have in support of a theory, the more likely we are to hold it as true – yet in empirical sciences there is always the chance that a conflicting piece of evidence is found. In this way, counterfactuals do not present insurmountable problems for the deflationist. Mathematics is different. Once we fix the axioms and rules of proof, the possibility of conflicting evidence having an effect on the theory ceases to exist. For the non-formalist this is not a problem: he can discuss the axioms and rules of proof in an informal metalanguage. The strict formalist, however, is immune to conflicting evidence and hence unable to avoid the problem of counterfactuals.

Can we accept such a state of affairs and be limited to whatever formal theory we have at hand? As far as the practice of mathematics is considered, this seems totally unthinkable. While the deflationist conception of truth always commits us to the provable sentences, Tarskian truth gives us the power to deal with the counterfactuals, the ability to discuss false sentences, as well as the concepts of truth and falsity of sentences. As we have seen, and will see, this is no small matter: in fact, it will be crucial whether we defend ourselves against the formalist arguments or (certain) Platonist ones.

7.2 Truth before reference or vice versa?

Counterfactuals will play a part later on, but let us now move back to even more basic aspects of truth in mathematics. My argument so far has been that extreme formalism leads to arbitrariness. Moreover, I have contended that when it comes to the ontology and epistemology of mathematics, *any* alternative to arbitrariness will do, and hence the question of final reference (that of pre-formal sentences) of T-instances can be left unanswered without damaging the case for Tarskian truth. Strict nominalism and the following deflationism fail; that is enough to warrant the use of Tarskian truth with its central concept of reference. I hope that by

now there does not remain much that is problematic with this conclusion. However, there remains one potential qualm the deflationist will have with this approach: the negative nature of the reasoning. That we are allowed to use reference and meaning in mathematics is not only a question of truth but a question of *ontology*, as well. With Tarskian truth we commit to the references of mathematical sentences. For formal mathematics this is no problem, the references of formal sentences being their pre-formal counterparts. But for the pre-formal sentences the commitment to references is clearly making an ontological commitment to mathematical objects, or at least the objectivity of truth-values of mathematical sentences. Obviously this is unacceptable for the nominalist, most likely up to the level in which arbitrariness is unacceptable for the non-nominalists. One could be accused of putting the cart before the horse: we are talking about reference and semantical truth before we know if there is anything for semantical truth to refer *to*.

The conclusion would seem to be the one that Chihara makes: we need to present a positive development as well as the negative one. This sounds highly worrying remembering how heterogenic and vast a field non-formalist philosophy of mathematics is. There is the danger that in order to use Tarskian truth we are required to present an ontological and epistemological theory of reference. Of course that amounts to nothing less than providing a full philosophical picture of mathematics, an endeavour problematic enough to make all the considerations so far seem insignificant. But is the situation really that dire? Is there not any way we can talk about truth *first* and reference *second*? This is the question I want to answer in this chapter.

Before we continue it must be pointed out that there are two different ways of discussing the order of truth and reference. While it is true that Tarskian truth presupposes reference, this must not be confused with presupposing something about the *nature* of this reference. All along I have argued that we can use Tarskian truth without having exact theories about the nature of mathematical objects. For Tarskian truth it is enough that there exists *some* reference for mathematical sentences. In this way, while Tarskian truth puts the existence of reference first and truth second, we can

characterize truth before we characterize reference. I will argue that there is nothing problematic in that. Nevertheless, it should be acknowledged that there is another way of discussing the order of truth and reference. In the *neo-Fregean* philosophy of mathematics we start directly from truth, and infer even the *existence* of reference from it. This is a considerably stronger argument than the one I make and if successful, it would make many of the considerations here unnecessary.

Frege's (1884) argument for the existence of mathematical objects was fundamentally very simple and it took the form of truth first and reference second. The stock example of this kind of "from truth to reference" inference is the use of existential quantifiers in mathematics. If a sentence of the form "there exists *A*…" is *true*, then the mathematical concept denoted by "*A*" *exists*. This is obviously a semantical connection where "*A*" stands in a referential position, and in Frege's account it would be proof enough for Platonism to show that (at least some) such mathematical statements are indeed true. Given the lack of commitment to outer criteria and reference, the most promising route for this was obviously showing that mathematical theories are consistent and complete. We know what happened to those hopes.

However, from Dummett (1956) to Wright (1983) and Bob Hale (Hale & Wright 2001), neo-Fregeanism (also called neo-logicism[192]) has had considerable popularity. The most interesting facet of it in the context here is the appeal to Platonist reference in mathematics. Matti Eklund (2006) has classified the different aspects of neo-Fregeanism and the one we are concerned with in this work is called *priority*. Priority is the idea that "truth is constitutively prior to reference" which of course mirrors the subject of this chapter. Given *truth*, what will we end up with regard to *reference*? It sounds ominous for the purpose of this work that the neo-Fregeans end up with Platonism, as we have been trying all along to

[192] Here we are not concerned so much with the logicist possibilities of the neo-Fregean program. That is a largely parallel subject, but the focus here is on the Platonist argument of neo-Fregeans. See Rayo 2005 for a good overview of the logicist argument.

minimize the ontological commitments – while being careful not to fall into arbitrariness. Platonism with its strict ontological demands is exactly the thing I want to avoid, while still retaining the privilege to speak about truth without a theory of reference to back it up.

In addition to targeting formalist mathematics, the neo-Fregean line of thinking sounds potentially damaging to the arguments in this work. If successful, it would give us a way to introduce truth without an antecedent theory of reference. By itself this would be very welcome because once it is done, we could use a Tarskian theory of truth without worrying about the counterarguments concerning reference. But problematically, it would give this based only on logical principles, thus retaining the formalist flavour of mathematical sentences having completely *internal* criteria for truth. In addition, the Platonist conclusion is a highly undesired one. If taking truth before a theory of reference implies either of these positions, my argument is in trouble. We will now examine these questions.

7.3 Neo-Fregeanism

The main idea of Frege's (1884) (as well as Whitehead's and Russell's) logicism was deriving the laws of arithmetic from the axioms of logic. However, the central concept of equinumerosity is not derivable from classical first-order logic – and without equinumerosity we cannot have a definition of a natural number. Frege's famous strategy was to introduce what George Boolos (1998, p. 171) calls *Hume's Principle (HP)*:

(*HP*): The number of *F*s is equal to the number of *G*s if and only if there exists a bijection (one-to-one correspondence) between *F*s and *G*s.

Following the so-called *Frege's Theorem*, from *HP* (a second-order sentence) and a set of definitions we can derive all the axioms of

second-order arithmetic[193]. This was the basis for Frege's logicist program. As we know, it was popular for quite a while until Gödel's proof of incompleteness came along. Second-order arithmetic is like PA in that the axioms are thought to give an implicit definition for the concept of "number". It is quite reasonable to require that such an implicit definition needs to be consistent as well as complete in order to be acceptable.

Incompleteness and unprovability of consistency being the behemoths they are, Frege's program was left in the shadows for almost a century before philosophers returned to it. While the lack of consistency proofs was a problem for many, the neo-Fregeans contended that we could have analytic truth nevertheless. *HP* as the explanation of a natural number was the perfect example of this. In Frege's (1884) *Grundlagen* §64 he writes about "carving up" the content of a concept in a different way and thus yielding a new concept. His example is that of the directions of lines:

(*D*): The direction of line *a* = the direction of line *b* if and only if lines *a* and *b* are parallel.

In a similar way, *HP* is considered to carve up the concept of number. Just like *D* is an analytic truth explaining the concept of direction, *HP* is the analytic truth explaining the concept of number.[194]

When we look at the question philosophically from the standpoint of truth and reference, we see that priority is the central theme. *HP* being the explanation of the concept of number, what is the order of truth and reference in it? If we take reference first and say that there exist such things as numbers, *HP* is quite obviously true as an explanation of numbers. This is something everybody can agree on. The neo-Fregean point, however, is that the inference

[193] One should remember that PA is a first-order theory of arithmetic, so the context is slightly changed with *HP*. However the philosophical conclusions made here will not depend on the stronger theory of arithmetic.

[194] See Hale & Wright 2001, pp. 91-116 for details.

can be reversed. In Wright's (Hale & Wright 2001, p. 153) most simple characterization:

> Objects are what singular terms, in their most basic use, are apt to stand for. And they succeed in doing so when, so used, they feature in true statements.

In a nutshell, the argument is that since *HP* is necessarily true and natural numbers are in a referential position in it, the natural numbers exist.

The first question one is bound to ask concerns the concept "necessarily true". This looks suspiciously too easy for the neo-Fregean since it seems like we are in fact talking about a *definition* – and not just an explanation – of the concept of number. *HP* is necessarily true as an analytic sentence, but there is a strong air of it being trivially so. If there did not exist such objects as numbers, *HP* would ultimately be a characterization of a fictional class. Or so would one obvious anti-neo-Fregean argument go.

The neo-Fregean, however, is not that easily repudiated. The crucial step for him is that we cannot discuss the concept of number outside the context of *HP*. This is based on Frege's *context principle*, according to which words get their meanings only in the context of sentence. According to the neo-Fregean, it does not make sense to ask anything about numbers in a context outside *HP*. Numbers exist in the way that *HP* says and any questions about their existence beyond or before *HP* are presented in a mistaken context.[195] The set of all sentences that use *HP* gives us the implicit definition of number, and we cannot discuss this definition in a context outside *HP*. In that account numbers get their meaning completely from their use in mathematics, and it does not make sense to talk about numbers in any other way. Hence the analytic truth of *HP*, and the fact that numbers are referred to in it, gives us the existence of numbers and truth has priority over reference.

---

[195] See Hale & Wright 2001, pp. 335-397. Eklund (2006) gives a good general - and critical – assessment of the neo-Fregean program.

This should not be understood as priority in the sense of existence, but rather in the sense of *conceptuality*: we cannot discuss the concept of arithmetical reference before the concept of truth because the latter defines the former. *HP* together with logic is enough to give us second-order arithmetic; it is analytically true, and gives us the context of natural numbers. This is all something we could be ready to accept, neo-Fregean or not. But the step to the objective existence of numbers is overwhelmingly more controversial. The most important question is how we can know that we are explaining an existing concept and not defining a new one? By explaining equinumerosity we carve up the same content as the explanation of number. Could we not just as well *define* natural numbers with equinumerosity and end up with *HP*? The neo-Fregean idea drawing from the context principle is that this is the case because we have no other way of speaking about *HP*. Yet also this seems like a criterion befitting both a complete explanation and a definition. It does not take much trust in Occam's razor to conclude that we are more likely to be discussing the latter. Something else is needed for the neo-Fregean account: something to distinguish between objective existence and fiction.

So what is the value of *HP* to the Platonist case? Hale & Wright (2005 p. 172) argue that the neo-Fregean position is important because it can avoid Benacerraf's dilemma. Since we do not postulate any causal connection from abstract numbers to *HP*, they argue, there does not exist any epistemological problem:

> the truth of [numerical] identities (and hence the existence of numbers) may accordingly be inferred. We thus have the makings of an epistemologically unproblematic route to the existence of numbers and a fundamental species of facts about them. (ibid.)

For one unacquainted with neo-Fregeanism, quotes like this one could be almost shocking to read from 21st-century philosophers. How can providing a definition for the concept of number be an "epistemologically unproblematic route" to their existence? We will soon see that there are many more or less technical arguments against neo-Fregeanism, but one cannot escape the initial

amazement of how easily a definition – what is in essence a *linguistic fact* – is used to derive a radical ontological thesis.

