The epistemology of abstractionism

Alexander Christoph Reinhard Oldemeier

June, 2012

Submitted in accordance with the requirements for the degree of PhD.

The University of Leeds

Department of Philosophy

Acknowledgements

First and foremost, I thank my patient main supervisor, Robbie Williams, for his great support. I have learned so much. I also thank my supervisor John Divers. Without his guidance, I might not have pursued this project in the first place.

My research has been made possible by the kind support of the Leeds Philosophy Department, the AHRC, the Jacobsen Trust of the Royal Institute of Philosophy, and the Northern Institute of Philosophy. I am deeply grateful for that.

I also thank all my fellow PhD students at Leeds, from whom I received a lot of valuable feedback and criticism. In particular: Thomas Brouwer, Michael Bench-Capon, Mirja Holst, Wouter Kalf, Stephan Kraemer, and Duncan Watson.

Moreover, I am deeply grateful for all the support I received from the Northern Institute of Philosophy (NIP). Thanks to the NIP, I am the proud owner of a Frege T-Shirt. I owe a great deal to its members. In particular: Aaron Cotnoir, Dylan Dodd, Sharon Coull, Filippo Ferrari, Andreas Fjellstad, Aidan McGlynn, Giacomo Melis, Grant Reaber, Björn Schotte, Paula Sweeney, and Crispin Wright.

I profited immensely from discussions with and constructive criticism from the following philosophers I did not yet mention: Philip Ebert, Ole Hjortland, Joachim Horvath, Nikolaj Pedersen, and Walter Swetly.

I am grateful to everyone who was involved in developing the free software that helped me complete and manage this project. It would have been much more difficult to produce all this without Mercurial, LaTeX, MiKTeX, LyX, JabRef, Dia, and FreeMind.

And I thank everyone who made my non-philosophy time in Leeds an excellent time. In particular, my long-time housemates: Billy, Hermione, Katie, Joe, Lucy, and Oli. Ben and Alasdair. And also: Joe, Martin, and Vasiliki for their hospitality in times of writing up.

Moreover, I thank all the people who made my time an excellent time in general. In particular, my good friends from Germany: Anja, Annika, Daniel, David, Georg, Isabella, and Moritz. Without your emotional support over many years, directly and indirectly, I would not have been able to complete this project.

The same holds true for my parents, Andrea and Matthias Oldemeier, who enabled me to study philosophy in the first place, for my brothers Dominik and Philipp, my sister Michelle, for Tina, and for Sofía. There is (much) more to say about you, but I do not know how to do justice to you in this context. Let me put it simply: you have been crucial.

Unfortunately, it is highly probable that this list is incomplete. If you would have liked to see your name on this list, but did not find your name on it, please accept my apologies.

Munich, 30 Oct 2012

Abstract

I examine the nature and the structure of basic logico-mathematical knowledge. What justifies the truth of the Dedekind-Peano axioms and the validity of Modus Ponens? And is the justification we possess reflectively available? To make progress with these questions, I ultimately embed Hale's and Wright's neo-Fregeanism in a general internalistic epistemological framework.

In *Part I*, I provide an introduction to the problems in the philosophy of mathematics to motivate the investigations to follow. I present desiderata for a fully satisfactory epistemology of mathematics and discuss relevant positions. All these positions turn out to be unsatisfactory, which motivates the abstractionist approach. I argue that abstractionism is in need of further explication when it comes to its central epistemological workings.

I fill this gap by embedding neo-Fregeanism in an internalistic epistemological framework. In *Part II*, I motivate, outline, and discuss the consequences of the framework. I argue: (1) we need an internalistic notion of warrant in our epistemology and every good epistemology accounts for the possession of such warrant; (2) to avoid scepticism, we need to invoke a notion of non-evidential warrant (entitlement); (3) because entitlements cannot be upgraded, endorsing entitlements for mathematical axioms and validity claims would entail that such propositions cannot be claimed to be known.

Because of (3), the framework appears to yield sceptical consequences. In *Part III*, I discuss (i) whether we *can* accept these consequences and (ii) whether we *have to* accept these consequences. As to (i), I argue that there is a tenable solely entitlement-based philosophy of mathematics and logic. However, I also argue that we can overcome limitations by vindicating the neo-Fregean proposal that implicit definitions can underwrite basic logico-mathematical knowledge. One key manoeuvre here is to acknowledge that the semantic success of creative implicit definitions rests on substantial presuppositions — but to argue that relevant presuppositions are entitlements.

Contents

Introduction

Motivation

We claim to know many things. We claim to be justified in believing many things. I know that I have two hands, that 1+1=2, and that Modus Ponens is a valid rule of inference. I am justified in believing that it is raining outside, because I *see* it. I am justified in believing that there are infinitely many prime numbers, because I can *prove* it.

Knowledge and justification matter to us. Knowledge is a guide to action. Using an example of Williamson's, suppose "a burglar spends all night ransacking a house, risking discovery by staying so long" (Williamson 2000, p. 62). Clearly, that the burglar *knows* that there is a diamond in a house explains this behaviour. Contrast the explanation that the burglar merely *hoped* that there is a diamond in the house. This would be very implausible as an explanation — at least under normal circumstances.

Justification enables us to give satisfactory *responses* to the doubts of others, and ultimately enables us to *convince* others. Only a network of justified beliefs will withstand critical scrutiny. Without our propensity to gather justified beliefs, modern science — with all its practical advantages — would be impossible.

Knowledge and justification are extended in the context of a plethora of *cognitive projects*: projects of finding out about the world, using a variety of cognitive and sensory capacities. Such projects can be very general, as the project of obtaining knowledge about the physical world or the project of discovering the structure of the natural numbers. However, cognitive projects can also be very specific, as the project of determining how many words this thesis has.

The success of some projects rests on the possibility of the success of other projects. For example, the project of determining what my fair share of a bag of Skittles is rests on the possibility of finding out about the external world by visual experience, and the possibility of dividing the number of sweets in the bag.

In this sense, logic and mathematics are *fundamental projects*, because a great many projects rest on their success (and the possibility of their success). Without mathematics — the project of discovering the mathematical facts — contemporary physics would be radically impaired. Without logic — the project of discovering what, in general, (deductively) follows from what — determining the (logical) consequences of our non-logical beliefs would be impossible, which would radically impair our cognitive lives in almost every respect. Not only because we would not be able to extend knowledge by means of logical reasoning. But also because logic is required for determining the consistency of our beliefs, and for rational (self-)criticism.

In this thesis, I shall provide epistemological foundations for mathematics (and some outlines for logic) by arguing for a particular option. In *Part I*, I provide an introduction to the problems in the philosophy of mathematics to motivate the investigations to follow. I present desiderata for a fully satisfactory epistemology of mathematics and discuss relevant positions. All these positions turn out to be unsatisfactory, which motivates the abstractionist approach. I argue that abstractionism is in need of further explication when it comes to its central epistemological workings.

I fill this gap by embedding neo-Fregeanism in an internalistic epistemological framework. In *Part II*, I motivate, outline, and discuss the consequences of the framework. I argue: (1) we need an internalistic notion of warrant in our epistemology and every good epistemology accounts for the possession of such warrant; (2) to avoid scepticism, we need to invoke a notion of non-evidential warrant (entitlement); (3) because entitlements cannot be upgraded, endorsing entitlements for mathematical axioms and validity claims would entail that such propositions cannot be claimed to be known.

Because of (3), the framework appears to yield sceptical consequences. In *Part III*, I discuss (i) whether we *can* accept these consequences and (ii) whether we *have to* accept these consequences. As to (i), I argue that there is a tenable solely entitlement-based philosophy of mathematics and logic. However, I also argue that we can overcome limitations by vindicating the neo-Fregean proposal that implicit definitions can underwrite basic logico-mathematical knowledge. One key manoeuvre here is to acknowledge that the semantic success of creative implicit definitions rests on substantial presuppositions — but to argue that relevant presuppositions are entitlements.

This proposal is an embedding of (classical) neo-Fregeanism in the epistemological framework discussed in Part II. Thus, I show how to kill two birds with one stone: I provide an explication of the epistemological workings of classical neo-Fregeanism, and thereby provide a non-sceptical internalist epistemology for mathematics, which *might* also be applicable to logic.

One final aspect of my proposal is worth stressing: this thesis, and its final proposal, is in the spirit of what has been called the *Traditional Epistemic Project*, i.e. the project — famously initiated by Descartes — of vindicating from scratch and from the armchair our right to claim knowledge of most of the knowledge we pre-theoretically take ourselves to possess, bracketing all antecedently held beliefs about the external world . I will engage in this project by telling an epistemological *Hero Story* — a story of a subject successfully engaging in this project. Of course, the idea is that Hero could be *you*.

Anticipation

The following paragraphs provide a brief, but slightly more detailed overview over the content of the chapters of this thesis.

Part I In *Part I*, I motivate the neo-Fregean position and argue that it is in need of further clarification when it comes to its exact epistemological workings.

Chapter 1 Chapter 1 provides a brief overview over some issues and options in the epistemology of mathematics. I begin with presenting Benacerraf's famous dilemma and Field's generalization of it. I then define some key terms, and extract some desiderata for a satisfactory epistemology for arithmetic. After that, I provide an (incomplete) overview over the space of options, by providing brief discussions of some of the most important Platonist positions: Gödelian Platonism, Frege's Logicism, and the indispensability argument. I argue that these positions fail to meet our constraints, or face other substantial difficulties. I then examine whether this motivates giving up Platonism. I argue that this is not so, because the nominalist positions also face substantial difficulties. This sets the stage for the position I investigate in chapter 2 — neo-Fregeanism, or: abstractionism.

Chapter 2 Chapter 2 is an opinionated survey of Hale's and Wright's neo-Fregeanism. I motivate and outline the position, and discuss three important objections to it. In the course of the discussion, it will become apparent that the proposal is in need of explication when it comes to its exact epistemological workings.

In some more detail: I sketch how neo-Fregeanism emerged from an analysis of Frege's failures and achievements. After that, I present the position's core components, and sketch how exactly they are supposed to work together. The epistemological core component of neo-Fregeanism — the proposal that apriori knowledge of abstraction principles can be obtained by means of implicit definition — will be investigated in some more detail. It shall transpire that it is unclear how exactly the postulated belief-forming process and the process of warrant generation are supposed to work. Among other things, the proposal relies on an unclear notion of ***warrant by default.*** Hale and Wright later suggest that it might be explicated by Wright's notion of *entitlement*, but how exactly the notion is to be applied remains open.

I close with presenting the three most pressing objections to classical neo-Fregeanism: the *Caesar problem*, the *Bad Company objection*, and *epistemic rejectionism.* Towards the end of the thesis, I will return to these problems, and show that my own proposal sheds light on them.

Part II The epistemological gaps in Hale's and Wright's proposal shall be closed by embedding their idea about implicit definition in a general epistemological framework. *Part II* contains a motivating introduction to as well as a detailed discussion of the framework I endorse for this purpose. For the sake of clarity, I endorse a particular explication of the type of framework I am advocating, building upon Wright's most recent epistemological works. My framework is based on three tenets: (access) internalism, (internalistic) warrant by default, and the transmission-failure diagnosis for Moorean argument. There is one chapter for each tenet.