However, in one sense the thesis is perfectly understandable: in the neo-Fregean account we *do* get all our knowledge of the numbers in an epistemologically unproblematic way, that is, by defining them. Most importantly, this immediately does avoid Benacerraf's dilemma. Unfortunately it also begs the original question of the existence of numbers. How can we get the meta-level knowledge that numbers defined this way correspond to anything *existing*; that we are not just creating fiction? Granted, *if* numbers exist, *HP* would seem to be a true statement about them. But that is only half of the picture. What if we define numbers as a fiction that corresponds to *HP*? Would we have any way of distinguishing these two positions? It seems highly implausible that *HP* alone could have relevance to this question. As we will see, the neo-Fregean argument for Platonism is as epistemologically problematic as anything in the philosophy of mathematics. In essence, we have just moved the problem to a deeper level, one where we have to ask whether *HP* explains or creates the concept of number. *HP* may be an analytic truth, but as Boolos (1998, p. 304) asks, how do we know that *HP* is not analytic in the same sense that the sentence "the present king of France is a royal" is analytic?

There is always something alarming about linguistic facts being used to derive ontological conclusions, and ultimately that is what the neo-Fregean does. We will soon turn to some of the more technical arguments against neo-Fregeanism, but we should first glance at the general problem of using linguistic facts in ontology. *HP* is an explanation of the concept of number and as such it is not fundamentally different from other analytic truths like "all bachelors are unmarried". While the existence of bachelors is uncontroversial, Field (1984) points out that the neo-Fregean approach has a notorious parallel in the history of philosophy: the ontological argument of Saint Anselm for the existence of God. This is a simple counterargument from Field, but initially it looks highly compelling. The jump from a linguistic matter into ontology is based on our language being infallible and having clear references. Certainly *HP* (or its equivalent) seems to be the only

way to explain the concept of equinumerosity, but similarly "nothing greater can be conceived" may seem like a valid explanation for the concept of God. The ontological argument is of course the culmination of *rationalism* in metaphysics. The neo-Fregean use of *HP* to argue for Platonism has the very same flavour. But in any even slightly empiricist-influenced philosophy mere words are cheap; surely we have the ability to create and explain fictitious entities. Consider the sentence "Polonius is father to Ophelia if and only if Ophelia is daughter to Polonius". Clearly this is an analytic truth and in the cast of characters for *Hamlet* this is all we learn about Ophelia. But it would seem very strange to make the ontological claim that there exists an Ophelia who is daughter to Polonius. For that we need something more: namely, the knowledge that Polonius exists. Of course analytic truths by themselves do not make objects exist, and the neo-Fregean case must be somehow fundamentally different from the proposed counter-examples here.

For that purpose, we can first take the approach (see Hale & Wright 2001, p. 163) of analysing sentences into a more detailed form such as "there exists a Polonius who is father to Ophelia…". Similarly, as Eklund (2006) points out, the neo-Fregean can claim that *some* definition of "God" could indeed refer to an existing entity and the ontological argument would be valid. Presumably, this could mean anything from a shared intersubjective idea of God to a metaphorical interpretation. These are good points, because *HP* is indeed different from the arguments concerning God and Ophelia. One can explain the concepts of God and Ophelia in other, non-equivalent ways, while *all* ways of explaining the concept of natural number seem to be equivalent to *HP*. In that way, as Wright (Hale & Wright 2001, p. 158) points out, unlike *HP*, the ontological argument is not a complete explanation of God, and therein lies the difference.

However, there is another way in which the above argument is irrelevant, and that is the context of counterfactuals. What if *HP* had developed a different meaning, referring not to the existing numbers, but something else? Would *HP* be replaced by a different explanation for the concept of number? Simply put, if *HP* was not the complete explanation of numbers it is thought to be, could we

somehow find this out? If priority holds, that could not be the case, since *HP* is true before reference. Hence, in addition of *HP* being a necessary truth; its reference, the concept of number, would also have to *necessarily exist*.

Conversely, what if *HP* did indeed explain the Platonist concept of number but we would not know it? *HP* may be an analytic truth but clearly not all equivalences where one side of the equivalence in *HP* is applied are analytic truths. As obvious as *HP* sounds, it is still picked from a whole class of possible definitions for the concept of number. Consider the following "Injection Principle":

(*IP*): The number of *F*s is equal to the number of *G*s if and only if there is an *injection* from *F*s to *G*s.

While *IP* has the desired consequence of $n(F) = 3$ being equal to $n(G) = 3$, it also has the undesired consequence of $n(F) = 3$ being equal to $n(G) = 4$. Considering that *HP* is a fact, what if we counterfactually claim that *IP* is true? Clearly a natural number as defined by *IP* is a useless concept, but can we not claim it to be true – or even *analytically* true – if we have absolutely no antecedent criteria for the concept of number? Now – remembering that at this point we can only speak about truth, and not reference – what reason do we have for choosing *HP* over *IP*? Of course there are some obvious criteria for choosing *HP*, like the lack of symmetry in *IP*, but how do we come to apply such criteria in the explanation (definition) of numbers? The purpose of *HP* was to define natural numbers for us, but we must not hold the illusion that without any reference there is something we hope to capture with such a definition. The neo-Fregean argument is that truth comes before reference and that from the analytic truth of *HP* – plus the fact that numbers are in a referential position in it – alone we can infer the existence of natural numbers. How can we say that from *IP* we do not infer the existence of some existing set of numbers?

All this seems to lead to the question of arbitrariness once more. As we remember, Wright's (1992, p. 48) contention is that we are dealing with such deep-entrenched – *superassertible* – standards in mathematical practice that they need no outer justification. But in Chapter 4.1 we already found this viewpoint unsatisfactory. *Why*

are certain truths superassertible? We simply cannot accept the appeal to conventionalism here. Certainly *HP* and its equivalents seem like the only acceptable explanation for the concept of number, but that already presupposes that we desire something from a concept of number.[196] In contrast to *IP*, we do not want 3 = 4 to hold. But as I have argued all along, in this philosophical context it is not acceptable to think of the truth (or assertability) of such concepts in modern-day terms. We need to imagine the unknown first steps of mathematical thinking, and ask whether it was an analytic truth that originally gave us the concept of number, or whether it had something to do with reference – say, with explaining observations of the physical world. In other words: did truth come before reference after all? My conclusion here should be clear by now: we can speak about truth *because* there is reference for mathematical sentences – the pre-formal idea of a number that has given us the criteria which make us define numbers as *HP* instead of *IP*. It is the existence of reference that gives us the ability to use Tarskian truth in mathematics. Priority in the neo-Fregean sense fails for the same reason as formalism: if we do not tie our concepts to anything objective, there is nothing to prevent us from changing their content.

Of course, given the problem of counterfactuals, this should be expected. Earlier in this chapter I argued against extreme formalism on the basis of counterfactuals, but with neo-Fregeans the matter is no different. Every time we take truth (or assertability) prior to reference, it is our set of true sentences that tells us what, if anything, exists. But from the problem of counterfactuals we see that we would still cling to these same true sentences even if there had been a change in their meaning, and hence also in their references. Either we must accept a kind of idealist account that our truths *create* their reference, or we must accept that reference can change our truths and priority in the neo-

[196] Indeed, picking up *HP* already makes a commitment to the underlying logic we want to use. Using the axioms of PA also gives us an equivalent definition of number. How are we to choose between such options?

Fregean sense is mistaken.[197] Not surprisingly, it is only in the philosophy of mathematics that scientifically minded philosophers are still ready to make the former claim. However, we should know enough by now not to make the leap from deeply entrenched conventions into *necessary truths*.

7.4 Bad company and neo-Fregean epistemology

I have presented above some general informal considerations concerning neo-Fregeanism. It has been my purpose to stress the problematic inference from analytic truths to ontological statements. In one way, I think that nothing further is needed, as I will argue later on. Still, there is one big difference between *HP* and any of the examples above: unlike *HP*, the other analytic truths (if they are that) are not anything we would wish from a theory of mathematics. The most mathematical one, *IP*, is not even an equivalence relation and as such totally hopeless as a definition of numbers. It was an extreme example of a flawed definition, and some would (quite rightly) say it was too extreme. Strictly speaking, however, without any prior appeal to reference the neo-Fregean does not even have the ability to distinguish between *HP* and *IP*. But let us grant that such considerations are possible, and we can speak about mathematical concepts in a non-arbitrary way. For that purpose we now move to a couple of more sophisticated counter-examples.

Considering the more technical issues about *HP*, it has been a common argument against neo-Fregeanism to claim that if *HP* is indeed an analytic truth, then so are all other second-order *abstraction principles* of the type:

$f(F) = f(G)$ if and only if $R(F,G)$,

[197] The third option being the skeptic's point of view that we are stuck with beliefs of which we can never know whether they refer to anything or not.

where *f* is a function from concepts to objects and R some relation of concepts. However, this by itself is false since there are abstraction principles that are not analytic truths. For example, one of Frege's laws (Basic Law V):

> The value-range of *F* = the value-range of *G* if and only if, for all *x* it holds that *F*(*x*) if and only if *G*(*x*).

is not satisfiable, because it causes Russell's paradox.[198] As such it cannot be considered an analytic truth. Clearly not all abstraction principles are analytic truths and the neo-Fregean must find a way to distinguish between *HP* and the undesired abstraction principles like Basic Law V. This is called the *bad company* objection to neo-Fregeanism in the literature.[199]

Abstraction principles as such are not enough to refute neo-Fregeanism, but they give us a good idea of how to find counterexamples. Surely *HP* cannot be the *only* such analytic truth, and there must be abstraction principles that have the same status as *HP*. While Frege's Basic Law V is known to be inconsistent, there are also abstraction principles that are not. George Boolos' (1998, pp. 214–215) example of *parity principle* is the most commonly used one in the literature:

> The parity of *F* = the parity of *G* if and only if *F* and *G* differ evenly (where two concepts differ evenly if an even number of things fall under one but not the other).

Boolos showed that the parity principle is satisfiable only in finite domains while *HP* is satisfiable only in infinite domains. Hence the two supposedly analytic truths are incompatible. How are we to know which one is actually true? The neo-Fregean answer to this can be introducing new criteria for acceptable abstraction principles – conservativeness was one suggestion – but against all such criteria there have been presented counterexamples.[200]

---

[198] See Boolos 1998, p. 173.
[199] Eklund (2007) presents a good overview of the bad company problem.
[200] See Eklund 2006, pp. 110-115.

Ultimately, we can think of analytic truths that by definition do not change anything mathematically, but are incompatible with *HP*. Eklund's (2006, p. 112) preferred counterexample is that of *anti-numbers*. Anti-numbers are abstract objects that do not change anything in mathematics but have the feature of ruling out the existence of numbers. Now if the neo-Fregean conclusion is moved into anti-numbers, we can infer that the anti-numbers exist, in addition to numbers existing due to *HP*. Both numbers and anti-numbers exist, but obviously they cannot exist at the same time. This is what Eklund calls the problem of *incompatible objects*. One might try some kind of indeterminacy move: that sometimes numbers exist and other times anti-numbers do. But the problem of counterfactuals gave us the conclusion that *HP* being a *necessary* truth, numbers would have to *necessarily exist*, which obviously goes against such indeterminacy.

One more problem is noted by Boolos (1998, pp. 313-314), and although it looks innocuous enough, in the end it could be the most serious one of them all. In the way the neo-Fregean program is carried out, zero is the number of things that are non-self-identical. By the same account there is also the number *anti-zero* $n[x\colon x = x]$, the number of things that *are* self-identical, in other words the number of all the things that there are. But in the usual set theoretic (ZF) conception of defining natural numbers as sets, there is the restriction that there is no set of all sets, as it causes Russell's paradox. So it seems that in addition to all the other problems, the neo-Fregean project seems to contradict with ZF set theory.

Perhaps there is a solution to the problem of anti-numbers, as well as the parity principle, although they both seem very problematic. Boolos' last objection definitely looks like something the neo-Fregean must answer and ditching ZF sounds like a Pyrrhic victory. But in any case, even if all that is solved, tinkering with the form of analytic sentences does not sound like a promising strategy in the long run. Even if the neo-Fregean could ultimately refute all the counterexamples in addition to the bad company of abstraction principles and incompatible objects, it would still not suffice to convince the anti-neo-Fregean when it

comes to the ontological side of the argument. The technical part of the debate is one matter, but the acceptance of the Platonist thesis goes way beyond that.[201] It could be an endless pursuit to create new counterexamples for the neo-Fregean's challenge, but there are also more general philosophical problems that we have to deal with. If one is not convinced that the truth of analytic sentences can imply the existence of their references, the particular formulations and differences between analytic sentences matter very little. Obviously mathematical sentences are the best candidate for the neo-Fregean strategy because they are maximally unambiguous and well-formulated. Nowhere else do we have the kind of complete explanations that *HP* gives us, and that can make neo-Fregeanism seem more appealing than it actually is.

However, if one is committed to any kind of empiricist and scientific epistemology, there is no argument that will be enough to convince that a linguistic fact can imply ontological facts in the neo-Fregean way. If, on the other hand, one is prepared to give room for such a rationalist ontology, then *HP* is merely an efficient way of stating this belief. In this way, I think that the structure – even if not the content – of Field's argument about the ontological proof of God's existence already carries much of the power against the neo-Fregean. The assumption of *epistemological rationalism* is much stronger than the conclusion of *ontological Platonism*. If we have a way of knowing objective mathematical facts via pure reason, surely it is not any more problematic to think that there exists a domain for such facts? It must be remembered that this is not only a question of the nature of *mathematical* objects (if any). Indeed, as Rayo (2005, Chapter 4) points out, to make the neo-Fregean Platonist claim from *HP* implies that there are infinitely many natural numbers, which in turn requires an infinite universe.

---

[201] It could be said that the technical side which concentrates on the abstraction principles is more about the *logicist* side of neo-Fregeanism, while the *Platonist* side is more or less independent of it. Wright (1983) seems to follow this distinction, as Eklund (2007) notes. Indeed, given the problem of absolutely no reference, the logicist side sounds like a much more promising venture – aside from the problem that it is not at all clear that *HP* is a sentence of logic (see Boolos 1998, p. 216).

We might or might not accept the infinity of the universe, but few of us are ready to do it based on a definition in mathematics. Clearly for such requirements as infinity we need *antecedent* justification – and at that point the neo-Fregean thesis is obviously redundant.[202]

As was noted earlier, the above problem has a lot to do with the neo-Fregean choice of terminology. In fact, I am convinced that most problems of neo-Fregeanism are more or less linked to the choice of the central term used to characterize *HP*. Wright (1983, p. 153) writes:

> ...the fundamental truths of number theory would be revealed as consequences of an explanation [*HP*]: […] a statement whose role is to fix the character of a certain concept.

It is quite clear that the second half of the quote is characterizing a *definition*. Then why does Wright talk about *HP* being an explanation? Also Boolos (1998, pp. 310-311) suggests that what *HP* really is for Wright is an analytic definition. Of course the flavour we get from the word "explanation" is that of explaining something existing, and that seems to be what Wright is after. If we call *HP* a definition, it suddenly gets the more constructivist air of fixing a concept, whether it refers or not. But a definition it is, there is little doubt about that. Ultimately, I do not think that *HP* can be understood as anything other than an implicit definition for the concept of natural number. Whether this definition refers to anything existing or not is another question, one that requires antecedent (to *HP*) justification for the reference. Considering the possibility that no mathematical objects exist, logically speaking, we would still be able to define numbers with *HP*. That could not be the case if the neo-Fregean argument were sound.

That is the case when we think of *HP* as a constructive principle defining numbers. But Hale & Wright (2001, pp. 147-148) have

---

[202] The ontologically unproblematic way of dealing with the infinity of natural numbers is to consider them to be a *potentially* infinite sequence. With the Platonist twist, however, we move from potential to *completed* infinity – an altogether more problematic position ontologically.

another question to ask. They respond to the demand of antecedent knowledge of reference by asking how *else* could we know that the natural numbers exist and have the characteristics they have; that, for example, they constitute an infinite series? Simply put, according to Hale & Wright, knowledge about natural numbers antecedent to *HP* is not possible. In their account, all knowledge of numbers follows from *HP* and this gives the neo-Fregean argument the power to avoid Benacerraf's dilemma. I think that this is a very good question, although one that is primarily targeted against the non-neo-Fregean *Platonist*. Of course here one has to remember that the axioms of PA will define the same numbers as *HP* does. We can convince ourselves about the characteristics of numbers in other ways than *HP*. But PA is no different from *HP* in any positive sense, either: it gives us an implicit definition for the concept of number. For the antecedent knowledge of the existence of numbers we would need something different.

However, it seems to me that Hale & Wright do not sufficiently recognize that our antecedent knowledge about numbers does not need to have *all* the characteristics that PA or *HP* give them. In an empirical account of the origins of mathematical knowledge, for example, we do not get knowledge of all the properties that the natural numbers have – but we do get knowledge of some of them. We can use PA (or *HP*) to explain our antecedent (perhaps empirical) pre-formal knowledge of natural numbers and go on to study what *other* qualities they have according to the axioms of PA. Of course not only can we do this, but most likely this is the way arithmetic has been *actually* developed: PA and *HP* are both quite modern ideas that follow sophisticated antecedent knowledge of natural numbers. Indeed, perhaps the best way to answer the question of Hale & Wright is a factual one: to see how we can get knowledge of the natural numbers outside of *HP*, we must look at how it is actually acquired before devices like *HP* come along. If *HP* is the only way of getting knowledge of the natural numbers, then before Frege (or Hume) we could not have had any knowledge of them. Likewise, those currently unfamiliar with *HP* could not be said to have knowledge of natural numbers. But that is just absurd. Even if we may lack a definition, we certainly can

have ways to use natural numbers to great effect, in practice and in theory. After all, most of arithmetic dates from times before Frege. In my account, we can count all this as antecedent knowledge.[203]

Finally, there is the question of *HP* being an analytic truth in the first place. In addition to such nominalist/Platonist problems as the phrase "…there exists a function…" in *HP*, the original problem with Frege's logicist program remains: we cannot know the theory of arithmetic to be *consistent*. I do not see this as much of a problem, as argued earlier, but my grounds for accepting arithmetic are different. Half the idea behind the neo-Fregean program is that arithmetic is derived purely from the truths of logic, plus the (according to Boolos) non-logical axiom of *HP*. In this respect the lack of consistency proof is definitely a drawback. Platonism or not, for *HP* to be an analytic truth, we should expect it to define a consistent system of arithmetic. It sure *looks* like an analytic truth, but is that enough in the strict logicist game that the neo-Fregean commits to? After all, we are talking about radical ontological statements made based on *HP* being *true* and without consistency that truth is not established.[204]

7.5 Two kinds of priority

I have claimed above that priority in the neo-Fregean sense fails. In addition to all the technical considerations, it is simply too difficult to accept the leap from linguistic facts to ontological conclusions. But this sounds potentially problematic for the arguments in this work, as well. After all, it is priority of truth that I have argued for all along: that we can use Tarskian truth and its appeal to reference

---

[203] Of course it must be noted that the *idea* of equinumerosity predates Hume and Frege, and is intimately present in our pre-formal conception of number. But already by noting this we are moving into a more allowing field of mathematical knowledge, and into a conception of natural numbers antecedent to *HP*.

[204] Boolos has written (1998, pp. 301-314) at more length about the problems concerning the supposed analytic nature of *HP*. See also Wright's answer to Boolos (Wright & Hale 2001, pp. 307-332).

even if we do not have a comprehensive theory for that reference. However, here we must remember the lesson of Chapter 6: it cannot be accepted that *all* theories of reference fail, because that leads to extreme formalism and arbitrariness. Some theory of reference is needed and this *demand* for reference is prior to truth. This way, the kind of priority I argue for is essentially different from the neo-Fregean priority. We can *characterize* truth before we characterize reference, but we cannot make the *ontological* claim of reference based on truth. This is a very important distinction to make. We can only believe in substantial Tarskian truth in mathematics if we have antecedent justification to believe that there exists a reference for mathematical sentences. The way I have argued for this has been to argue against extreme formalism and its arbitrariness.

However, that argument cannot be applied by the neo-Fregean, because Tarskian truth is much too weak for his purposes. With Tarskian truth we do not need to commit to any particular characterization of reference: most importantly, there is no mathematical sentence that is necessarily true *due to Tarskian truth*. The problems of neo-Fregean strategy cause no damage to the arguments of this work when we remember to distinguish between priority in the neo-Fregean sense and the Tarskian sense. In this work I speak of truth before reference, but it only means that I believe in the possibility of mathematicians finding out true sentences that refer to something non-arbitrary. This is a very weak form of priority, and most definitely not priority in the conceptual or ontological sense.[205] With neo-Fregeanism the case is vastly different.

When we look at the big picture, neo-Fregeanism actually bears a lot of resemblance to extreme formalism. In both theories we neglect the origins of mathematical thinking and start explaining the current mathematical theories. However we arrived at them, we *have* arrived at them, and that is enough for a starting point.

---

[205] In fact, when fully interpreted languages are considered, priority between truth and reference in the conceptual sense would not seem to matter much.

This way both the formalist and the neo-Fregean[206] try to find out criteria for assertable sentences, whether they are axioms of formal systems or analytic truths like *HP*. These criteria can include consistency and conservativeness, but they are always *internal* conditions and never refer to anything outside mathematical theories.[207] That the neo-Fregean chooses to call these conditions "truth" and the formalist "assertability" is, in this way, only a minor difference.

It is only when we consider the ontological conclusions that we see the major difference. While an extreme formalist will consider *HP* assertable, the neo-Fregean considers it to be an analytic truth. So far, this is hardly more than a question of terminology, even directly translatable ones. The difference in ontological conclusion, however, could not be more drastic – and it must be said that here the extreme formalist comes across as much less problematic. Priority in the neo-Fregean way needs the existence of reference to escape arbitrariness – and if we are ready to accept arbitrariness, we should have no problem in rejecting Platonism. However, as always, we must be careful not to throw the baby away with the bathwater. When rejecting Platonism, we cannot be accepting extreme formalism. Simply considering the content of *HP*, it may seem tempting to say that if not grounds for Platonism, it must be grounds for *fictionalism*. After all, *HP* is an instantly acceptable definition of natural numbers. There is a temptation to conclude that if numbers do not exist based on *HP*, then numbers do not exist *at all*. This is the problem we will examine next.

---

[206] See Hale & Wright 2001, pp. 117-150 for an example.

[207] Here I must once again neglect Field's criterion of usefulness; it being so obviously against the doctrine of non-reference in the strictly fictionalist mathematics.

7.6 Non-Platonist reference: Linnebo

My argument against neo-Fregeanism has been that we need to first assume a domain of mathematical objects in order to make ontological conclusions concerning the reference of mathematical sentences. Obviously this does not mean that we should dispense with the Fregean notion of reference, or the objectivity of mathematical truths that comes with it. In fact, Frege's theory of truth before reference fits perfectly well with the general Tarskian account I have been trying to defend in this work. What I claim is simply that we cannot make the ontological claims the neo-Fregeans do based on the theory of reference.