Chapter 3 Chapter 3 is on internalism. I begin by outlining two debates in epistemology to which the issue of internalism vs. externalism is relevant. After drawing some distinctions, I argue for a particular version of what I call *Relevance Internalism*: the claim that every satisfactory epistemology has to explain how we can possess ***rationally claimable warrants*** — warrants whose possession is available on the basis of self-knowledge and introspection. I argue for Relevance Internalism by arguing for the claims that (i) the most interesting sceptical challenges are directed against our right to claim warrants; (ii) externalistic notions cannot be used to provide dialectically stable responses to simple closure-based sceptical challenges; (iii) that successfully pursuing what has been called the

Traditional Epistemic Project requires accounting for the possession of rationally claimable warrants in the sense above. I argue that some of these considerations can also be applied to the arithmetical case.

Chapter 4 Relevance Internalism invites scepticism. In chapter 4, I argue that the internalist can avoid scepticism, but that he or she needs to invoke internalistic *warrants by default,* i.e. internalistic warrants one can possess without having done any prior evidential work. I endorse Wright's entitlements of cognitive project to render the envisaged response to scepticism explicit.

After drawing some relevant distinctions, I argue for a position that has been called conservativism. Conservativism about perception and deduction makes it hard to see how we can avoid certain radical forms of scepticism. This is because it renders it hard to see how we can acquire first evidential warrants for relevant propositions at the basic level, and it makes it impossible to claim any warrants because second-order arguments become viciously circular. I present a very general argument to the effect that, in order to avoid scepticism, every access internalist needs to invoke a notion of an internalistic warrant by default at the basic level, i.e. a notion of an internalistic warrant one can possess without having done any prior evidential work.

I render this response to scepticism explicit by endorsing Wright's notion of entitlement of cognitive project. After motivating and introducing the basic idea, I provide two models of how exactly entitled presuppositions might serve the generation of evidential warrants sufficient for claimable knowledge.

Chapter 5 The moral of chapter 5 is that it matters a lot in terms of the consequences of our epistemological framework, what we regard to be the canonical structure of justification in relevant areas of cognitive enquiry.

The argument goes as follows. First, I point out some aspects in which Wright's entitlements are weak warrants. In particular, I agree with Wright in that entitled true belief cannot amount to claimable knowledge. One might think that one can avoid this consequence by epistemically upgrading entitled basic propositions by some form of bootstrapping. However, secondly, I argue that such bootstrapping fails. This is due to the phenomenon of *failure of warrant transmission*, which Wright uncovers in his reflections on Moorean arguments. Responding to scepticism by invoking entitlements is inevitably concessive in the sense that certain basic propositions — what I call the *presuppositions* of basic belief-forming methods — can never be claimed to be known. This might yield sceptical consequences.

As an example, I consider the logical case. I argue that, because of conservativeness about deduction, rule-circular arguments are just as bad as Moorean arguments. Thus, if validity claims were entitlements, then they could never be claimed to be known. This would be a revisionary sceptical consequence.

This motivates looking for basic belief-forming methods that allow for justifying validity claims without invoking validity as a presupposition. In general, what the basic belief-forming methods are determines what the presuppositions are. Finding a suitable structure of justification might avoid revisionary sceptical consequences.

Part III What has just been said is obviously relevant to the mathematical case. For example, if axioms are entitlements, then they cannot be claimed to be known. This would also be a revisionary sceptical consequence. In *Part III*, I answer two questions regarding these limitative results. Firstly, can we bite the bullet and live with the revisionary consequence that we cannot claim to know arithmetical truths? Secondly, can we hope to find a non-sceptical solution, i.e. a way to apply the framework in such a way that mathematical axioms can be claimed to be known after all? Chapter 6 provides a positive answer to the first question, and chapter 7 provides a positive answer to the second question.

Chapter 6 It is not obvious that it is devastating to our cognitive lives if it turns out that we only possess entitlements for important basic propositions. Wright wholeheartedly accepts such limits in the external world case. However, it seems that, in the mathematical case, entitlement at the basic level entails that all of mathematics turns out to be merely entitled.

I will argue that this result is not devastating to our epistemology overall. The idea is that we could still fruitfully apply mathematics and logic in other cognitive projects, such as the sciences. I will discuss three versions of the fallback position, and argue that even the most concessive one is a viable and interesting position.

Chapter 7 In chapter 7, I embed the neo-Fregean thesis about implicit definition in the framework discussed in *Part II*, in such a way that we can avoid the consequence that mathematical axioms are entitlements and cannot be claimed to be known. As mentioned above, the proposal promises to kill two birds with one stone, because it also provides an explication of the central epistemic workings of the idea that apriori knowledge can be obtained on the basis of implicit definitions which has been requested in *Part I*.

The basic idea is that implicit definition can be associated with a basic belief-forming method with entitled presuppositions, whereas the presuppositions are just the relevant preconditions of definitional success. I explain the account by sketching its application to the arithmetical case. I investigate the presuppositions of definitional success for the case of Hume's Principle, and shed light on the three objections to neo-Fregeanism discussed in chapter 2.

I then sketch, in broad brushstrokes, how the account might be applied to the logical case, and thereby provide the outlines of a unified epistemology for mathematics and logic that I imposed as a desideratum at the beginning of the thesis. Moreover, I argue that my account reveals a flaw in Boghossian's notion of *epistemic analyticity*, and I shall suggest a new explication of an epistemic notion of analyticity.

Part I

Old solutions to old problems

1 Epistemology of mathematics: issues and options

1.1 Puzzles about arithmetical knowledge

We know many mathematical truths, and we know a lot about arithmetic in particular. Every educated person knows that there are infinitely many prime numbers, and every child knows that 1+1=2. Moreover, we know many things about the world which we describe with the help of mathematical concepts. Do not most of us know that the number of planets is 8, and that many believe falsely that the number of planets is 9?

Mathematical knowledge matters to us: there is a vast number of Mathematics departments around the world — dedicated to extending mathematical knowledge. Without knowing a great deal of mathematics, contemporary physics would be impossible. Without mathematical knowledge, there would never have been space shuttles.

Mathematical knowledge is good for many things. So much the worse that some straightforward philosophical reflection makes mathematical knowledge look rather puzzling. We just need to ask the following questions: how exactly are mathematical beliefs justified? How exactly do we know mathematical truths? And how exactly can we rationally claim this knowledge?

I first show how these questions lead to philosophical puzzles, and then discuss some classical responses to them. The questions are hard enough for the case of arithmetic. I mostly stick to this particular case, although I say a little bit more about other cases along the way.

So why are these questions about arithmetical knowledge so hard to answer? After all, for many arithmetical beliefs — or other beliefs involving arithmetical terms — the questions of how exactly they are justified seem to admit of straightforward answers. However, there are at least two strategies to create puzzles about arithmetical knowledge. The first is to ask the "How?" question one time too many. The second is to look at what exactly arithmetical statements say. I discuss both strategies in turn.[1]

[1]Since this thesis is also concerned with basic logical knowledge, it is worth mentioning that at least the first puzzle also arises for logical knowledge. It is not clear whether the second puzzle arises as well (Field 2005, section 6).

1.1.1 Puzzle 1: how are the axioms justified?

Consider the question of how we can justify that the numbers of knives on a table is equal to the number of planets (assume there are 8 knives on the table). The answer seems simple: we know that the number of planets is 8, so if we *count* 8 knives, there is not much more to say.

Another example is the proposition that there are infinitely many primes. How can we justify it? Again, the answer is straightforward: by a proof, of course. For any prime number p, the product of all prime numbers "$\leq p$" plus 1 is prime, and greater than p. So we always find a greater prime number. q.e.d.

However, there will be some mathematical facts for which it is far less obvious how we are to justify them. And we do not have to consider complicated number-theoretic conjectures. Interestingly, an answer to our question gets harder for very simple mathematical truths. Here is an example: how can we justify the belief that 1+1=2? Whereas most non-experts lack a good answer (or simply say "this is obvious" or "I learned that"), mathematicians might refer to the Dedekind-Peano axioms (henceforth: Peano axioms). In Peano Arithmetic (PA), the deductive closure of these axioms, the equivalent of "$1+1=2$" in the language of PA ("$S0 + S0 = SS0$") is a theorem. We can thus still obtain an *inferential* justification from more basic principles.

However, we philosophers can easily puzzle the experts as well. We simply ask the "how" question one time too many. We ask: how are the Peano axioms justified? At this point, a further appeal to a proof from more basic principles is useless. After all, the Peano axioms are *axioms*. Mathematicians regard them as mathematical *basic beliefs* — fundaments of the tower of canonical mathematical knowledge. In actual mathematical practice, the axioms are just introduced and accepted. So are we just *assuming* the axioms? It would be worrying if we could not say more: for we can assume anything. Obviously, mere assumptions cannot amount to knowledge.

1.1.2 Puzzle 2: incompatible constraints

Benacerraf's challenge The second puzzle is a notorious challenge by Paul Benacerraf (1973). Benacerraf argues that it is hard to combine a reasonable semantics for arithmetical statements with a reasonable epistemological background picture. In other words, Benacerraf imposes a condition on a successful philosophy of mathematics which appears (almost) trivial, and shows that it faces a dilemma. The condition is:

(Benacerraf's Triviality) A successful philosophy of arithmetic has to combine a reasonable semantics for arithmetic with a reasonable epistemological theory in such a way that arithmetic knowledge is possible.

According to Benacerraf, this generates a dilemma. On one horn of the dilemma, we assume the most reasonable semantics for arithmetical statements. Then it is hard to see how mathematical knowledge is possible. On the other horn of the dilemma, we assume an epistemological theory that makes it easy to see how arithmetical knowledge is possible: then it does not give the right picture of what mathematical statements mean or *really* are about. Thus, we either have to give up our most reasonable semantics, or we have to give up our most reasonable epistemology.

In order to render his dilemma more concrete, Benacerraf uses the prevalent assumptions of his time. This yields the following tenets: the semantics for arithmetic must be based on Tarski's theory of truth (Tarski 1935), and the most plausible epistemological background picture is a *causal theory of knowledge*. Now consider the sentence:

(NOP1) There are 8 planets in the solar system.

It seems to be equivalent to:

(NOP2) The number of planets in the solar system is (identical to the number) 8.

The second sentence expresses an identity. In particular, "The number of planets" and "8" are singular terms. Thus, according to Tarski's theory of truth, the sentence can only be true if and only if both terms refer to the same object. The general point to note here is of course that arithmetic is about arithmetical *objects* — the (natural) numbers. Arithmetical statements are true if and only if they express facts about these objects.

But what kinds of objects are numbers? They certainly aren't objects like tables and chairs, located in space and time. We cannot *see* the number 7, and it does not make sense to suppose that it ceases to exist in 2000 years. It is easy to arrive at the picture of a distinct Platonic realm of mind-independent abstract objects.