However, while not necessary for the introduction of truth, a theory of reference is undoubtedly something we should wish to have – and there Frege's work is potentially helpful. Neo-Fregeanism *does* avoid Benacerraf's dilemma and that is no small feat for a realist strategy in the philosophy of mathematics. There exists a sizable literature on the subject among neo-Fregeans as well as their critics – much of it concentrating on the technical matter of finding a satisfactory theory of reference. But the technical details of such pursuits are not crucial for the matters in this work. It is more interesting to see whether a Fregean pursuit of reference can be satisfyingly constructed without the Platonist conclusion, but just as importantly, without succumbing to nominalism and fictionalism. It must be remembered that nominalism is the obvious first alternative that we have for explaining *HP*. As we saw, strict nominalism and neo-Fregeanism are actually closely related viewpoints until the ontological conclusions are made. Clearly numbers defined by *HP* refer to some set of objects. If not Platonist, it is natural to argue that the set is fictional. Indeed, the counterarguments so far have been destructive more than constructive and fictionalism has been the big winner. But once again, this is to make the wrong dichotomy of Platonism and nominalism. Let us now look at the possible ways of including a Fregean theory of meaning before reference without ending up with Platonism *or* nominalism.

Øystein Linnebo (2006b, 2007) in his Fregean project presents a theory of reference that resembles the neo-Fregean ones, but does

not end up with a "thick" Platonist ontology of mathematical objects. However, Linnebo does not end up with a nominalist account, either, as his theory allows reference and objective truths of arithmetic. Since his conclusions resemble my arguments here, we will now examine Linnebo's argument as an ontologically sensible alternative in the field of Fregean mathematical reference.

Linnebo's (2007) strategy is to use Frege's argument for Platonism as the starting point to develop a kind of structuralist conception of reference for natural numbers. To establish this, Linnebo argues that the nature of mathematical reference is fundamentally different from the standard physical conception of reference. His example is the concept of "roundness" in physical objects. The truth of the statement "this body is round" depends on the statement having a particular semantic content, but also on the *non-semantical* state of affairs in the world. But the truth of a mathematical statement like "2 directly precedes 3" is, Linnebo argues, different. *Everything* we use to decide the truth of "2 directly precedes 3" is semantical, from the places of "2" and "3" in the natural number structure to the meaning of the expression "directly precedes". While the truth of physical statements depend both on semantical and non-semantical facts, the truth of mathematical statements does not depend on any completely non-semantical facts.

In addition, the beliefs that make mathematical beliefs *true* also explain why we *have* those beliefs in the first place, thus answering Benacerraf's dilemma. To show this, Linnebo compares the accounts of reference for physical bodies to those of natural numbers. His approach is to think of the way a robot learns to assess statements of the physical world. With physical bodies the robot gets information perceptually and uses some equivalence relation to conclude whether two parts belong to the same body. Similarly, Linnebo suggests, the robot uses the equivalence relation of Frege's equinumerity to conclude that two numerals (names of numbers) refer to the same natural number. The difference, as was stated above, is that with mathematical truth the equivalence relation by itself is sufficient, and non-semantical facts do not need to come into the picture.

So far this pretty much follows the usual Fregean strategy, but Linnebo makes interesting conclusions out of it. Continuing with the Fregean tradition Linnebo holds that, unlike physical descriptions, mathematical descriptions are by themselves enough to give all the information of their references. While a statement "the mass of x is m" cannot be decided by observing x alone, statements concerning the natural numbers can be completely decided from their numerals. What we can know of "4" in the decimal sequence of numbers is *all* there is to its reference, the number 4 of the natural number structure. This is obviously reductionist and if we reject neo-Fregeanism, it looks like a victory for the nominalist. Indeed, if the names (numerals) are enough, why do we need to postulate the reference (numbers) anymore?

However, Linnebo disagrees with this reduction. He claims that we must dispense with the physical notion of reference here, and turn to one more suited for mathematics. This reference he calls the *semantic values* of numerals, and it corresponds to the familiar Fregean (1892) idea of difference between sense and reference. Famously, the expressions "morning star" and "evening star" have a different sense, but the same reference, the planet Venus. So the reference of a concept is distinct from its sense. This reference Linnebo calls the "semantic value" of an expression. In a true Fregean way, it is the principle of compositionality that tells us when the semantic values are the same. If the expression "Louis XIV" has the same semantic value as the expression "The Sun King", then "Louis XIV had the longest tenure of any European monarch" must have the same semantic value as "The Sun King had the longest tenure of any European monarch".

This concept of semantic value saves Linnebo's Fregean argument from nominalism. Now the numerals "5" and "V" have the same semantic value, and they function in a manner similar to how "Louis XIV" and "The Sun King" function in the English language. Clearly the latter two names have a reference, so why not the numerals? Against this, it could be argued that now that we have the reduction to numerals, why do we need to bother with the semantic values? But Linnebo (2006b, Chapter 4) rejects this objection based on the possible difference between the type of reduction and the semantic analysis given here. We may have

arrived at the reference – and the following reduction – via some *other means* such as perception. As such, we cannot equate between the reduction of the *reference* and the reduction in the *semantic analysis* of numerals.

So what ontological conclusions can be made out of Linnebo's analysis? In his own account Linnebo states that while natural numbers are "thinner" than physical objects, via semantic values they still have legitimate references and hence do not conflict with Platonism. This is of course to define Platonism in a very weak way. Such a thin notion of natural number is hardly what Frege was after. For the purposes of this work, however, the conclusion sounds perfect. With legitimate reference we can instantly move to Tarskian truth, and the semantical arguments for the substantiality of truth apply. I see only one major gap in Linnebo's argument so far and that it is the answer to the reductionist objection. If the semantic analysis gives us the semantic values, must we continue using them if *another* analysis shows that these semantic values can be reduced to numerals? One must appreciate the (potential) reductionist argument that it does not matter which way we arrive at the reduction, as long as the reduction *can be done*. The semantic values of numerals are of course natural numbers, and if numerals are enough, one could argue that the semantic analysis is no longer needed.

What we need to complete Linnebo's account is some argument to the effect that semantic values are indeed needed and that they are never completely redundant. Fortunately, it seems to me that Linnebo has had this in his argument all along: it is the very fact that "5" and "V" *have the same semantical value,* that of the natural number 5. That "5" explains everything there is to 5 can be redundant, but that we can *explain this* is not. In this way, to use an understatement, "5" simply happens to be a very good explanation of the natural number 5. Linnebo (2006b, Chapter 2.3) seems to be after the same thing with his distinction between semantics and *meta-semantics*. To make sense of the whole phenomenon of numerals referring to natural numbers is a question of meta-semantics, and in this sense the natural numbers are not redundant.

The best argument for objective semantic values seems to be that without them we cannot explain the phenomenon of "3", "III" and "three" having the same semantic value. This problem can be seen in all its seriousness when we consider teaching numerals in a language that does not have them. How does a missionary explain the numeral "5" to people whose indigenous language does not contain such numerals? To state an example closer to home: how can a child learn to use the numeral "five" if all there is to "five" consists of just an independent circle of translations between "5", "V", "five", etc.? Of course the answer is familiar to everybody: it is done ostensively by pointing out groups of five objects and letting the student realize the connection between them. Frege defined natural numbers this way and it corresponds to our most basic understanding of them. That is why *HP* works as a definition of natural numbers, and that is also why the neo-Fregean plan has whatever appeal it has. *HP* seems to be so obviously true even in our most primitive thinking, and numbers defined by it seem to refer to an objective natural number, a place in a natural number structure, or in the very least an objective semantic value. However, if we reject that objective reference, what are we left with? Nothing other than an empty system of translatable languages without references; languages which can somehow be learnt without any prior knowledge of them, developed to include new concepts, and used to great effect in scientific as well as in direct applications.

That system of translations is obviously the fictionalist plan, and we know all the serious problems behind it. But as such, Linnebo's arguments do not necessarily conflict with it. He shows that it makes sense to speak about semantic values even if they are already completely explained by the numerals, but obviously this by itself is not enough to show that there *are* semantic values. The nominalist accepts all kinds of talk, like that of truth, but he cannot accept referring to any objectively *existing* mathematical truths. Field's thesis is not that it does not make sense to speak about truth, but that truth is deflationary and as such philosophically superfluous. Similarly, I see nothing in Linnebo's argument that by itself shows that the semantic values of numerals are not deflationary. That way, as well as obviously being against

Platonism, it could also be seen – alarmingly – as proposing nominalism and fictionalism. If we hold all talk about reference to be fiction, then everything in Linnebo's strategy can be given that notorious nominalist term "useful fiction". Going deeper into philosophy we need something more, and here the old questions of theory choice and arbitrariness come in handy. Once again, deflationary semantic values would mean arbitrary mathematics. We could change the semantic values of numerals as we like, and aside from it conflicting with our conventions, the deflationist would not be able to point out a problem with this approach. For mathematics, however, it would be disastrous.

It seems that wherever we go with philosophy, the backbone of anti-nominalism is the denial of arbitrariness. This should not be a surprise: if we did accept arbitrariness, then clearly nominalism and fictionalism could be instantly accepted. But it is important to realize just how strong the problem of theory choice is. *Any* reason for choosing one theory over another contradicts arbitrariness, and hence any such reason already takes us away from extreme formalism. One cannot stress enough the need to tie semantic values into something objective. Looking at a flock of birds the size of n, we must be able to count the birds from 1 to n without gaps in the semantic values of the numerals. But if the semantic values are not objective, what is there to prevent us from stating that "3" and "4" have the same semantic value 3, and between that and 5 there is no semantic value? In a self-standing system of translatable numerals such changes are possible. *That* is why we need to postulate objective semantic values for numbers.

However, that is only why we *need* semantic values, not why we *have* them. As I see it, Linnebo's arrival at thin but existing natural numbers cannot be all there is to natural numbers. We avoid such problems by choosing *HP* or PA to the define natural numbers, but this choice must be explained somehow. There remains the question of arriving at the natural number structure, and the equinumerity to explain it, in the first place. The origins of mathematical thinking are left alone in Linnebo's approach. Still, Linnebo does give an argument for the need of something objective – the semantic values – in a Fregean project of mathematics, without making the full Platonist claim. Between

neo-Fregeanism and nominalism – with a little tinkering – we now have a third option which gives us justification for the use of reference, and as such for Tarskian truth. At this point it seems obvious enough that the Fregean strategy can be used to make all kinds of philosophical conclusions, and the most important problems will most often not be related to the technical details, but rather to larger epistemological and ontological considerations. Linnebo's case shows that the project can be carried out with minimal ontological burden, yet without losing the need for reference of mathematical concepts. It is not a complete picture, but it does give us a Fregean framework to develop moderate non-nominalist projects like the one described in the end of Chapter 6.

7.7 Neo-Fregeanism and Quine

When we consider the neo-Fregean ontology, one important observation to be made is that the set of true sentences tells us what exists. Rather than being an argument for Platonism, as the neo-Fregean thinks, I have claimed that this is better understood as *demanding* a Platonist ontology. But there is one troubling thing in my approach here, and that is the apparent conflict between mathematical and *other scientific* truths. When rejecting neo-Fregeanism we seem to be rejecting that the true mathematical statements tell us what exist. In any realist ontology that respects the achievements of science it is commonplace to think that it is indeed the set of true scientific sentences that tells us what exists. If not truth, then a milder concept like verisimilitude or probable knowledge takes this place. In any case, with advancements in science we come closer and closer to true sentences about world, and this way truth has priority over reference. In this work we have used the scientific applications of mathematics as an argument against arbitrary formal theories, so it is reasonable to demand that this putative difference between mathematical and physical truth is explained.

Of course the simplest way to deal with the relationship between mathematics and physics is the Quinean strategy of thinking of it all as one theory. We return soon to this approach,

but let us assume for now that there is a difference (as well as a connection) between mathematical and physical knowledge, and it makes sense to speak of them independently of each other. My argument is that Tarskian truth fits both disciplines, and whatever differences there may exist, they are differences in the ontological nature of the subject matter. So how can we reserve for physics the role of telling us what exists, while refusing it for mathematics? The short answer is that we *do not*, and the whole conception of such difference is mistaken. As was said in the previous chapter, reference comes before truth, but only in the sense that there *exists* a reference. No specific characterizations concerning it were made. Some interpretations of quantum theory notwithstanding, there is a clear consensus in physics that there is a world "out there" that scientists are trying to explain. Just what means and theories are the best for this job is up to physics to decide, but the existence of an objective world is an implicit assumption that is ubiquitously made.

I argue that with mathematics the situation is exactly the same. Mathematical theories have evolved to include new domains like complex numbers and there are (perhaps valid) doubts about the existence of them. But there can be valid doubts about the existence of, say, electrons, as well. The important question is whether there can be valid doubts about the existence of the *whole* domain of the subject matter. Here the difference between physical and mathematical truth no longer seems acceptable. It is a return to archaic philosophy to insist that philosophers *qua* philosophers can explain which physical objects exists. However, a philosopher of physics can study the assumptions made in physics and the existence of an objective reality is one assumption that cannot be dismissed. Similarly, if we are not ready to accept arbitrariness, we cannot deny the existence of some reference for mathematical theories. Scientific theories, both in physics and mathematics, tell us what they refer to – as long as we assume that they refer to *something*. Philosophically there is nothing problematic in that, and we reserve the same role for mathematics that the neo-Fregean does; only the Platonist conclusion is replaced by a more allowing array of ontological alternatives.

How about the Quinean (1995) view that the physical sciences and mathematics actually form one theory that tells us which sentences refer to existing objects? Eklund (2006) has clarified the distinction between the Quinean and the neo-Fregean ontology. Both theories endorse the viewpoint that true sentences tell us what exists, but the difference can be seen when we think about what the theories say about *arithmetic*. The neo-Fregean obviously thinks that the natural numbers exist. The Quinean, on the other hand, will hold that the natural numbers exist *if our best theory of science needs them*. Unlike the neo-Fregean, the Quinean is not committed to any particular ontological statement. In this way, the existence of natural numbers is something for science to find out, not a trivial question to be answered based on an analytic truth.

How does this Quinean conclusion compare to the theses in this work? Obviously the big difference lies in whether we include mathematical terms *at all* in our best theory of science. Unlike Quine claimed in his indispensability argument, for the Quinean it actually seems possible to end up with the Fieldian possibility of rejecting the existence of mathematical objects. In fact, this was the whole motivation for Field: he tried to show that the Quinean theory of indispensability fails and we *do not* need to include mathematics in physics. Field's achievements aside, with everything we know by now about his project, we cannot accept this possibility. Not that it is likely that our best theory of physics could ever *actually* be void of mathematics, but starting from basic arithmetic and geometry, mathematics has too many applications outside the developed theories of physics to be just arbitrary fiction. By all the considerations in this work, we must include a theory of reference in the philosophy of mathematics. Of course this does not mean that we cannot have different conceptions over *which* mathematical sentences refer to existing objects – not to mention the nature of this reference. But those are whole other questions.

8. Loose ends

8.1 Non-standard models

One subject that must be addressed in a work like this is the existence of non-standard models. As we know from the Löwenheim-Skolem theorem, the classical first-order languages considered in this work – in addition to having the standard (intended) infinite models – also have non-standard ones (that is, ones not isomorphic with the standard model), in fact uncountably many of them. In PA the main feature of non-standard models is that, as well as satisfying all the true statements of the standard model of PA, they also satisfy new ones. This has obvious consequences when we think of the semantic arguments concerning PA expanded with Tarskian truth. In addition to establishing the true sentences of the standard models, in order to be completely adequate, the truth definition would have to establish all the true sentences of the non-standard models. In essence, the axioms of first-order PA do *not* give us only the model of arithmetic that we want (pre-formally), but also an (uncountably) infinite number of models of arithmetic that we do not want. It goes without saying that if we have no way of distinguishing between the standard and non-standard models, it would – among other things – have a lot of relevance to the question of truth.

Jody Azzouni (1999, pp. 543-544) has used the existence of non-standard models to argue against Shapiro's semantical argument. According to him, Shapiro's argument, especially the part concerning mathematical induction over formulas containing the truth predicate, only holds if the Peano axioms somehow pick out the standard model. There are non-standard models, Azzouni argues, for which such induction will not apply. Hence the arithmetical truth that Shapiro is talking about is the truth of the standard model and – in order to accept Shapiro's argument – the deflationist would need to accept that the Peano axioms somehow implicitly pick out the standard model, which of course they do not.

For the non-formalist, however, one question is immediately raised: why should we ever come to the strange conclusion that we could not distinguish between the standard and non-standard models – unless we are *already* committed to strict formalism? After all, the whole question of non-standard models is discussed all the time in both mathematics and philosophy, with little thought spent on whether we can formally separate them from the standard model. As Shapiro (1997, p. 133) has noted, we have in the *informal* (in my terminology, pre-formal) language of mathematics the resources to distinguish between the two. For a strict formalist who does not have these resources (or who rather, under my interpretation, *pretends* not to have them) this might indeed be a problem. For the rest of us, the fact that non-standard models are spoken with ease in higher-order languages, as well as in pre-formal languages, should be enough.[208]

That is the reason why there was no need to bring up the question of non-standard models earlier as part of the Field-Shapiro-Azzouni debate. After all the problems of formalism we have seen, Azzouni actually seems to reveal another flaw in extreme formalism: the fact that we cannot distinguish formally (in PA) between the standard and non-standard models. This is of course, like many aspects of strict formalism, in a clear conflict with the actual practice in mathematics. That is why we should be able to restrict all the results here to concern the standard models of arithmetic without damaging the arguments.

All that considered, perhaps in place of "truth" we should be talking about "truth in the standard model of PA" to avoid confusion. However, that seems like an unnecessary complication.

---

[208] Indeed, the *second-order* version of Peano Axioms only has the intended model. However, this is the case only if we apply *standard* second-order semantics, which have the problematic feature of assuming the entire power set of the universe in discourse. This subject is discussed in Shapiro 2000b, pp. 80-96. At any rate, such a choice of standardness is again something for which we have no purely formal criteria, so the conclusion made here from the Löwenheim-Skolem theorem holds also for the higher-order languages. We never arrive at the standard model purely formally, yet pre-formally we always manage to pick it out while practising mathematics.

All along I have been arguing that we are after a particular, non-arbitrary, set of sentences when we study truth in mathematics. When it comes to arithmetic, to clarify that to mean our intended model of arithmetic is hardly necessary. Indeed, it should not be unreasonable to ask that the ones researching truth also in non-standard models use the extra clarification. In any case, it is the formalist who finds himself lacking in means to distinguish between the standard and non-standard models. For the arguments presented here, the existence of non-standard models is – if anything – more ammunition.

8.2 Another semantical argument

The Gödelian argument of Shapiro and Ketland is not the only semantical argument for the substantiality of truth. Hyttinen & Sandu (2000) show that introducing a truth predicate into a *Henkin-hierarchy* of languages also causes an undefinability result and a hierarchy of languages that will not collapse.[209] To see this we need to define a hierarchy of languages in the following way:

Let $L_{\omega\omega}$ be a classical first-order language. Hyttinen and Sandu (ibid., p. 520) define a hierarchy of languages $L^n(H)$ as follows:

$L_{\omega\omega} = L^0(H)$

$L_*^{n+1}(H) = \{Hx_0x_1y_0y_1\varphi : \varphi \in L^n(H)\}$

$L^{n+1}(H) =$ the closure of $L_*^{n+1}(H)$ under negation, disjunction and existential quantifier.

Here $L_*^1(H)$ is then a classical first-order language extended with Henkin quantification over the classical first-order formulas. As we remember, having the added expressive power of Henkin quantifiers, and only them, $L_*^1(H)$ is equivalent to an IF-language.

---

[209] To be exact, the truth predicate is a way to *show* that such a hierarchy of languages does not collapse.

The important thing here is, as we remember from IF-logic, is that a language $L_*^1(H)$ is *not* closed under contradictory negation.

The following table represents the hierarchy thus defined:

Level		
0	$L_{\omega\omega} = L^0(H)$	
1	$L_*^1(H)$	$L^1(H)$
2	$L_*^2(H)$	$L^2(H)$
.	.	.
.	.	.
.	.	.
$n+1$	$L_*^{n+1}(H)$	$L^{n+1}(H)$
.	.	.
.	.	.

The language on the left side on Level 1 is defined by extending a classic first-order language with Henkin quantification, that is, over the formulas of the language on the Level 0. The language on the right hand side on Level 1 is defined as the language on the left side under the closure of negation, disjunction and existential quantifier. The language on the left on Level 2 is defined by Henkin quantification over the formulas of the language on the right on Level 1, and so on, *ad infinitum*.

What Hyttinen and Sandu prove (ibid., pp. 520-521) is that on every level, the language on the left side defines its own truth predicate[210] while the language on the right side does not. One corollary of this (ibid., p. 522) is that the hierarchy of languages $L^n(H)$ does not collapse: for no *m, n* such that $n > m$, is it the case that $L^n(H) \equiv L^m(H)$. The language on the left includes its own truth predicate, but once we close the language under negation, it

[210] Of course we must remember that we cannot recognize the truth predicate as such in the language itself, as noted in Chapter 5.2 of this work.

cannot contain its own truth predicate any more. Either we must dispense with the closure under negation or we commit to a hierarchy of languages.

The above result is not restricted to Henkin-hierarchies. The same has been proved for many other hierarchies. For example, Azriel Levy (1965) has proved that for a hierarchy of formulas in set theory, we can have a truth predicate for each level of the hierarchy, under the condition that the class of formulas is not closed under negation. If we close the level under negation, a hierarchy is needed for the truth predicate. The Tarskian hierarchy is of course the most obvious case where the truth predicate and closure under negation cause the liar's paradox and the hierarchy of languages will not collapse. For an extreme formalist this is definitely a problem. Infinite non-collapsing hierarchies of languages are not the kind of mathematical theories that formalists are after.

Yet dispensing with the negation is equally problematic: essentially, we must deal with a many-valued, or incomplete, logic. The proof procedure for such logics is bound to be incomplete, and we must commit to using game-theoretic semantics, the intractable logical consequence of second-order logic, or other such alternative. In any case, the proof procedure is bound to be essentially *semantical*, not the syntactical formal one of classical first-order logic. This is why the argument presented in this chapter can be called another semantical argument. The truth predicate requires a hierarchy of languages when the languages are closed under negation, and it requires a semantical proof procedure when they are not. In both cases truth is substantial: because of the truth predicate formal languages must be expanded from the classical first-order ones.

The conclusion seems clear enough: either by the liar's paradox we commit to a hierarchy of languages or else we must dispense with contradictory negation to avoid the liar's paradox. Whether we can ultimately avoid the liar's paradox even that way is not always clear (see Ketland's point in Chapter 5.4), but in the very least we are committing to much less *formal* mathematics than the extreme formalist could accept. After all, formalism in the Hilbertian tradition meant totally syntactical, consistent and

complete way of proving theorems according to rules of classical logic. Consistency and completeness proved to be impossible. From all we know about the possibility of formalist programs by now, it seems that also classical logic and syntactical proof must be abandoned – or expanded – and even so we are left with a variety of problems. The most important of these is the question of theory choice. Whatever appeal strict formalism still retains, it is a long way from its original ideals, and philosophically as unfeasible as ever.