However, this picture is hard to reconcile with the causal theory of knowledge. Assume that arithmetic is indeed the science of natural numbers, and the numbers are abstract objects in a Platonic realm. Then how could we ever be in causal contact with them? But this is just what the causal theory of knowledge demands. According to the causal theory of knowledge, we possess knowledge of some fact F if and only if our belief that F stands in the appropriate causal relationship to the fact that F. For example, so the thought goes, we can obtain knowledge of the existence of a barn in front of us if there is in fact a barn in front of us reflecting some of the incoming sunlight, which in turn causes an image on our retina which finally causes an experience and the forming of the appropriate belief. No such causal story seems to be available in the arithmetical case.[2]

To sum up: we either have to give up Tarski's semantics for arithmetic or the causal theory of knowledge. Now, of course this is only a dilemma if we want to uphold both Tarski's semantics and a causal theory of knowledge. And although the first assumption can be defended, the second is contentious, to put it mildly.

How can we defend Benacerraf's first assumption? First, we note that Tarskian semantics is suitable for other areas of discourse. For example, when we interpret a discourse about tables and chairs, it makes sense to say that singular terms refer to objects and sentences expressing an identity are true if and only if the terms on both side of the identity refer to the same object. Second, we note that it at least appears as if number terms are singular terms: we assign numbers to concepts,[3] we talk about numbers being identical, etc. Now, the following principle is very plausible (Benacerraf 1973, p.670):

(Uniformity Principle) We should aim at a uniform semantic theory for all our factual discourses.

And this enables us to infer that we should interpret number terms as referring to objects, since mathematical discourse is clearly factual.

As I said, Benacerraf's second assumption is not so easily defended. A reason in favour of the causal theory of knowledge is that it provides a response to Gettier cases (see e.g. Goldman 1967). However, the causal theory is not without problems, and alternatives have been suggested, such as reliabilism (Goldman 1979) and sensitivity-based accounts (Nozick

[2]Note that the problem is only an instance of a general problem with knowledge of abstract objects. A similar puzzle could be created for properties, musical works, etc.

[3]This observation is, of course, due to Frege. It famously appears in *Grundlagen* §46.

1988). Such accounts do not generate the same problems for knowledge of abstracta as the simple causal theory does, at least not obviously so (Hale 1994*a*, p. 170).

Field's challenge However, the general idea behind the challenge remains untouched by the fact that the simple causal theory does not withstand critical scrutiny. According to Field (2005), the real challenge is not about a particular epistemological theory. It is about explaining how our beliefs can reliably match mathematical truth. For without being able to explain how this can happen, we cannot rationally claim any mathematical knowledge.

The thought is that for any subject matter M we want to claim knowledge of, and whatever our conception of knowledge is exactly — the following condition holds:

> **(Field's Constraint)** We cannot rationally claim knowledge about M if it is impossible to explain how we can reliably form M-beliefs, i.e. how it can come about that many (or most) of our beliefs about M correspond to the M-facts.

The immediate problem is that it is very hard to see how such an explanation might look, if we assume the above-described semantic picture. Field (1989) presents a dilemma which emerges by dividing up the space of possible explanations into causal and non-causal explanations. The explanation cannot be causal, so the thought goes, because the Platonist is forced to hold that numbers are not in causal contact with anything. And the explanation cannot be non-causal either because, given the mind and language-independence of mathematical entities, "it is very hard to see what this supposed non-causal explanation could be" (Field 1989, p. 231).

Field does not offer an argument for this claim. Divers and Miller (1999) interpret Field as just seeing no option for someone who holds that mathematics is mind-independent in the sense that the existence of mathematical facts does not depend on the existence of minds. One option Divers and Miller consider and dismiss on behalf of Field is that the truths of mathematics are constituted by mental states. Although this would render the explanation of reliable mathematical belief-formation a lot easier, it contradicts our picture of mind-independent mathematical facts, for whatever is constituted by mental states is mind-dependent.

If Field is correct, then, since both the causal and the non-causal route to an explanation of the reliability of our mathematical beliefs are closed, and these alternatives are exhaustive, there is no explanation of reliable mathematical belief-formation and arithmetic violates **(Field's Constraint)**. In this case, we should not (or cannot) claim any arithmetical knowledge.

1.2 Desiderata for a solution

To put more carefully the upshot of the discussion of **(Field's Constraint)**, it is that we should not (or cannot) claim any arithmetical knowledge as long as we uphold a Tarskian semantics, and a picture of mind-independent mathematical facts and numbers as mind-independent abstract objects. Furthermore, it is necessary to develop this thought in detail and to define some of the key terms involved.

1.2.1 Arithmetical Platonism

I have already explained how, given some plausible background assumptions, the problematic semantic picture naturally arrives from a reflection on arithmetical discourse. The following three claims are very plausible:

(Minimal Arithmetical Realism)

1. The surface grammar of arithmetical propositions has to be *taken at face value*. In particular, number terms are singular terms — terms whose semantic role is to refer to objects. So, for example, the sentence "1+1=2" expresses a genuine identity, namely that the object denoted by "2", and the value of the function denoted by "+", applied to the objects denoted by "1" and "1", are in fact one and the same object.
2. Arithmetical discourse is *factual*. For example, people have genuine disputes about whether there are numbers with certain features. Arithmetical discourse is about what the *facts* about the arithmetical objects (the numbers) are.
3. Arithmetic is a body of *truths*. Given 2, this entails that its distinctive objects — the numbers — (really) *exist*.

With *surface grammar*, I mean the logical structure that is suggested by paraphrasing the sentence literally, i.e. the logical structure of the most straightforward paraphrase of the sentence to standard (second-order) logic. The surface grammar of a sentence can be contrasted with its *real logical form*, which is the logical structure of the best paraphrase of the sentence to standard (second-order) logic — roughly, the paraphrase which captures the meaning of the sentence in an important sense, which captures its commitments, and so on. To say that the surface grammar of a sentence S has to be taken at face value is to say that the surface grammar of S and the real logical form of S are identical.[4]

To repeat: the first claim can be motivated by looking at actual mathematical practice and Benacerraf's **(Uniformity Principle)**. We talk as if numbers exist. For example, when mathematicians *find* a new large prime number, they claim nothing short of existence. Number terms seem to meet the syntactic criteria for singular termhood (for an analysis of such criteria supporting this claim, see Hale 1994*b*, 1996). By the **(Uniformity Principle)**, we should treat singular terms in mathematics just as in ordinary external world discourse, i.e. as terms whose semantic role is to refer to objects.

For the purposes of this project, I will assume that there are singular terms, terms whose semantic role is to refer to objects. I will also assume epistemic transparency in the sense that someone possessing the relevant mathematical and philosophical concepts and skills — in a sense that includes the reader of this thesis — is able to justify claims about the real logical form of relevant sentences apriori, in a sense of apriority that does not rule out empirical defeasibility. I cannot rule out here that linguistic investigations reveals that ordinary number terms should not be treated as singular terms — this is certainly an epistemic possibility — but I assume that this is false. I briefly come back to these assumptions in 2.2.2, 2.3.5, and 7.1.8.

The second claim — the factuality of arithmetical discourse — is easily motivated by observing that we take arithmetic (very) seriously. In particular, we neither treat arithmetical discourse relativistically nor does it seem as if we treat arithmetical discourse fictionally. It is not a matter of taste whether three is a prime number. And saying that three is a prime number does not seem to be like saying that Sherlock Holmes is a detective.

The first part of the third claim — that arithmetic is a body of truths — is a piece of *common knowledge*: of course, arithmetic is true. It can be motivated further by noting that it is hard to see how it can be so usefully applied if it is untrue. Of course, this does

[4]Sometimes, the surface grammar of a sentence is different from the logical form of the best paraphrase of the sentence. For example, one might think that "It is raining" is not best translated in the same way as "It is a blue car".

not mean that we do not have false arithmetical beliefs from time to time. However, at least the experts' beliefs are largely correct.

A stronger claim than **(Minimal Arithmetical Realism)**, but still a popular claim, is the following:

> **(Arithmetical Realism)** **(Minimal Arithmetical Realism)** holds, and the objects of arithmetic — the natural numbers — are *mind- and language-independent.*

That the natural numbers are neither mind- nor language-dependent is to mean that their existence does not counterfactually depend on the existence of minds or languages.[5] This claim can be motivated by noting that the following counterfactual seems to be true: if there were no minds or languages, it would still be the case that $1+1=2$. It is hard to see how the minimal realist can avoid **(Arithmetical Realism)**.

Now, **(Arithmetical Realism)** in turn leads to the following, even stronger position:

> **(Arithmetical Platonism)** **(Arithmetical Realism)** holds, and the numbers are (pure) *abstract objects*, i.e. they do not exist in space and time (and we cannot be spatiotemporally related to them).

Here is one way to motivate the claim that numbers do not exist in space and time. It seems clear that they are not physical objects like tables and chairs, objects that can be seen, measured, or destroyed. If they were located in space and time, the question of where the number 7 is would be meaningful, where "where" is not deviantly interpreted as in "Where is the number 7?" — "Well, in-between 6 and 8, of course".[6] Moreover, the question of when the number 7 exists would be meaningful. However, both questions sound absurd.

Henceforth, I will also say "numbers" instead of "natural numbers" and "Platonism" instead of "Arithmetical Platonism". Someone rejecting either **(Minimal Arithmetical Realism)** or **(Arithmetical Realism)** will be called an *anti-realist about arithmetic.* Someone rejecting **(Arithmetical Platonism)** will be called a *nominalist.*

(Arithmetical Platonism) is the position that Benacerraf and Field assume to be the standard picture of arithmetic. In order to generate problems as above, we just need

[5] I thus follow Divers and Miller (1999).

[6] Note that if numbers were mind-*dependent* objects, then "in our heads" might be an eligible answer.

to impose it as a constraint for our epistemology of arithmetic that it has to work together with **(Arithmetical Platonism)**. Let me note, however, that the best version of Field's challenge is one that does not straightforwardly assume **(Arithmetical Platonism)**, but which just assumes that it is a plausible position whose denial needs motivation and explanation. This opens up a lot of new space for manoeuvre. Suppose that we just cannot find any way to reconcile our epistemological constraints with **(Arithmetical Platonism)**. Then one might be able to find an epistemology for arithmetic which — although it does not work for full-blown **(Arithmetical Platonism)** — at least works for **(Minimal Arithmetical Realism)** or **(Arithmetical Realism)**. A fortiori, if it really can be made plausible that we are even mistaken in assuming **(Minimal Arithmetical Realism)**, we are entitled to give up even this basic semantical constraint. If, on the other hand, the position giving up any of the principles above has problems of its own, this adds up to the initial implausibility of denying one of the plausible semantic constraints.

Simply put: I believe that in the philosophy of mathematics, we ultimately have to carry out a cost-benefit analysis. In this thesis, I argue that a cost-benefit analysis reveals that we do not have to give up **(Arithmetical Platonism)**. In particular, I argue that nominalist positions have problems on their own, and that there is a plausible account of knowledge of arithmetic, Platonistically construed.

1.2.2 Reconstructing arithmetical knowledge

We are left with two challenges for **(Arithmetical Platonism)**:

- We need to respond to Field's worry, i.e. we need to account for the possibility of reliably forming beliefs about the arithmetical realm.
- We need to account for a way to justify (obtain knowledge of) arithmetical axioms.