8.3 Gödelian fallacies

One would imagine that a work titled, "Truth, Proof and Gödelian arguments" is bound to trigger skepticism in some circles, especially as the Gödelian arguments are seemingly used to make deep philosophical conclusions. Generally speaking, this is a healthy attitude: one must remember that Gödel's incompleteness theorems are a sophisticated result of mathematical logic which is reached very much on purpose by a known paradox-raising trick – the diagonal procedure. One must be extremely careful about making far-reaching conclusion from Gödel's theorems into other areas of mathematics, let alone areas where consistent formal systems containing arithmetic are nowhere to be found.

Alas, among philosophers, both lay and professional, there has been a constantly surfacing trend to read too much importance into Gödel's incompleteness theorems.[211] Gödel himself did exactly that by believing that he had found evidence for Platonism[212]; others have made similar mistakes after him. It is not uncommon to read allusions to "limits of mathematical knowledge", "crisis in mathematics" or even "arbitrariness of mathematics" with reference to Gödel's theorem in the non-philosophical literature. This is not helped by the widely known fallacious Gödelian arguments by philosophers. It must always be remembered that

[211] See Franzén (2005) for a good book-length introduction to the subject.

[212] Gödel's views on mathematical truth, intuition and Platonism are best described in Gödel 1964b.

Gödel's incompleteness theorems are a result concerning consistent formal systems of mathematics containing arithmetic – *and only them*. One has to be very careful with the philosophical conclusions, especially when we acknowledge the rather vague connection between mathematical logic and the subjects it has been claimed to concern. These range from the mechanical model of mind (Lucas) and artificial intelligence (Penrose) to poetry (Kristeva) and sociology (Debray).[213]

While the latter two mainly provide comic relief for any serious student of Gödel's theorems, the first two arguments were well constructed and concerned areas that are not too far from formal mathematical systems. In fact, both of them are based on the direct application of the concept of Turing machine to Gödel's theorem. From Alan Turing's work it is commonplace to believe that what we mean by formal systems (in this work, primitive recursive functions) can be identified with algorithmic, mechanical procedures and thus with the so-called *Turing machine*, an ideal model of a digital computer. This contention is known as the *Church-Turing thesis* and it is as widely accepted as anything non-proven in mathematics. Turing showed that we can never know from a Turing machine that it can prove all the theorems of a given formal system. The result is obviously very much like Gödel's incompleteness theorems. What we gain from Turing is that the same problem can be understood in terms of mechanical procedures, and hence, machines.[214]

The important point in both John Lucas' (1961) and Roger Penrose's (1989 & 1994) arguments is that a human being can "beat" the Turing machine by seeing the truth of Gödel sentences, in a manner seemingly very much like Shapiro's and Ketland's. Lucas uses this to conclude that the human mind cannot be mechanical, Penrose to claim that a computer cannot even in

---

213 For the Gödelian misuse by Julia Kristeva and Régis Debray see Kristeva 1969, pp. 189-190 and Debray 1983, pp. 169-170. See Sokal & Bricmont 2003 for a good overview of such arguments.

214 For details, see Penrose 1989, pp. 40-97; Wang 1987, pp. 169-170 and Gödel 1964a, pp. 369-370. For Turing's original article, see Turing 1937, and for the philosophical conclusions of it, Turing 1950.

principle simulate a human brain completely – the inescapable difference being in the Gödel sentences. Naturally this does not mean any Gödel sentence, since there is no problem in mechanical models and computers proving Gödel sentences of *other* formal systems. The weight of Lucas' and Penrose's arguments lies in the contention that a human being can beat any Turing machine (formal system) that is supposed to completely represent our mind, the human thought. Here the healthy dose of skepticism must be taken in. How could a result of mathematical logic possible imply that a computer could never be a complete model of the mind? From what we know by now, the flaw of both Lucas and Penrose is not hard to see. When it comes to Gödelian incompleteness, we are always talking about *consistent* formal systems. Obviously we like to believe that human thinking is sound, but to assume that the human brain (Lucas) or human thinking (Penrose) is consistent is scientifically and philosophically fantastic. We still know quite little about the workings of the brain, but in the sense of comparing it to formal mathematical systems we basically know nothing. That is why Lucas' line of thinking is bound to fail until there is empirical evidence for it. We just cannot hope to answer questions about the consistency of the brain (or the mind). Until that, Lucas' argument gets the form "*if we know the human mind to be consistent*, then by Gödel's incompleteness theorems…" – obviously as speculative a conditional as there can be.

Penrose's argument has somewhat more potential since it concerns the output of the brain, the human thinking. It seems much more plausible that human thinking could be consistent. In areas like mathematics there are people who have come very close to consistency, perhaps even reached it. However, that is considering the mathematical output, and human thinking as a *whole* is a different thing. We are talking about the philosophical question of artificial intelligence, ultimately the ability of computers to simulate human thinking completely. That is a phenomenon as complex as the human behaviour in all its facets, not just that of mathematical knowledge. How could we ever know that this is consistent? Furthermore, how could we ever now from a proposed formal system that it actually captures the human

thinking? Until these questions are answered, also Penrose's argument takes the form of a conditional where the antecedent is extremely speculative.

We see the danger in applying Gödel's incompleteness theorems in fields other than formal mathematics. What about Shapiro's and Ketland's semantical arguments? Are we not falling for the same thing? Mathematical truth, after all, is not a formal question – it is a deep philosophical one. However, although the conclusions here are philosophical, there is an important difference from the arguments of Lucas and Penrose: we are still only concerned with consistent formal mathematical systems containing arithmetic. The question with semantical arguments is not whether Gödel's theorems can have relevance to something physiological (the brain) or psychological (human thinking). It is whether Gödel's theorems have relevance on the *very thing they concern*: formal mathematical systems. Obviously we should be quite surprised if the incompleteness of all formal systems did *not* have any philosophical importance, given that it has such a big mathematical importance.

As it happens, if the semantical arguments were not valid, that philosophical importance would be all the more drastic. If formal systems are all there is to mathematics, then by Gödel's result mathematics (presented as a single system, which is the only way extreme formalism can be ultimately conceived of) is incomplete: there are fundamentally undecidable sentences. As far as results in the philosophy of mathematics go, there cannot be many more radical ones. This is something that has been obscured along the decades after Gödel presented his result, but it must be stressed here. Compared to fundamental undecidability, the philosophical conclusion of the semantical arguments – the substantiality of truth – seems much weaker, and much easier to accept. There is a difference between truth and proof, which is due to the incompleteness of formal mathematics. However, this does not imply that there are fundamentally undecidable sentences: different axiomatizations have different Gödel sentences. It is the formal systems, not mathematics as a whole, that are affected by Gödel's incompleteness theorems – which is *exactly* what the actual theorems tell us. To get a fuller, practical and a more realistic

picture of mathematics, we must expand beyond the formal systems. This is what we do in the semantical arguments and Tarskian truth gives us a convenient way of completing the picture when it comes to truth and proof. As far as mathematical thinking is concerned, Gödel's incompleteness theorems do not show that we can beat formal systems: they simply show that formal systems cannot be everything there is to mathematics – which is something we should have suspected anyway, Gödel or no Gödel. So when it comes to the philosophy of mathematics, the semantical Gödelian arguments actually draw *weaker* conclusions from the incompleteness theorems than extreme formalism does. There should be no danger of a Gödelian fallacy here.

8.4 Conclusion: what does "substantial" truth mean?

Throughout this work we have (following earlier discussion) used the term *substantial* (as well as *robust*) for the point of view opposing deflationism. If mathematical truth is not deflatable, it is not metaphysically thin, and thus it is substantial. This has been the line of thinking. The negative nature of this definition is obvious and has clearly worked in favour of the theses presented here. It has been claimed that mathematical truth is indeed substantial, but the exact nature of this apparently metaphysical property has not been clarified. Indeed, what can be the nature – the *essence* – of the property of mathematical truth?

We should examine this problem now. Tennant's whole argument was based on the view that a robust property cannot be derived from other properties, or to be exact, a property is not robust if everything we can achieve with it can also be achieved with other properties. Tennant claimed that Ketland's argument for the robustness of truth does not hold because the Gödel sentence could also be asserted with the help of a soundness principle. Since Tennant's argument is so far the strongest one for the deflationist cause, we can take this as our starting point.

In Tennant's account, with the soundness principle we could assert the Gödel sentence which is not assertable in the formal system. Hence, by assuming a soundness principle we could

derive all the assertable sentences we could with truth. We can derive the extension of mathematical truth from the soundness principles[215], and in Tennant's account, soundness principle is thus a robust property and truth is not. However, this could be only a temporary state of affairs, since the soundness principle may also be derivable from other properties. Indeed, as we saw from Ketland's and Shapiro's work, the soundness principle can be derived from a Tarskian notion of truth, which was thought to be robust. But there is something wrong with this picture: when it comes to the choice between soundness and Tarskian truth, no matter which concept we adopt as the robust one, extensionally speaking, we can derive the other from it. Basically, there seem to be two options: either both of the properties can be robust, or neither one can.

However, there is also a third option: that our concept *robust* needs fixing. I think we have seen enough to conclude that this indeed is the case. Ketland argued that Tennant misunderstood his argument. His point was not that a robust notion of truth was the only way to establish the truth of Gödel sentences. Truth is robust because it is the *best*, most natural, way. This is the crucial point. Of course we could adopt a soundness principle, but as I argued earlier, adopting a Tarskian notion of truth is a more plausible extension. This seems to be the only means we have of defining robustness in any satisfactory way: a property can be called robust, not because it is not derivable from other concepts, but because it is the *most plausible* of the competing alternatives. This could be understood ontologically, epistemologically or pragmatically. But it must not be understood only *logically*. Truth is, at best, a concept with the extension of the quasi-logical T-sentences. The equivalent extension can be achieved also with soundness principles, but that is not the question here. It is the *intension* of the preferred concept that we must base our decision on, and that intension must fit together with a satisfactory epistemological and ontological account of mathematics. In this work, the decision in favour of Tarskian truth is based on two arguments. First, I have argued for

[215] The intensional difference is obvious, however, and quite clearly the one that matters more here, as we will see.

the existence of our pre-formal thinking, and that semantical truth is not an expansion at all in that fuller picture of mathematical thinking. Second, I claim that the epistemological alternative is arbitrariness, which we cannot accept of our mathematical theories.

The basic terminological problem here is that robust truth is often misunderstood to imply a radically realist philosophy of mathematics, usually an archaic version of Platonism. This is something we have to rid the philosophy of mathematics of. The question of truth in mathematics is not the same question as the existence of mathematical objects, or even that of objective truth-values. Of course truth *can* be conceived as the correspondence between mathematical statements and mathematical objects in Platonism, just like truth can obviously be seen as the property we refer to with objective truth-values. However, the question of truth comes up *before* all these metaphysical questions when we try to construct a comprehensive philosophical account of mathematics. Everything considered in this work points to that: Tarskian undefinability, Gödel's incompleteness theorems and the semantical arguments based on them, Hintikka's and Kripke's truth, second-order logic, pre-formal thinking, the impossibility of extreme formalism and the hierarchies that follow from truth predicates. At no point have we had the need to make any metaphysical or epistemological considerations even remotely resembling Platonism. This is an extremely important point: we have arrived at the need for a robust (once again, defined as non-deflationary) truth predicate with the absolute minimum of presuppositions – ultimately insisting only on that mathematical theories are not arbitrary. The highly problematic metaphysical and epistemological problems concern the *nature* of truth. Those are of course some of the most important questions in the philosophy of mathematics – but they are not the question whether truth and proof are the same concept. We need robust truth primarily because deflationary truth fails in mathematics. Perhaps we could arrive at the robustness of truth in other ways – possibly more constructive ones – but the destructive argument here on extreme formalism seems just as valid nevertheless.