These formulations are highly ambiguous. Firstly, it is not clear which notions of justification and knowledge are presupposed. For example: is it the ordinary, everyday notion of knowledge, or some precisified, philosophical one? I leave this question open here, because I do not think that much hangs on it.[7] Secondly, the challenges can be understood as a challenge about *actual* knowledge. However, one might also have the more modest aim to

[7] I expand on this issue in chapters 3 and 4.

explain how it is *possible* to reliably form mathematical beliefs, and how it is *possible* to justify axioms, remaining silent about the actual situation.

If we understood the challenges as being about actual mathematical beliefs, and actual mathematical knowledge, then we would surely have to use an ordinary notion of knowledge, and conceive of our two challenges as challenges to the following claim:

> **(Actual Knowledge)** Most of our arithmetical beliefs are items of knowledge.

However, the use of "our" here is suboptimal. In fact, we can focus on mathematicians without loss of generality. For if the mathematicians have a lot of arithmetical knowledge, knowledge of the non-mathematicians could be explained, among other things, as acquired by testimony. We thus obtain the following, more precise claim:

> **(Actual Knowledge')** Most of the mathematicians' arithmetical beliefs are known by them.

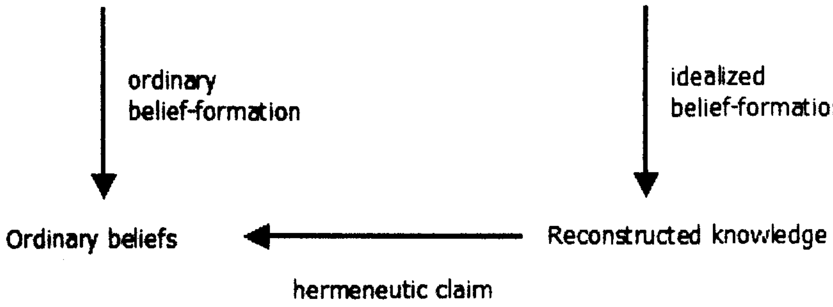
In this thesis, I shall not attempt to establish **(Actual Knowledge)** and **(Actual Knowledge')**. Rather, my primary aim is just to establish the following:

> **(Possible Knowledge)** *There is* a route to acquiring knowledge of the arithmetical facts.

Arithmetical facts should be understood as the facts about the numbers that fully competent users of number talk and mathematicians talk about. They are expressed by ordinary — and, when it comes to the mathematicians, sometimes technical — statements about numbers, which can be found in mathematical, scientific, and everyday discourse of sufficiently competent speakers. I call the theory consisting of the true ordinary statements about numbers the *ordinary theory* (or *ordinary arithmetic*), and I call the number terms of ordinary arithmetic *ordinary number terms.*

The phrase "there is a route" is ambiguous. It is important that it shall be understood in a sense which is not too demanding. I shall argue that every non-defective epistemic agent, in a sense of "non-defective epistemic agent" which includes the readers of this thesis, is able to acquire knowledge of the truths of ordinary arithmetic.

In particular, I shall engage in a *hermeneutic reconstructive epistemological project.* A hermeneutic reconstructive epistemological project about a region of thought X is a project of explaining a canonical process by which a non-defective epistemic agent can vindicate — from scratch — all relevant X-knowledge. It is *reconstructive* because it is not required that the described process has anything to do with *actual* belief-formation. It is *hermeneutic* because the reconstructed knowledge is knowledge of the same subject matter.[8] We can visualize such projects as follows:



One can conceive of the project as involving a two-step process. The first step consists in construing a route of how our epistemic agent can come to acquire knowledge of some artificial theory, suitable for the reconstructionist's purposes. The second step consists in arguing for a hermeneutic claim, i.e. that the terms and statements of the artifical theory have the same meaning as the terms of the discourse to be reconstructed, and the ordinary theory in particular. Ideally, it will also be possible to argue that the hermeneutic claim is available to Hero as well, but note that this is not part of the hermeneutic claim. Given the above assumption that the fact that the surface grammar of ordinary arithmetical statements has to be taken as face value is available apriori to sufficiently competent agents (see 1.2.1), this claim can be argued for. I will briefly discuss this further in 2.3.5 and 7.1.8.

There is an ambiguity as to what exactly the hermeneutic claim involves. Using the Fregean distinction between sense and reference, we can distinguish the following two claims:

(Weak Hermeneutic Reconstruction) The terms[9] of the reconstructed theories have the same *referents* as the corresponding terms of the ordinary theories.

And:

[8]This terminology is inspired by a similar terminology endorsed by Burgess and Rosen (1997).

[9]Note that this claim is not restricted to singular terms. I mean all terms, including predicates. I assume here that predicates refer to properties.

(Strong Hermeneutic Reconstruction) The terms of the reconstructed theories have the same sense / express the same concepts as the corresponding terms of the ordinary theories, and corresponding statements express the same thoughts.

The weak hermeneutic claim already entails that the reconstructed theories have the same subject matter, and that the statements of the ordinary theory are true just in case corresponding statements of the reconstructed theory are true. Note that, if arithmetic is about necessarily existing abstract objects, the weak hermeneutic claim will already entail that it is necessarily the case that the propositions expressed by the statements of the ordinary theory are true just in case the propositions expressed by the corresponding statements of the reconstructed theory are true. The strong hermeneutic claim entails this in any case.

There will be a variety of strong hermeneutic projects, if there is a variety of ordinary concepts of number. There might be different concepts of number possessed by subjects with different levels of sophistication (mathematical, conceptual, and otherwise).[10] The question then arises of which of the concepts I am concerned with. If there are different concepts of number, then I am most interested in the ordinary concept of number that is possessed by someone who is a fully competent user of ordinary number talk (in a sense that includes mathematicians[11], scientists, and the reader of this thesis). For the purposes of this project, I assume that there is only one such concept.

As Burgess and Rosen (1997) point out, *hermeneutic reconstructions* should be distinguished from what we may call *revolutionary reconstructions,* i.e. reconstructions that provide an epistemically kosher replacement for the old theory. Revolutionary reconstructions replace the hermeneutic claim by a weaker claim to the effect that the missing hermeneutic link is harmless. One plausible candidate for a weaker claim is that we can replace the ordinary theory by the reconstructed theory without (explanatory) loss because the reconstructed theory can play the same role in the sciences and elsewhere:

(Revolutionary Reconstruction) The reconstructed *entities* can be used to replace the ordinary entities without (explanatory) loss, and the reconstructed *theories* can be

[10] I am indebted to Andrew McGonigal for raising this issue.

[11] Mathematicians might possess different concepts of number. I always mean the concept that they employ in ordinary number talk, and not different artificially introduced concepts they might use in the math classroom. Moreover, I am of course only concerned with cardinal numbers here, not with other types of numbers such as real numbers.

used to replace the ordinary theories without (explanatory) loss.

Arguing for hermeneutic claims requires substantial further work. In the arithmetical case, it is not even clear what exactly the ordinary theory is. For example: is it a pure arithmetical theory, such as PA, or is it a theory implicit in our everyday arithmetical practice (or both)? I will return to this issue in 2.3.5.

In this thesis, I will argue mainly for **(Weak Hermeneutic Reconstruction)**, but I take it to be a live option that my proposal also accounts for **(Strong Hermeneutic Reconstruction)**. I conceive of **(Revolutionary Reconstruction)** as a fallback option.

Thus, regarding arithmetic, the claim to be argued for is the following:

> **(Arithmetical Knowledge)** The readers of this thesis can acquire knowledge of a theory which is a weak hermeneutic reconstruction of ordinary arithmetic (the theory containing the true ordinary arithmetical statements).

It is desirable to argue for **(Arithmetical Knowledge)** because, among other things, it enables us to make claims like the following:

- The mathematicians, the scientists, and sufficiently competent users of mathematics in everyday discourse have true arithmetical beliefs. This is because the truth values of relevant ordinary arithmetical propositions (including "1+1=2", but also suitable versions of the Peano axioms such as "Every number has a successor") are the same as the truth values of corresponding theorems of the reconstructed theory, and we know that the reconstructed theory is true.

- Given that we can show that the reconstructed theories are about mind-independend abstract objects, the arithmetical objects the users of arithmetic seem to fiddle around with uncritically all the time turn out to really exist as mind-independent abstract objects, and theorems of arithmetic mathematicians make every effort to prove turn out to be truths about these objects. For it can be shown that these objects are the same as the objects our reconstructed theory is about, and we know that the reconstructed theory is true.

This is a huge advantage, psychologically as well as philosophically. If we can only argue for **(Revolutionary Reconstruction)**, we cannot justifiably make these assertions. The status of the propositions of the old theories, and the status of the beliefs of those who do not use the new theories remain open.

1.2.3 Additional desiderata for the reconstructive project

I now carve out some additional desiderata for our reconstructive project.

Additional epistemological desiderata I will eventually argue that we need an internalistic notion of knowledge, i.e. a notion of knowledge whose possession is reflectively available by the subject. Details will be provided in chapter 3. What is more important for my current purposes is that arithmetical knowledge is apriori, and our epistemology should account for this fact.

In short, a belief is justified apriori if the justification for the belief is not essentially based on experience, and apriori knowledge is knowledge based on an apriori justified belief. A belief that is justified, but not justified apriori, is justified aposteriori. Aposteriori knowledge is knowledge based on aposteriori justified belief.

The phrase "essentially based" allows for apriori knowledge in cases where acquiring or possessing relevant concepts requires experience. For example: that nothing is red and green all over should count as knowable apriori, although acquiring the concepts red and green requires experiencing red and green things. Moreover, it allows for the generation of apriori knowledge by proofs which are carried out using pen and paper.

We obtain the following constraint:

> **(Apriority Constraint)** There is a way to acquire apriori knowledge of arithmetical facts.[12]

A further epistemological desideratum does not concern the epistemic status of arithmetic in general, but of a special class of arithmetical statements: the Peano axioms. In 1.1.1, we saw that it is especially puzzling how we are justified in believing these axioms. If we account for their knowability, we will have made a big step towards solving the problem of canonical arithmetical knowledge, which is based on axioms and proofs. It is worth making the special role of axioms explicit:

[12] A further distinction can be made here between *weak* and *strong* apriori knowledge (see e.g. Field 2005). I come back to this distinction in a footnote in 7.1.3.

(Arithmetical Foundationalism) We must be able to explain how we can know arithmetical axioms in particular, and how all other arithmetical knowledge can rest on this knowledge.

Applicability A desideratum of a different type emerges from the fact that mathematics has an important role to play in the sciences, and indeed also in ordinary life. In the sciences, it serves at least three purposes (Bueno & Colyvan 2011, section 5.4):

- Explanation: mathematics is used in scientific explanations. We do not need to restrict this claim to physics. For example: the prime lifecycles of cicadas can be explained by the fact that prime lifecycles minimize intersections (Baker 2005).
- Prediction: it is obvious that mathematics plays an important role in scientific predictions. As I said at the beginning of this chapter: without mathematics, there would not have been space shuttles.
- Unification: new mathematical theories can help with unifying scientific theories.(Bueno & Colyvan 2011, p. 351; Colyvan 2002).

Every satisfactory epistemology of mathematics needs to say something with respect to the question of how mathematics (and mathematical knowledge) can serve these roles. However, in this thesis, I cannot focus on all these roles in detail. What I take to be central to all these cases, is that mathematics helps with the extension of knowledge in other areas (I will come back to this issue in chapter 6). This general possibility every satisfactory epistemology of mathematics needs to explain. This yields our final constraint:

(Applicability Constraint) We must explain how our reconstructed arithmetical knowledge can be used to extend knowledge in other areas of cognitive enquiry, e.g. the sciences.

This almost completes my discussion of desiderata for a satisfying epistemology of arithmetic. Before I look at some Platonist positions in the philosophy of mathematics and examine to which extent they meet these desiderata, a note is in order regarding the role of logic in a reconstructive project for mathematics.

Two notes on logic Mathematics cannot be done without logic. Proof is a canonical belief-forming method in all of mathematics, and proofs are logical deductions from mathematical axioms. Any epistemology of mathematics rests on the premise that logical reasoning extends knowledge, in such a way that some special epistemic status is preserved, such as apriority. This is often only tacitly assumed, and it is worth making it explicit:

> **(Extension of apriori knowledge)** Ceteris paribus, logical reasoning can be used to extend apriori knowledge.

A second (optional) desideratum arises from a comparison between logic and mathematics. Both logic and mathematics are apriori, both are regarded as especially certain, and both serve the extension of knowledge in a wide range of cognitive projects. This suggests they have something important in common. One plausible idea — for which I will argue in the last chapter of this thesis — is the following:

> **(Same Source)** Our knowledge of mathematics and our knowledge of logic rest on the same epistemological source.

I take this claim to be clear enough for my purposes here, but I will precisify it later. Frege famously defended it. It was only after he discovered that his programme in *Grundlagen* was destined to failure that he gave up the claim that both mathematical and logical knowledge rest on what he called a *logical source*, and that he conjectured that the former rests on both logic and what he called the *geometrical source* of knowledge (Frege 1979*c*,*b*).

Note, however, that the **(Same Source)** desideratum does not amount to assuming that mathematics is nothing else than logic. Rational intuition, for example, would also be a source delivering logical as well as mathematical knowledge.

In any case, the **(Same Source)** desideratum is optional. We cannot use it to rule out an otherwise successful epistemology of mathematics.

1.3 Unsatisfying approaches

I first discuss three approaches to arithmetical knowledge which preserve **(Arithmetical Platonism)**, and show that they face strong objections. I then (very) briefly look at nominalistic positions. It turns out that these positions also confront huge problems, so there is no epistemological reason to give up **(Arithmetical Platonism)**.

1.3.1 Gödelian Platonism

Gödel, in a supplement to his 1947 paper on the continuum hypothesis (Gödel 1964), endorses a generalization of **(Arithmetical Platonism)**, and notoriously combines it with the idea that we possess a faculty of *mathematical intuition*. Gödel claims: (i) that there are reliable belief-forming methods associated with this faculty; (ii) that these can deliver apriori justified true beliefs about mathematical axioms, and (iii) these amount to knowledge of mathematical axioms. Gödel's writing even suggests that the workings of the relevant belief-forming processes are similar to those of sense perception. For example, Gödel speaks of mathematicians *perceiving* mathematical objects (Gödel 1964, p. 268). This suggests that Gödel believes that we can be in some kind of causal contact with mathematical objects.

Unfortunately, it remains entirely unclear how exactly all of this is to work. Gödel never develops his theory of mathematical intuition any further. However, Bonjour (1998) develops a general theory of *rational intuition* which seems to be applicable here. Except for Bonjour's claim that one should not appeal to a special cognitive *faculty* of rational intuition (Bonjour 1998, p. 109), and that he rejects the claim that one can be in any direct causal contact with abstract objects (Bonjour 1998, section 6.2, pp. 159f), Bonjour's proposal seems to be compatible with Gödel's. Since Gödel's remarks are too thin to draw on, I will discuss an application of Bonjour's proposal to the mathematical case.

Bonjour argues that all apriori knowledge must ultimately rest on primitive *rational insights* (or *rational seemings*) to the effect that something is (necessarily) true (Bonjour 1998, section 4.3, p.106). These insights can confer apriori justification upon their target beliefs (Bonjour 1998, p.107). However, the relevant belief-forming method is fallible: it is possible that something clearly seems (necessarily) true to us, but turns out to be (necessarily) false in the end (Bonjour 1998, 4.4). It might help to conceive of rational seemings as an apriori analogue of perceptual seemings. Hallucinations and optical illusions can occur, but clearly they are the exception rather than the rule.

As to the application to the mathematical case, Bonjour discusses the proposition that 2+3=5. According to Bonjour, everyone understanding and entertaining this proposition will be confronted with a rational seeming of its (necessary) truth (Bonjour 1998, p.104). It is worth noting that Gödel's proposal is first and foremost meant to apply to mathematical axioms (Parsons 1995, pp. 59f). However, there is a very natural way in which Bonjour's proposal accomodates this. For example, it seems natural for Bonjour to say that everyone who understands what natural numbers are, and understands the proposition that every natural number has a successor, will be confonted with a rational seeming of its truth.

Of course, this is not to say that everyone has the same rational insights. The experts will have much more such seemings, and much more adequate ones, than the layman. Moreover, it seems that even the experts will not have clear seemings regarding every mathematical truth. Often, they will need to help themselves with constructing proofs, based on axioms or rules that are clear cases of rational insight.

Is such a position a candidate for a satisfactory solution to our epistemological problem? It is surely initially attractive, because it meets **(Arithmetical Platonism)**, **(Arithmetical Foundationalism)** and the **(Arithmetical Knowledge)** by design. Since the faculty of mathematical intuition could also provide knowledge of axioms, we would also meet **(Arithmetical Foundationalism)**.

The problem is that it is not clear *how* such knowledge is possible, because it is not clear how precisely the belief-forming method of rational intuition is supposed to work. Where do rational insights come from? Why do they carry any epistemic force? The position is quite spurious when it comes to its central epistemological workings.

For example, one might worry whether the position is really compatible with **(Arithmetical Platonism)**: is it possible to explain rational intuition without postulating some kind of causal contact with mathematical objects?

This points towards a more general problem: rational intuition might be so spurious that it cannot count as a proper *explanation* of how our mathematical beliefs can track mathematical truth. Thus, **(Field's Constraint)** might kick in and we might be unable to claim arithmetical knowledge after all. To be sure, the proposed faculty of mathematical intuition would certainly entail that we can be in touch with the mathematical facts. However, the question is whether this is effectively just taking it to be a brute fact that we can be in touch with mathematical reality (Field 1989, p. 28).

Moreover, and this is related to the unclarity of the postulated belief-forming method, more needs to be said about how the knowledge delivered is *apriori*. Whether it should count as apriori depends on two factors:

- The range of the term "experience" in the definition of apriority.
- How seriously we have to take the analogy between rational intuition and sense perception.

If rational intuition is a kind of perception, and this kind of perception counts as experience, then the postulated belief-forming method might render mathematical knowledge aposteriori, and the **(Apriority Constraint)** would be violated.

In sum, the Gödel-Bonjour position is replacing our initial puzzles with other puzzles, and no satisfactory account of its exact epistemological workings has been produced. I share this opinion with many philosophers of mathematics (Boghossian 2001, p.6; Wright 2004*a*, p. 156). We should look for alternatives.

1.3.2 Fregean logicism

In his seminal *Die Grundlagen der Arithmetik* (*Foundations of Arithmetic*, Frege 1884, henceforth *GL*), Frege argued that arithmetic — platonistically interpreted — can be based on logic and (explicit) definitions alone, as opposed to some form of intuition.[13] Whereas *GL* contains the philosophical groundwork of his programme, he carries out the logico-mathematical component of his programme in full detail and rigour in *Grundgesetze der Arithmetik* (*The Basic Laws of Arithmetic*, henceforth *GG*), using a further developed version of the logical system he introduced in his earlier work *Begriffsschrift*. The programme is well-known as *Frege's Logicism*.[14]

Frege's philosophical aim in *GL* is to provide epistemological and semantic foundations for arithmetic. After dismissing several rival positions, Frege argues that number talk is genuine object talk. According to Frege, numbers are objects belonging to concepts (*GL* §46). For example, the statement "There are three knives on the table" really expresses the claim that *the number three* — an abstract object — belongs to the concept of *being a knife on the table*.

Frege then asks how we can obtain apriori knowledge of the existence and the nature of these objects. His answer is that arithmetical truths are analytic in the following sense:

[13] In particular, Frege attacks Kant's epistemology of arithmetic, which also rests on some form of intuition.

[14] Often, Logicism is only identified with the epistemological claim that mathematical knowledge is logical knowledge. In this thesis, Logicism is understood as including **(Arithmetical Platonism)**.

(Frege Analyticity) A statement S is *Frege-analytic* if and only if it follows from definitions and logic.

Frege thus explicitly opposes Kant's view that arithmetic consists of *synthetic* apriori truths, which would imply that they must be justified with the help of some sort of intuition. What are the definitions underlying arithmetical knowledge? First and foremost: what is the definition of number? At the heart of Frege's first train of thought in *GL* lies his notorious *Context Principle*, which he lays down in the introduction *GL*, and which he repeats in *GL* §62 (see also Wright 1983, p.6):

(Context Principle) Never (...) ask for the meaning of a word in isolation, but only in the context of a proposition.

This enables Frege to argue that fixing the meaning of all sentential contexts including a hitherto undefined term suffices to fix the meaning of this term. Moreover, according to Frege, the meaning of a sentence can be fixed by fixing its truth conditions. Hence, so the thought goes, fixing the truth conditions for a sufficient range of sentences containing a hitherto undefined term suffices to fix the meaning of this term.[15]

Number terms are singular terms. Thus, in order to apply the Context Principle here, Frege needs to single out a suitable class of sentences whose truth values must be fixed in order to introduce new singular terms. He does that in *GL* §63, where he argues that we can introduce a new sort of object into discourse by laying down an *identity criterion* for this kind of object.

This is where (Fregean) abstraction principles enter the picture. *Abstraction principles* are universally quantified bi-conditionals of the following form, where $\vec{\alpha}$ and $\vec{\beta}$ are blocks of m different variables of the same order n, Σ is a term-forming operator (the "abstraction operator"), and $Eq\left(\vec{\alpha}, \vec{\beta}\right)$ is an equivalence relation between the $\vec{\alpha}$s and the $\vec{\beta}$s:

(AP) $\forall\vec{\alpha}\forall\vec{\beta}\left(\Sigma(\vec{\alpha}) = \Sigma\left(\vec{\beta}\right) \leftrightarrow Eq\left(\vec{\alpha}, \vec{\beta}\right)\right)$

Frege's idea is that abstraction principles might be laid down in order to fix the truth conditions of identity statements for new sorts of objects and thus provide a means to

[15] Of course, the whole issue becomes much more complicated with the distinction between "sense" and "reference" that Frege draws after having written *GL*. In particular, the question arises of whether to interpret the **(Context Principle)** as a claim on the level of sense or on the level of reference. I cannot go into these issues here.

refer to them and to acquire knowledge of them. Let me go through an example which Frege uses to demonstrate how this might work in *GL* §§64-66: the "*Direction Abstraction*". The idea is that directions are objects belonging to directed objects (e.g. lines), and that the directions of two directed objects are identical if and only if the directed objects in questions lie parallel. This yields the following first-order[16] abstraction principle:

(DA) $\forall x \forall y \, (Dir\,(F) = Dir\,(G) \leftrightarrow \forall x \forall y \, (Fx \| Gx))$

In *GL* §63, Frege suggests applying the same idea to the number case, claiming: "Once we possess a means to grasp and recognize a certain number by this procedure, we are allowed to assign it a proper name" (*GL* §62, own translation). Thus, according to Frege, what needs to be done is to find suitable identity criteria for numbers, where numbers are objects belonging to concepts.