Of course the negative nature of the definition of robustness remains, but this should not be seen as a problem. The exact nature of mathematical truth is bound to continue puzzling philosophers, whether it is thought in terms of Platonism, conventionalism, naturalism, empiricism or in some other way. What mathematical truth is, and how the manifold physical and direct applications of mathematical theories can be accounted for, are probably the two most difficult questions in the philosophy of mathematics. However, although quite clearly relevant, ultimately they both fall outside the scope of this work. The problem studied here is whether mathematical truth can be viably seen as the same concept as formal proof, and from all we have seen, the answer is no. It should suffice here to conclude that the road of explaining mathematical truth is indeed the one to take, rather than trying to deflate mathematics completely into formal systems, and ominously arbitrary soundness principles.

References

Ayer, Alfred Jules
1946 *Language, Truth & Logic*, Second Edition, Dover, New York 1952.

Azzouni, Jody
1994 *Metaphysical myths, mathematical practice*, Cambridge University Press, Cambridge.

1999 "Comments on Shapiro", *The Journal of Philosophy*, Vol. XCVI, pp. 541-544.

Barrow, John D.
1992 *Pi In The Sky – Counting, Thinking and Being*, Black Bay Books, New York.

Barwise, Jon (ed.)
1977 *Handbook of Mathematical Logic*, North Holland Publishing, New York.

Belnap, Nuel D.
1962 "Tonk, Plonk and Plink", *Analysis* Vol. 22, pp. 130-134. Printed in Strawson (ed.) 1967.

Benacerraf, Paul
1973 "Mathematical Truth", *The Journal of Philosophy*, Vol. LXX, pp. 661-679.

Benacerraf. Paul & Putnam, Hilary (ed.)
1964 *Philosophy of Mathematics*, Prentice-Hall, Englewood Cliffs, New Jersey.

1983 *Philosophy of Mathematics*, Second Edition, Cambridge University Press, Cambridge.

Berger, U. & Schwichtenberg, Helmut (eds.)
1999 *Computational Logic*, Springer-Verlag, Heidelberg.

Boolos, George
1998 *Logic, Logic and Logic,* Richard Jeffrey and John P. Burgess (eds.) Harvard University Press, Cambridge, Massachusetts.

Bourbaki, Nicolas
1970 *Eléments de Mathématiques, Théorie des Ensembles*, Hermann, Paris.

Boyer, Carl B.
1985 *A History of Mathematics*, Princeton University Press, Princeton, NJ.

Brandom, Robert
1994 *Making It Explicit*, Harvard University Press, Cambridge, Massachusetts.

Burgess, John P. & Rosen, Gideon
1997 *A Subject with No Object: Strategies for Nominalistic Interpretation of Mathematics*, Clarendon Press, Oxford.

Cantor, Georg
1883 *Grundlagen eine allgemeine Mannigfaltigkeitslehre,* translated as *Foundations of a General Theory of Manifolds*, in Ewald, William (ed.) 1996.

Carnap, Rudolf
1956 "Empiricism, Semantics and Ontology" in Benacerraf & Putnam (eds.) 1983, pp. 241-257.

Chihara, Charles
1973 *Ontology and the Vicious Circe Principle*, Cornell University Press, Ithaca, NY

1990 *Constructibility and Mathematical Existence*, Oxford University Press, Oxford.

2005 "Nominalism", in Shapiro, Stewart (ed.) 2005, pp. 483-514.

Church, Alonzo
1956 *Introduction to Mathematical Logic*, Princeton University Press, Princeton.

Cohen, Paul J.
1963 "The Independence of the Continuum Hypothesis", *Proceedings of the National Academy of Sciences of the United States of America* 50 (6), pp. 1143-1148.

Crowell, Richard & Fox, Ralph
1963 *Introduction to Knot Theory*, Ginn and Co., Boston.

Curry, Haskell B.
1954 "Remarks on the definition and nature of mathematics", *Dialectica* 8. Printed in Benacerraf & Putnam 1983, pp. 202-206.

Dales, H.G. & Oliveri, G. (eds.)
1998 *Truth in Mathematics*, Clarendon Press, Oxford.

Davis, Robert B.
1984 *Learning Mathematics*, Croom Helm, Sydney, Australia.

Debray, Régis
1983 *Critique of Political Reason*, Translated by David Macey, New Left Books, London.

De Rouilhan, Philippe
2002 "On What There Are", *Proceedings of the Aristotelian Society* 102, pp. 183-200.

De Rouilhan, Philippe & Bozon, Serge
2003 "The Truth of IF: has Hintikka really exorcised Tarski's curse" in Hintikka 2006, pp. 683-705.

Detlefsen, Michael
1980 "On a Theorem of Feferman", *Philosophical Studies* 38, 129-140.

1986 *Hilbert's Program*, D. Reidel Publishing, Doordrecht, Holland.

Dummett, Michael
1956 "Nominalism", *Philosophical Review* 65, pp. 491-505.

1976 *The Logical Basis of Metaphysics*, Harvard University Press, Cambridge, Massachusetts.

1977 *Elements of Intuitionism*, Oxford University Press, London.

1978 *Truth and Other Enigmas*, Duckworth, London.

Eklund, Matti
2006 "Neo-Fregean Ontology", *Philosophical Perspectives*, Vol. 20, Issue 1, pp. 95-121.

2007 "Bad Company and Neo-Fregean Philosophy", *Synthese*, published online.

Enderton, Herbert B.
1977 *Elements of Set Theory*, Academic Press, Cambridge.

Ewald, William (ed.)
1996 *From Kant to Hilbert*, Clarendon Press, Oxford.

Feferman, Solomon
1991 "Reflecting on Incompleteness", *Journal of Symbolic Logic*, 46, pp. 1-49.

1998 "Infinity in Mathematics: Is Cantor Really Necessary?", in *In The Light of Logic*, Oxford University Press, Oxford, pp. 229-248.

2006 "What kind of logic is 'Independence Friendly' logic?" in Hintikka 2006, pp. 453-469.

Field, Hartry
1972 "Tarski's theory of truth", *Journal of Philosophy* 69, pp. 347-375.

1980 *Science Without Numbers*, University Press, Princeton.

1984 "Critical Notice of Crispin Wright: *Frege's Conception of Numbers as Objects*, in Field 1989, pp. 147-170.

1989 *Realism, Mathematics and Modality*, Oxford University Press, Oxford.

1998 "Which undecidable mathematical sentences have determinate truth values?", in Dales & Oliveri (eds.) 1998, pp. 291-310.

1999 "Deflating The Conservativeness Argument", *The Journal of Philosophy*, Vol. XCVI, pp. 533-540.

2001 *Truth and The Absence of Fact*, Clarendon Press, Oxford.

2006 "Truth and the Unprovability of Consistency", *Mind* Vol. 115, pp. 567-506.

Forster, Thomas

2006 "Deterministic and Nondeterministic Strategies for Hintikka games in First-order and Branching-quantifier logic", *Logique et Analyse* Vol. 195, pp. 265-269.

Franzén, Torkel

2005 *Gödel's Theorem: An Incomplete Guide To Its Use And Abuse*, A K Peters, Wellesley, Massachusetts.

Frege, Gottlob

1884 *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl*. Translated as *The Foundations of Arithmetic: A logico-mathematical enquiry into the concept of number*, 2nd ed. Blackwell 1974.

1892 *Über Sinn und Bedeutung*, translated as *On Sense and Reference*, in A.W. Moore (ed.) *Meaning and Reference*, Oxford University Press, Oxford.

Friedman, Harvey

1976 "Systems of second order arithmetic with restricted induction", *Journal of Symbolic Logic* Vol. 41, pp. 557-559.

Gentzen, Gerhard

1936 "Die Widerspruchfreiheit der reinen Zahlentheorie'. *Mathematische Annalen* 112, 493–565. Translated as "The consistency of arithmetic" in Szabo 1969.

Girard, Jean-Yves
1999 "On the meaning of logical rules I: syntax vs. semantics, in Berger and Schwichtenberg (eds.) 1999, pp. 215-272.

Goble, Lou (ed.)
2001 *The Blackwell Guide to Philosophical Logic*, Wiley-Blackwell, Oxford.

Goldman, Alvin
1967 "A Causal Theory of Knowing", *Journal of Philosophy* 64, 357-372.

Goodman, Nelson
1956 "A World of Individuals" in Benacerraf. Paul & Putnam, Hilary (ed.) 1964, pp. 197-210.

Goodman, Nelson & Quine, W.V.O.
1947 "Steps toward a Constructive Nominalism", *Journal of Symbolic Logic* 12, pp. 97-122.

Gupta, Anil & Belnap, Nuel
1993 *The Revisionist Theory of Truth*, MIT Press, Cambridge, Massachusetts.

Gödel, Kurt
1931 "On formally undecidable propositions", *Collected Works Volume I*, pp. 145-195. Oxford University Press, New York 1986.

1932 "Completeness and consistency", *Collected Works Volume I*, pp. 235-236. Oxford University Press, New York 1986.

1940 *The Consistency of the Continuum-Hypothesis*. Princeton University Press, Princeton, NJ.

1946 "Remarks before the Princeton bicentennial conference on problems in mathematics", *Collected Works Volume II*, pp. 150-153, Oxford University Press, New York 1990.

1951 "Some basic theorems on the foundations", *Collected Works Volume III*, pp. 304-323. Oxford University Press, New York 1995.

1958 "On a hitherto unutilized extension of the finitary standpoint", *Collected Works Volume II*, pp.241-251. Oxford University Press, New York 1990.

1964a "On undecidable propositions of formal mathematical systems, Postscriptum 1964", *Collected Works Volume I*, pp. 369-370. Oxford University Press, New York 1986.

1964b "What is Cantor's continuum problem", printed in Benacerraf & Putnam 1983, 470-485.

Haack, Susan
1978 *Philosophy of Logics*, Cambridge University Press, Cambridge.

Hadamard, Jacques
1954 *The Psychology of Invention In The Mathematical Field*, Dover Publications, Mineola, NY.

Halbach, Volker
1999 "Conservative Theories of Classical Truth", *Studia Logica* 62, pp. 353-370.

2001 "How Innocent is Deflationism?", *Synthese* 126, pp. 167-194.

Hale, Bob
1977 "Realism and its Oppositions" in Hale & Wright (eds.) 1997, pp. 271-308.

1987 *Abstract Objects*, Basil Blackwell, Oxford.

Hale, Bob & Wright, Crispin
2001 *The Reason's Proper Study*, Clarendon Press, Oxford.

2005 "Logicism In The Twenty-First Century" in Shapiro, Stewart (ed.) 2005, pp. 166-202.

Hale, Bob & Wright, Crispin (eds.)
1997 *A Companion to the Philosophy of Language*, Basil Blackwell, Oxford.

Hart, W.D. (ed.)
1995 *The Philosophy of Mathematics*, Oxford University Press, Oxford.

Hellman, Geoffrey
1989 *Mathematics without numbers*, Oxford University Press, Oxford.

2005 "Structuralism", in Shapiro, Stewart (ed.) 2005, pp. 536-562.

Heyting, Arend
1931 "The intuitionistic foundations of mathematics", printed in Benacerraf & Putnam 1983, 52-61.

Hilbert, David
1900 *Grundlagen der Geometrie*, translated as *The Foundations of Geometry*, Open Court, Chicago 1902.

1925 "Über das Unendliche", *Mathematische Annalen* 95, 161-190. Translated "On The Infinite" in Benacerraf & Putnam 1983, pp. 183-201.

1970 *Grundlagen der Mathematik I*, 2nd edition Springer, Berlin.

Hintikka, Jaakko
1996 *Principles of Mathematics Revisited*, Cambridge University Press, Cambridge.

2001 "Introduction and postscript", *Synthese*126, pp. 1-16.