Such a criterion is readily available, and can be uncovered by considering examples. How can we come to know that the number of forks on the table is identical to the number of knives on the table? One way is to observe that we can align each fork with exactly one knife. Why? Because this shows that the knives on the table and the forks on the table are *equinumerous*, i.e. there is a *bijection* between the knives on the table and the forks on the table.

The equinumerosity of two concepts is not only a sufficient, but also a necessary condition for the numbers belonging to these concepts to be identical. We thus obtain a criterion of identity: the numbers belonging to concepts are identical if and only if the respective concepts are equinumerous. Frege notes that this criterion has already been suggested by Hume in his "Treatise of Human Nature" (*GL* §62). We may call it *Hume's Principle*[17]. It can be formalized as the following second-order[18] abstraction principle:

(HP) $\forall F \forall G \, (\#\,(F) = \#\,(G) \leftrightarrow \exists R \, (Bij\,(R, F, G)))$

"$Bij\,(R, F, G)$" stands for "R is a bijection between the Fs and the Gs", which is expressible in second-order logic.[19] "#" is called the "number operator"; read "$\#F$" as "the number belonging to the concept F".

[16]The abstraction principle is first-order because the outer quantifiers are first-order. They range over (directed) *objects*.

[17]The name "Hume's Principle" has been introduced by George Boolos.

[18]The abstraction principle is second-order because the outer quantifiers are second-order. They range over concepts.

[19]Formally: $Bij\,(R, F, G) =_{def} \forall x \, (Fx \rightarrow \exists! y \, (Gy \wedge Rxy)) \wedge \forall y \, (Gy \rightarrow \exists! x \, (Fx \wedge Rxy))$

If **(HP)** is a good definition of the number operator, then it is easy to obtain the concept of (cardinal) number by the following explicit definition:

(Number) $\forall x\,(Number\,(x) \leftrightarrow \exists F\,(\#F = x))$

However, directly after suggesting **(DA)** and **(HP)** as definitions introducing new kinds of objects into discourse, Frege rejects them both, because he thinks such definitions suffer from a quite general problem.

For example, according to Frege, one cannot — by means of **(DA)** — "decide whether the direction of the Earth's axis is (identical to) England" (*GL*, §66, own translation). This gives rise to the infamous *Caesar problem.* Similarly, so the thought goes, **(HP)** will not enable us to decide whether the number of planets is identical to Julius Caesar.[20] To be sure, the question of whether the number of planets is identical to Caesar may seem absurd (or trivial), and certainly does not actually arise in any non-philosophical context, but the problem that we do not seem able to decide the question by means of **(HP)** suggests that abstraction principles alone cannot introduce new kinds of abstracts in a fully satisfactory way, since they fix the identity conditions only between abstracts of the same sort (e.g. when both sides have the form "$\#\phi$").

Frege's own formulation suggests that the Caesar problem is an epistemic problem. Abstraction principles, so the thought goes, must give us means to *decide* all questions that could arise with regard to the newly introduced objects, and this includes mixed identity statements. However, there is also a semantic reading: that abstraction principles such as **(HP)** do not suffice to determinately fix a concept of number. For example, so the thought goes, it has not been fixed whether Caesar is identical with some number or not. It has been argued that, in the light of what Frege says about the same problem in *Grundgesetze*, it must have been the semantic problem that Frege worried about most (Schirn 2003). However, I do not have to (and cannot) decide this issue here. I will come back to different versions of the Caesar problem when I discuss the analogous worry for *neo*-Fregeanism in 2.4.1.

For now, it suffices to note that Frege regarded the Caesar problem as a devastating objection to using **(HP)** as a definition of number. According to Frege, the moral of this

[20] In Grundlagen §56, Frege gives this example in a slightly different context, but it is now usually used in this context, to point out Frege's worries about abstraction principles.

problem is that one has to define "the number of Fs" *explicitly*, using the notion of extension (think of extensions as sets), and to derive **(HP)** from the definition. As a definition for "the number of Fs", Frege suggested "the extension of the concept 'equinumerous with the concept F"' (*GL* §68). With this definition, Frege must have thought to have solved the problem of *what kinds of things* numbers are (see *GL* §67). This also suggests that "definition" in Frege's definition of analyticity should be understood in a narrow sense. The Caesar problem can be regarded as an objection against a more liberal conception of definition.

The new definition presupposes extensions, which Frege has available in his logical system. The notion of extension is governed by two axioms, corresponding to both sides of the following abstraction principle — the famous infamous *Basic Law V*:

(BLV) $\forall F \forall G \left(\epsilon\left(F\right) = \epsilon\left(G\right) \leftrightarrow \forall x \forall y \left(Fx \leftrightarrow Gx\right)\right)$

(BLV) says that the extensions of two concepts are identical if and only if the respective concepts are co-extensional. Using the above-mentioned explicit definition of cardinal numbers, Frege was able to prove **(HP)**. Moreover, Frege was able to (explicitly) define the remaining arithmetical notions, and to prove the axioms of arithmetic in turn (for the basic idea, see Frege 1884, §§70-83). Interestingly, his derivation of the Peano axioms from **(HP)** does not make any use of **(BLV)**. The result that arithmetic is derivable from **(HP)** alone has been called *Frege's Theorem.*[21]

With cardinal numbers defined as extensions the Caesar problem does not arise for numbers, *assuming* that we have solved it for extensions. However, the question arises of why **(BLV)** is not subject to the same sort of criticism as **(HP)**. In other words: how does Frege avoid the Caesar problem for extensions? This is one of the most difficult questions for Frege exegesis. We can only speculate, but I think the best way to answer this question is this: **(BLV)** is *not* meant to be a definition. Hence, it is not subject to the same standards. Rather, **(BLV)** is meant to be a logical basic law which must only meet the standard of logical basic laws, i.e. it must be self-evident, absolutely general, and not capable of further proof. Moreover, Frege takes the objects **(BLV)** is about — the extensions — to be logical objects which are already well-understood. In particular,

[21] For a more precise formulation, see 2.1.

Frege assumes that we already know that Caesar is not a logical object and thus cannot be identical with any extension.

Be that as it may, the resulting logical system is too strong, for **(BLV)** is inconsistent in Frege's logical system. It leads to *Russell's paradox*, which Russell communicated to Frege in a famous letter in 1902.

One can conceive of Frege's logical system without **(BLV)** as second-order logic with unrestricted comprehension. The paradox arises by defining the *Russell predicate* as the property of "being an extension which does not have itself as a member". By the right-to-left direction of **(BLV)**, there is an extension of this property. Call it the *Russell set*. Using the left-to-right direction of **(BLV)**, we can show that the Russell set has itself as a member if and only if it does not. Contradiction!

Thus, assuming the logical background system is correct, **(BLV)** cannot possibly be true, and thus cannot serve as a logical axiom. Interestingly, Frege was never absolutely sure about the status of **(BLV)**, which he mentions in a letter to Russell (Frege 1980, letter XV/7). After the inconsistency had been discovered, he tried to fix his system, but did not manage to do so. Frege became convinced that his efforts to defend logicism had failed.

Many attempts have been made to properly understand just why Russell's Paradox arises. This was not because Frege's system was popular or widely accepted, but because Russell's paradox looms in our naïve conception of set.[22] Set theory began to evolve as the foundational theory for (most of) mathematics, so it was of extreme importance to properly understand the problem that Russell had discovered.

Russell himself analysed Frege's system and came up with his own version of Logicism. He saw the reason for the inconsistency in Frege's system as arising from the *vicious circularity* of the definition of the Russell set (Burgess 2005, p.36).[23]

In their *Principia Mathematica* (Russell & Whitehead 1910-1913), Russell and Whitehead tried to ban circularity and — as far as we know — succeeded in constructing a consistent system. However, some of the axioms of Principia look far less like logical ax-

[22]Many set theories that were modelled on the basis of the naïve view were subject to the paradox: this includes Russell's own preferred system at the time of his discovery (Burgess 2005, p.32).

[23]It is obviously circular in the sense that the definition of the set talks about the set itself. It is vicious because this leads to paradox. A similar vicious circularity arises for properties: define the Russell property as the property of all properties which do not fall under themselves. This leads to a similar paradox, which has been called the *heterological paradox*.

ioms than Frege's. In particular, in order to obtain something as strong as arithmetic, they had to make use of the so-called *axiom of infinity*, which says that there are infinitely many individuals. Surely, such an axiom cannot count as *logical*. Thus, Russell's and Whitehead's version of logicism cannot count as logicism proper. They did not succeed in putting arithmetic on logical foundations, but just found another axiomatization for a theory that interprets arithmetic. In particular, their system cannot be used to solve our epistemological problems: for the question of how to justify the axiom of infinity is just as difficult to answer as the question of how to justify the Peano axioms.

Still, logicism remained one of the main contenders in the philosophy of mathematics. This is due to the influence of the *Vienna Circle*, according to which theories such as arithmetic are analytic in the sense that they are true *in virtue of meaning alone*. However, and especially because of the mentioned notion of analyticity, logical positivism was commonly deemed to be mistaken after Quine's battery of objections to the notion (e.g. Quine 1951). By the middle of the 20th century, attention had turned away from logicism.

Until recently, it was common ground that we have to draw this negative conclusion about logicism, and about Frege's philosophy of mathematics in particular. Dummett, for example, wrote that "Frege, as a philosopher of mathematics, is indisputably archaic" (Dummett 1991, p. xx).

However, *if* Frege's system had not been inconsistent, and his axioms had been self-evident logical axioms, Frege's logicism would have met all of our constraints. Frege would have provided a route to apriori knowledge of numbers, based on definitions and self-evident logical axioms, whilst accepting **(Arithmetical Platonism)**.[24]

1.3.3 The indispensability argument

What else could justify the existence of numbers, and reveal their properties? There is a third option, which is based on much more general considerations in metaphysics and epistemology. It is the option of justifying platonistically interpreted arithmetic on the basis of its being indispensable to our best scientific theories. This route to knowledge of abstract objects has been famously defended by Quine (e.g. Quine 1953) and Putnam (e.g. Putnam 1979). Contemporary versions of the indispensability argument — which are

[24]The claim that basic logical laws are self-evident needs a sustained defence. This is another big gap in Frege's programme.

based on inference to the best explanation — are taken very seriously even by nominalists (see e.g. Field 1989).

I will just discuss one particular version of the argument, building up on Field's excellent discussions in (Field 1980, 1989). The argument uses two premises to establish an epistemological point: that we are justified in believing that some mathematical theory is true. It is analoguous to arguments for the truth of scientific theories about unobservable entities, such as quarks. The first premise is that there is a belief-forming method of *Inference to the Best Explanation* (IBE) which can be used for justifying whole theories, regardless of what these theories are about:

(IBE) For any theory T, if we use T in our best explanations of the data, we are justified in believing that T is true.

It is not easy to say what exactly should count as the data. For my purposes here, think of it as the sum of our experiences. The second premise is a claim about how mathematics features in our best scientific theories:

(Indispensability) Some mathematical theories are indispensable to our best scientific theories.[25]

Since our best scientific theories are used in our best explanations of the data, we are justified in believing that the mathematical theories which are indispensable to these scientific theories are true. Since we also assume that we should take the surface grammar of mathematical statements at face-value, and mathematical discourse is factual, we are justified in believing in the existence of the referents of the mathematical terms used. The argument is meant to be analogous to arguments for the existence of unobservable entities. Consider quarks, for example. Postulating their existence might be indispensable to our best physical theories. Thus, we can believe in their existence, although they are *unobservable* in an important sense of term.

Obvious candidates for indispensable mathematical theories are arithmetic and real analysis. Quine thought that at least some set theory is indispensable (Quine 1953).

[25]One might think that the mere fact that mathematical theories *feature* in our *best* theories already entails that some mathematical theories are *indispensable* to our best scientific theories. For, presumably, simplicity is a theoretical virtue featuring in our account of what the best theories are. If they were dispensable, then the theory doing without mathematics would — all other things being equal — be better because it is simpler.

However, any such claim is in need of further argument. In fact, that our best theories include mathematical theories is very contentious. *Dispensabilists* such as Field (1989) accept **(IBE)** but deny **(Indispensability)**. They argue that we can rephrase our scientific theories in such a way that they do not quantify over mathematical entities, but still yield the same empirical consequences, and that this shows that mathematical theories are not indispensable to our best scientific theories. This does not mean that dispensabilists do not think that mathematical theories have an important role to play. Field, for example, still thinks that they can be used as means to shorten certain lines of reasoning. However, when the chips are down, we do not need them.

Carrying out the dispensabilist project requires substantial technical work. There are attempts to carry out this programme in detail for physics. For example, in *Science Without Numbers* (Field 1980), Field argued that we can dispense with (real) numbers in Newtonian physics. And there is some research about other parts of physics (e.g. Arntzenius & Dorr forthcoming).

Moreover, a lot depends on what theoretical virtues are important to single out our best theories. Indispensabilists might reject the claim that the theories offered by dispensabilists are more virtuous than the theories that make reference to mathematical entities. Virtues different from parsimony might be decisive.

In recent years there has been a lot of attention paid to what one might call *genuine mathematical explanations*. The thought is that mathematics does not only play a role because it is indispensable to obtain certain empirical predictions, but that it is indispensable to good scientific explanations — good answers to "Why?" questions. For something to be a good explanation, so the thought goes, more is required than just parsimony of the background theory: we also need to come to *understand* what has to be explained. It seems plausible that mathematics has an important role to play here.

The issue of mathematical explanation and their role in indispensability arguments is wide open. It seems that we are not even in possession of a suitable theory of (genuinely) mathematical explanation. Without being in possession of such a theory, it is hard to decide on which side the burden of proof lies.

Independently, a dispensabilist might respond that the indispensability of mathematics in good explanations does not suffice for us being justified in believing in the existence of numbers. The thought is that, for such purposes, a fictionalist interpretation of mathematical discourse suffices. This points towards a strategy to deny **(IBE)** for mathematics without denying it in plausible cases such as the quark case. One might argue that only those parts of theories used in our best explanations that say how the world is are subject to **(IBE)**, and that mathematics is only needed to formulate these descriptions in an elegant way, but not part of the description (Melia 2000).

Be that as it may, there are some general considerations that seem to show that the whole debate about indispensability cannot address all issues about Platonism in mathematics anyway. Firstly, even if the indispensabilist can defend **(IBE)** and **(Indispensability)**, it still seems as if mathematics and the empirical sciences should not be treated as being on a par. For example:

- We often believe mathematical propositions with a much higher degree of confidence than propositions about the physical world.
- Mathematical facts seem to obtain necessarily, whereas physical facts just contingently obtain. It is not clear whether indispensability theorists can account for this difference, and whether they can account for the necessary truth of mathematics in particular.

We need to explain such differences, and the picture underlying the indispensability argument makes it hard to see how this might be done. On this picture, mathematical entities such as numbers have exactly the same status as other theoretical entities such as quarks or strings.

Secondly, there are interesting *pure* mathematical theories which are not applied in the sciences. However, we want to acquire knowledge of these theories as well. Maybe the questions are not as pressing because the theories do not have the same practical importance; but we will have to address them sooner or later. The indispensabilist, however, cannot say anything about these theories (or even has to concede that he or she is not able to account for their literal truth).

Finally, even if the indispensability argument does give us *justification* to believe in some mathematical theories, it does not address **(Field's Constraint)**, at least not directly. There are responses to this worry, but there are also objections to these responses (see e.g. Field 1980, pp. 28ff). I cannot go into this here. I leave it with the remark that it is not at all clear who wins the debate.

In sum, it would be nice if we had an account that is less contentious and that does more justice to the particularities of mathematics. Be that as it may, I am happy to regard indispensability arguments as the Platonist's "last resort", albeit a contentious one.

1.3.4 Nominalistic positions

One might think that the failings of Gödelian Platonism and Fregean Logicism, as well as the existence of dispensabilist reformulations of scientific theories suggest that our main assumption — **(Arithmetical Platonism)** — is misguided. Maybe we can do better without it. In the end, a philosopher of mathematics has to appeal to cost-benefit considerations. If it turns out that accounting for mathematical knowledge requires giving up **(Arithmetical Platonism)**, then we have to do so. However, in this section, I shall argue that nominalists also face pressing objections, and thus are unable to motivate giving up **(Arithmetical Platonism)** by appealing to cost-benefit considerations.

Dispensabilists take the possibility that we *have to* quantify over mathematical objects very seriously. They just think that it does not obtain. Thus, they think they can avoid commitment to the truth of mathematical theories by providing nominalistic reformulations of our best scientific theories, i.e. reformulations that do not quantify over mathematical objects. Let us focus on the case of arithmetic again. Because arithmetic is dispensable to our best scientific theories, so the thought goes, we are not committed to **(Arithmetical Platonism)**.

Some dispensabilists hold that **(Arithmetical Platonism)** is false: for their background view entails that one should deny the existence of entities one does not need in order to obtain a sound world-view — this is one reading of *Occam's Razor*. Other dispensabilists are agnostic about the existence of mathematical entities. Both kinds of theorists need an error theory which explains how we could fruitfully apply arithmetic even if it was — strictly speaking — false.

Fictionalism To this end, Field (1980, 1989) endorses a fictionalist picture of mathematical discourse and combines it with an explanation of how mathematical fictions can be useful.

Very roughly, fictionalism about X is the view that X-talk is nonfactual talk like talk about a story containing non-existent characters. For example, we should be fictionalists about the Sherlock Holmes story. Sherlock Holmes does not *really* exist, but — according to the Sherlock Holmes story — he does. Fictional talk can serve communication and the expansion of our knowledge, so the thought goes, no matter whether the entities in question really exist. One trivial example is knowledge of literature. If Hero, who does not know much about Sherlock Holmes, is told that Sherlock Holmes did this and that, Hero can acquire knowledge about the Sherlock Holmes story. Hero does not take the story seriously — he knows this is only fictional discourse — but still learns something about the real world.

Field's thought is that mathematical theories such as arithmetic can be interpreted in a similar fashion. Even if mathematical theories were false when taken literally, so the thought goes, we could conceive of them as fictions about mathematical entities, which can still be used to extend our knowledge about the world. That mathematical theories can serve this purpose is established by a technical result: the conservativeness of arithmetic over the relevant empirical theories. The notion of conservativism is this:

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

Field shows that arithmetic and real analysis are conservative over his nominalized physical theories in this sense (Field 1980, pp.16-19). Thus, so the thought goes, we can endorse these theories as useful fictions since they enable us to shorten trains of reasoning, but cannot possibly yield anything undesired, such as new consequences about the real (physical) world that we could not already obtain from our physical theory alone (MacBride 1999, p. 434).

However, one of our stated epistemological aims was that we can acquire *arithmetical* knowledge, and not only additional scientific knowledge. According to Field's position, it is possible that we do not possess any arithmetical knowledge, where "arithmetical knowledge" is read as "knowledge of the natural numbers". However, there is an alternative conception of arithmetical knowledge — knowledge which is distinctive of experts on arithmetic — that Field (1984*c*) wants to account for. The idea is, very roughly, that such knowledge is logical knowledge: knowledge about consistency and knowledge about what follows from arithmetical axioms with necessity. Note that Field needs modal notions here. One reason why this is important is that Field cannot analyze modal notions in the usual way: for if modal talk is possible world talk, Field cannot avoid abstract entities (MacBride 1999, p. 446). However, assuming that Field can meet this and other challenges about his analysis of what mathematical knowledge consists of, he will at least be able to account for a suitably modified version of **(Arithmetical Knowledge)**.

So does Field win the debate? A lot depends on whether the dispensabilist programme above can be carried out for all relevant cases, and the indispensabilist responses can be rebutted. However, even if we assume that the technical programme of providing nominalistic reformulations succeeds, it is not clear whether Field really has an epistemological advantage over the Platonist.

For Field suggests to replace the mathematical version of Newtonian physics by a version that quantifies over infinitely many space-time points. Now, why are infinitely many space-time points any less problematic than numbers? In particular:

- How do *these* entities meet **(Field's Constraint)**? Is the claim that we can reliably form true beliefs about such objects any less problematic than the claim that we can reliably form true beliefs about numbers? Although there is a sense in which space-time points exist in space and time, they are not like tables and chairs. We need an additional story of how we can refer to them and acquire knowledge of them. In this respect, Field does not have any initial explanatory advantage over the Platonist.

- Field's reformulation presupposes the existence of infinitely many concrete objects. This is problematic. Intuitively, it should be possible that only finitely many concrete objects exist. However, if there were only finitely many objects, Field's reformulated theory could not be true.[26]

In sum, it is not clear that Field's position does any better with regard to our initial puzzles. Field discusses some such worries (Field 1984*a*), but I cannot pursue this issue any further here.

[26] Also, suppose that there are infinitely many nominalistically acceptable objects. Then why not take numbers to be space-time points and say that numbers exist after all? This will enable us to preserve at least **(Minimal Arithmetical Realism)**.

Structuralism Some nominalists have tried to provide paraphrases for the mathematical theories that do not carry commitments to the existence of infinitely many objects of any sort. Paraphrases can either be understood as providing a kosher replacement of Platonist theories which are, strictly speaking, false; or as displaying the real content of mathematical statements. Only the latter strategy will allow for some form of hermeneutic reconstruction. The first can at best achieve what I called (**Revolutionary Reconstruction**). I only consider hermeneutic strategies here.

According to *eliminative structuralism*, arithmetical claims should not be read as particular propositions about natural numbers, but as universal propositions about all natural number structures. For example, the claim that 1+1=2 should be read as "For every natural number structure S[27], the interpretation of "+" in this structure maps the sum of its first element and its first element to its second element". It is easy to see that accepting this sentence does not commit one to the existence of any entity, since it is a universally quantified conditional.