2006 *The Philosophy of Jaakko Hintikka*, Open Court, USA.

Hintikka, Jaakko (ed.)
1969 *The Philosophy of Mathematics*, Oxford University Press, Oxford.

Hintikka, Jaakko & Sandu, Gabriel
1996 "A Revolution in Logic?", *Nordic Journal of Philosophical Logic*, Vol. 1, No. 2, pp. 169--183.

Horwich, Paul
1998 *Truth*, Clarendon Press, Oxford.

Hyttinen Tapani & Sandu, Gabriel
2000 "Henkin quantifiers and the definability of truth", *Journal of Philosophical Logic* 29, pp. 507-527.

2004 "Deflationism and arithmetical truth", *Dialectica* Vol. 58, pp. 413-426.

Jané, Ignacio
2005 "Higher-order logic reconsidered", in Shapiro (ed.) 2005, pp. 781-810.

Jones, W.T.
1980 *A History of Western Philosophy: The Classical Mind, Second Edition*, Harcourt Brace Jovanovich, USA.

Ketland, Jeffrey
1999 "Deflationism and Tarski's Paradise", *Mind* 108, pp. 69-94.

2003 "Can a many-valued language functionally represent its own semantics?", *Analysis* 63, pp. 292-297.

2005 "Deflationism and the Gödel Phenomena: Reply to Tennant", *Mind*, Vol. 114, pp. 75-88.

Kirkham, Richard L.
1992 *Theories of Truth*, MIT Press, Cambridge, Massachusetts.

Kitcher, Philip
1983 *The Nature of Mathematical Knowledge*, Oxford University Press, New York.

Kremer, Michael
1988 "Kripke and the logic of truth", *Journal of Philosophical Logic* 17, pp. 225-278.

Kripke, Saul
1975 "Outline of a Theory of Truth", *The Journal of Philosophy* 72, 1975, pp. 690-716.

Kristeva, Julia
1969 *Séméiotiké: Recherches pour une sémanalyse*, Éditions du Seuil, Paris.

Kuhn, Thomas
1962 *The Structure of Scientific Revolutions*, Chicago University Press, Chicago.

Lakatos, Imre
1978 *Philosophical papers Volume 2: Mathematics, science and epistemology*, Cambridge University Press, Cambridge.

Lavine, Saughan
1994 *Understanding the Infinite*, Harvard University Press, Cambridge, Massachusetts.

Levy, Azriel
1965 *A Hierarchy of formulas in set theory*, Memoirs of American Mathematical Society 57.

Lewis, David
1993 "Mathematics is megethology", *Philosophia Mathematica* (3) 1, pp. 3-23.

Linnebo, Øystein
2006b "Frege's Context Principle and Reference to Natural Numbers", forthcoming in *Logicism, Intuitionism and Formalism – What Has Become of Them?*, eds. S. Lindström et al, Springer.

2007 "The Nature of Mathematical Objects", forthcoming in *Current Issues in the Philosophy of Mathematics from the Perspective of Mathematicians*, ed. B.Gold, Mathematics Association of America.

Lucas, John
1961 "Minds, Machines and Gödel", *Philosophy* 36 pp. 112-127.

Löb, Martin
1955 "Solution of a problem of Leon Henkin", *The Journal of Symbolic Logic* 20, pp.115-118.

Maddy, Penelope
1984 "Mathematical Epistemology: What Is the Question", *The Monist* 67, pp. 46-55.

1997 *Naturalism in Mathematics*, Oxford University Press, Oxford.

Meschkowski, Herbert
1973 *Hundert Jahre Mengenlehre*, Deutscher Taschenbuch Verlag, München.

Mill, John Stuart
1874 *A System of Logic*, Harper & Brothers, New York.

Niiniluoto, Ilkka
1994 "Defending Tarski against his Critics" in Twardowski & Wolenski (eds.)1994, pp. 48-68.

Omnès, Roland
1999 *Understanding Quantum Mechanics*, Princeton University Press, Princeton.

Penrose, Roger
1989 *The Emperor's New Mind*, paperback edition, Oxford University Press, New York.

1994 *Shadows of the Mind*, Oxford University Press, London.

2004 *The Road to Reality*, Random House, London.

Plato
1992 *The Republic*, second edition. Translated by G.M.A Grube, Hackett Publishing Company, Indianapolis.

1996 *Parmenides*, Translated by Mary Louise Gil and Paul Ryan, Hackett Publishing Company, Indianapolis.

Prior, A.N.
1960 "The Runabout Inference Ticket", *Analysis* Vol. 21, pp. 38-39. Printed in Strawson (ed.) 1967.

Putnam, Hilary
1961 "Minds and Machines" in Hook 1961, 142.

1967 "Mathematics without foundations", *Journal of Philosophy* 64, pp. 5-22. Printed in Benacerraf. Paul & Putnam, Hilary (ed.) 1964, pp. 295-311.

1971 *Philosophy of Logic*, Harper and Row, New York.

1975 "Do True Assertions Correspond to Reality", in *Mind, Language and Reality, Philosophical Papers Vol. 2*, pp. 70-84. Cambridge University Press, Cambridge.

1980 "Models and reality", *Journal of Symbolic Logic* 45, pp. 464-482.

Quine, W.V.O.
1951 "Two Dogmas of Empiricism", *Philosophical Review* 60/1, pp. 20-43. Printed in Hart, W.D. (ed.) 1996, pp. 31-51.

1966 "The Scope of Language of Science", in *The Ways of Paradox and Other Essays* pp. 215-232. Random House, New York.

1986 *Philosophy of Logic*, Second edition, Prentice-Hall, Englewood Cliffs, New Jersey.

1990 *Pursuit of Truth*, Harvard University Press, Cambridge, Massachusetts.

1995 *From Stimulus to Science*, Harvard University Press, Cambridge, Massachusetts.

Ramsey, Frank P.
1927 "Facts and Propositions" in D.H. Mellor (ed.) *F.P. Ramsey: Philosophical Papers*, Cambridge University Press, New York, pp. 34-51.

Rayo, Agustín
2005 "Logicism Reconsidered", in Shapiro (ed.) 2005 pp. 203-236.

Reid, Constance
1970 *Hilbert*, Springer, Berlin.

Resnik, Michael D.
1981 "Mathematics as a science of patterns: Ontology and reference", *Nous* 15, pp. 529-550.

1982 "Mathematics as a science of patterns: Epistemology, *Nous* 16, pp. 95-105.

1988 "Second-order logic still wild", *Journal of Philosophy* 85, pp. 75-87.

Robinson, Raphael M.
1950 "An Essentially Undecidable Axiom System" in *Proceedings of the International Congress of Mathematics*: pp. 729-730.

Rosser, J. Barkley
1936 "Extensions of some theorems of Gödel and Church", *Journal of Symbolic Logic* Vol.1, pp. 87-91.

Russell, Bertrand
1912 *The Problems of Philosophy*, Williams and Norgate, London.

1920 *Introduction to Mathematical Philosophy*, Second Edition, Dover, New York 1993.

Sandu, Gabriel
1996 "IF First-Order Logic, Kripke, and 3-Valued Logic", in Hintikka 1996, pp. 254-270.

Shapiro, Stewart
1983 "Conservativeness and Incompleteness", The *Journal of Philosophy*, 80, pp. 521-531.

1997 *Philosophy of Mathematics: Structure and Ontology*, Oxford University Press, New York.

1998 "Proof and Truth: Through Thick and Thin", *The Journal of Philosophy*, Vol. XCV, pp. 493-521.

2000a *Thinking About Mathematics*, Oxford University Press, Oxford.

2000b *Foundations Without Foundationalism: A Case for Second-order Logic*, Oxford University Press, Oxford.

2005 "Higher-order Logic", in Shapiro (ed.) 2005, pp. 751-780.

Shapiro, Stewart (ed.)
2005 *The Oxford Handbook of Philosophy of Mathematics and Logic*, Oxford University Press, New York.

Simpson, Stephen G.
1999 *Subsystems of Second Order Arithmetic*, Springer-Verlag, Berlin.

Smorynski, C.
1977 "The Incompleteness Theorems", in Barwise 1977, pp. 821-865.

Smullyan, Raymond
1992 *Gödel's Incompleteness Theorems,* Oxford University Press, Oxford.

2001 "Gödel's Incompleteness Theorems", in Goble, Lou (ed.) 2001, pp. 72-89.

Sokal, Alan & Bricmont, Jean
2003 *Intellectual Impostures*, Second Edition, Profile Books, London.

Stewart, Ian
2006 *Letters to a Young Mathematician*, Basic Books, New York.

Strawson, P.F. (ed.)
1967 *Philosophy of Logic*, Oxford University Press, 1967, London.

Suppes, Patrick
1959 *Introduction To Logic*, D. van Nostrand Company, New York.

Szabo, M.E. (ed.)
1969 *The Collected Works of Gerhard Gentzen*, North-Holland, Amsterdam.

Tall, David (ed.)
1994 *Advanced Mathematical Thinking,* Kluwer, Dordrecht, The Netherlands.

Tarski, Alfred
1936 "Das Wahrheitsbegriff in formalisierten Sprachen", *Studia Philosophica* I, pp. 261-405. Translated in English by J.H. Woodger (1956) as "The Concept of Truth in Formalized Languages", in *Logic, Semantics, Metamathematics,* pp. 152-278, Hackett, Indianapolis 1983.

1944 "The Semantic Conception of Truth and the Foundations of Semantics", *Philosophy and Phenomenological Research* 4, 341-376.

1959 "What Is Elementary Geometry" in Hintikka, Jaakko (ed.) 1969.

1969 "Truth and Proof", *Scientific American* 220, pp. 63-77. Printed in *Collected Papers Vol. 4*, pp. 399-424. Birkhäuser, Basel 1986.

Tennant, Neil
2002 "Deflationism and the Gödel Phenomena), *Mind,* Vol. 111, pp. 551-582.

2005 "Deflationism and the Gödel Phenomena: Reply to Ketland", *Mind*, Vol. 114, pp. 89-96.

Turing, Alan
1937 "On Computable numbers, with an application to the Entscheidungproblem", *London Mathematical Society 42*, 230 –265.

1950 "Computing Machinery and Intelligence", *Mind* 1950, pp. 433-460.

Twardowski, Bartlomiej & Wolenski, Jan (eds.)
1994 *Sixty Years of Tarski's Definition of Truth*, Philed, Krakow.

Väänänen, Jouko
2007 *Dependence Logic: A New Approach to Independence Friendly Logic*, Cambridge University Press, Cambridge.

Wang, Hao
1974 *From Mathematics To Philosophy*, Routledge, London.

1987 *Reflections on Kurt Gödel*, MIT Press, Cambridge Massachusetts.

Whitehead, A.N. & Russell, Bertrand
1956 *Principia Mathematica*, abridged edition, Cambridge University Press, Cambridge.

Williams, Michael
1999 "Meaning and deflationary truth", *The Journal of Philosophy*, Vol. XCVI, pp. 545-564.

Wittgenstein, Ludwig
1976 *Lectures on the Foundations of Mathematics*, Harvester Press, Hassocks, Sussex.

1983 *Remarks on The Foundation of Mathematics (Revised Edition)*, MIT Press, Cambridge, Massachusetts.

Woleński, Jan
2001 "In Defense of The Semantic Definition of Truth", *Synthese* 126, pp. 67-87.

2006 "Tarskian and Post-Tarskian truth", in Hintikka 2006, pp. 647-672.

Wright, Crispin
1983 *Frege's Conception of Numbers as Objects*, Aberdeen University Press, Aberdeen.

1987 *Realism, Meaning and Truth*, Basil Blackwell, Oxford.

1992 *Truth & Objectivity*, Harvard University Press, London.