However, this lack of commitment backfires. If there is not at least one natural number structure, the universally quantified paraphrase will be vacuously true (Parsons 1990, p. 310). However, it is a necessary condition for there to be a natural number structure that there are infinitely many objects. And how is that claim to be justified?

There is a variant of eliminative structuralism which tries to avoid this problem by using modal notions — so-called *modal structuralism* (Hellman 1989). The idea is that the problem can be avoided by demanding that the universally quantified conditional needs to obtain necessarily. Even if there actually are no natural number structures, it is certainly *possible* that there are such structures. Thus, the sentence "Necessarily, for every natural number structure S, the "+"-function in this structure maps the sum of its first element and its first element to its second element" will not be vacuously true.

However, the use of modal notions backfires as well. The modal structuralist might be able to avoid ontological commitment to natural numbers, but only by increasing his ideological commitments. For he needs to explain the used modal notions, and this explanation should not involve reference to abstract objects. Possible world talk, for example, is not an option. It is plausible that modal notions have to be taken as primitive.

[27]That is: for every structure S that satisfies the Peano axioms.

It is not clear whether the resulting position is better off in terms of explaining how we acquire mathematical knowledge. For the modal structuralist not only needs to explain his modal vocabulary, but also needs to account for modal knowledge, such as knowledge of the possible existence of infinitely many objects (Ebert 2005*a*, pp.31f).

1.4 Intermediate conclusion

Let me take stock. I presented two puzzles about arithmetical knowledge, followed by a list of desiderata for a fully satisfactory epistemology of mathematics. I argued that classical Platonist positions face substantial difficulties, and violate one or more of our epistemological constraints. A version of the Indispensability Argument has been identified as the Platonist's last resort. I then temporarily bracketed **(Arithmetical Platonism)** and examined nominalist positions, to see whether the epistemological problems can be avoided. It transpired that relevant nominalist positions face substantial difficulties as well. Thus, there is no reason to give up **(Arithmetical Platonism)** already. We could just as well reconsider Platonist positions. This sets the stage for yet another Platonist position: ***abstractionism*** (or ***neo-Fregeanism***).

2 Neo-Fregeanism (abstractionism)

In this chapter, I examine yet another Platonist position, and argue that it promises to meet all our epistemological constraints, although there is some unclarity about its exact epistemological workings. The position is based on a train of thought that Frege entertained, but dismissed in *GL* — that **(HP)** is a proper definition of number and can underlie our knowledge of arithmetic.

2.1 Frege's Theorem

In (Parsons 1965), Parsons made explicit what is nowadays known as *Frege's Theorem*: the astonishing above-mentioned technical result, implicit in Frege's work, that the concept of extension is not needed to obtain arithmetic once one has **(HP)** available, and thus that **(HP)** as a single axiom suffices for a derivation of the Peano axioms in second-order logic (Parsons 1965, p. 194). Call the deductive closure of **(HP)** under second-order consequence *Frege Arithmetic* (*FA*). The result can be expressed as follows:

(Frege's Theorem) FA *interprets* second-order Peano Arithmetic.

The notion of interpretation I use here is the notion of *relative interpretability*, introduced by Feferman (1960). In short, a theory T_1 relatively interprets another theory T_2 just in case a definitional expansion of T_1 (syntactically) entails a version of T_2 with relativized quantifiers. In our case, the relevant result is that a definitional expansion of FA (syntactially) entails the second-order Peano axioms, relativized to a number predicate.[28] For example, the successor axiom a definitional expansion of FA needs to entail is not:

$$\forall x \exists y\, (Sxy)$$

But:

$$\forall x\, (Number\,(x) \rightarrow \exists y\, (Number\,(y) \wedge Sxy))$$

[28] A complete statement of a "relativized" version of the axioms can be found in (Wright 1983, p.158), for example.

For the purposes of this chapter, nothing hangs on the notion of interpretability, and the reader can think of the theorem as the result that **(HP)** *entails* (second-order) PA.[29]

Almost twenty years later, Wright (1983) provided a proof of Frege's Theorem in modern notation and conjectured that FA is consistent. This conjecture was later confirmed by Boolos (1987), who proved that FA is consistent if and only if second-order arithmetic — which is deemed consistent by the whole mathematical community[30] — is consistent.

These technical results motivate a new position in the philosophy of mathematics, when combined with some optimism about the epistemic and semantic status of **(HP)**. Could we not base arithmetic on **(HP)** directly, as opposed to proving **(HP)** from some other principle (that, in Frege's case, led to inconsistency)?

Parsons already notes that "Frege *does* show that that the logical notion of one-to-one correspondence is an essential constituent of the notion of number" (Parsons 1965, p. 203). Wright (1983) suggests that some Fregean abstraction principles, and **(HP)** in particular, have a better epistemic and semantic standing than Frege thought they had in *GL*, based on considerations about Frege's Context Principle. In particular, Wright argues that the Caesar problem can be solved and that **(HP)** is a proper explanation of the number operator. Thus, Frege made a major mistake in abandoning **(HP)** as a definition, trying to define the concept of number in a different way, and proving **(HP)** from this definition. The results above show that it is only because of this further move that Frege's system becomes inconsistent.

This position became known as *neo-Fregeanism*, a name emphasising the origins of the position. Later, it also became known as *abstractionism*, which emphasises the central role of abstraction principles, and the possibility to generalize the position.[31] In this chapter, I will be concerned with the details of this position.

Before I turn to the details, a note is in order concerning the scope of my presentation. Firstly, although Wright's original line of thought rests a particular interpretation on Frege's Context Principle, I will not be concerned with an exegesis of this principle

[29] I am, however, aware of the fact that it is very important to be precise here, and that the notion of interpretability one uses has philosophical consequences (Walsh 2010). Unfortunately, I cannot discuss these issues in this thesis.

[30] There are always exceptions. However, so far every attempt to show that arithmetical theories are inconsistent — notably a recent attempt by Edward Nelson (2011) — has failed.

[31] Sometimes also "neo-Logicism". However, I prefer to use "neo-Logicism" to single out a particular component (aim) of the neo-Fregean position, namely that arithmetical knowledge can be based on logic and definitions.

here. I will only sketch the mature position as it is explicated and defended today. My primary resources will be Hale's and Wright's collection on neo-Fregeanism (Hale & Wright 2001*a*), their survey essay (Hale & Wright 2005), and MacBride's survey essay (MacBride 2003).[32] Secondly, I will keep focusing on the arithmetical case. Discussing neo-Fregean reconstructions of other mathematical theories is beyond the scope of this thesis.

2.2 Neo-Fregeanism: two (Fregean) aims

Neo-Fregeanism (for arithmetic) has two components — or aims — which exactly correspond to the two major elements of Frege's philosophy of arithmetic (Hale & Wright 2005, pp. 166f). The first component may be called *neo-Fregean Platonism*.[33] It is the conjunction of a semantic and a metaphysical claim:

(Neo-Fregean Platonism)

- Number terms refer to objects: the numbers.
- Numbers are mind-independent abstract objects.

Note that this claim is just terminologically different to what I dubbed **(Arithmetical Platonism)** in 1.2. The second component may be called *neo-Fregean Logicism*, and is an epistemological claim:

(Neo-Fregean Logicism) There is a route to acquiring apriori knowledge of arithmetical truths on the basis of logic and definitions.

However, both "logic" and "definitions" have to be construed relatively widely here: for the neo-Fregean logic comprises second-order logic, and the definitions include Fregean abstraction principles such as **(HP)**.[34] I now examine both claims in some more detail.

[32] Although there is more recent work on neo-Fregeanism (e.g. Hale & Wright 2009; Wright 2009), the later work introduces new lines of thought which sometimes seem to be opposed to the original ideas. Moreover, there seems to be some divergence between Hale and Wright in most recent work. I will come back to some of these issues later in this thesis.

[33] I endorse the terminology that Hale and Wright use in the introduction to their (2001*a*).

[34] This yields two points of divergence from Frege's own view. Firstly, neo-Fregean logic is weaker than Frege's, because it does not include set theory. Secondly, the neo-Fregean notion of definition is wider than Frege's: Frege just accepted *explicit* definitions in his framework, whereas neo-Fregeans need an additional notion of *implicit* definition. I expand on these two conceptions of definition below.

2.2.1 Neo-Fregean Logicism

The idea underlying **(Neo-Fregean Logicism)** is as follows. First, according to the neo-Fregean, **(HP)** can be known apriori:

> **(Abstraction)** A range of (good) abstraction principles — including, of course, **(HP)** — can be known apriori.[35]

Secondly, a priori knowledge is *transmissible over second-order consequence.*[36] That is:

> **(Transmission)** If we know ϕ apriori and ψ is a second-order deductive consequence of ϕ, then, ceteris paribus[37], we can come to know ψ apriori by virtue of deriving it from ϕ by means of second-order logic.

Thus, we can acquire apriori knowledge of FA, on the basis of logic and definitions. By **(Frege's Theorem)**, we can obtain apriori knowledge of a theory that interprets second-order arithmetic on this basis.

However, there is a potential gap between the claim that one reconstructed knowledge *of a theory interpreting* arithmetic and the claim that one reconstructed *arithmetical* knowledge. So the neo-Fregean also needs to claim:

> **(Same Subject Matter)** Knowledge of the FA-interpretation of arithmetic is knowledge of arithmetic.

This claim is ambiguous. Remember the discussion of hermeneutic reconstruction in the last chapter. Focusing on the case of pure arithmetic, the claim could either be that the number terms of FA refer to the same objects as the number terms of second-order PA or ordinary number terms, or that relevant statements of FA even express the same thoughts as corresponding theorems of second-order PA or ordinary mathematical statements. I will discuss these hermeneutic claims in due course, and temporarily assume the weaker claim for the sake of the argument. Given this assumption, the neo-Fregean will be able to meet both **(Arithmetical Knowledge)** and **(Arithmetical Foundationalism)**.

[35]Note that the claim is not just a claim about **(HP)**. This more general claim is required for extending the programme to other parts of mathematics, and I will come back to it below.

[36]In fact, the neo-Fregean does not require second-order logic with full comprehension in order to prove a version the Peano axioms. Π^1_1-comprehension suffices (Linnebo 2004).

[37]The reasons for invoking a ceteris paribus clause will become apparent in chapters 4 and 5. However, for the purposes of this chapter, the reader can safely ignore this complication.

How can **(Abstraction)** and **(Transmission)** be justified? The neo-Fregeans do not say much about logic, but the thought seems to be that (i) any proper logic[38] transmits apriori knowledge, and (ii) the following holds (see e.g. Hale & Wright 2001*a*, p.430):

(Logicality) Second-order logic is logic proper.

This entails **(Transmission)**. A similar thought applies to the claim that the relevant abstraction principles are (implicit) definitions. Surely, so the thought goes, any *proper* definition can be known apriori. Thus, neo-Fregeans also claim:

(Definition by Abstraction) The relevant abstraction principles — including, of course, **(HP)** — are proper (implicit) definitions of the respective abstraction operators.

(Logicality) and **(Definition by Abstraction)** provide the basis for **(Neo-Fregean Logicism)** — the claim that mathematics follows from logic and definitions. However, the connections between logic and mathematics the neo-Fregeans postulate might run deeper. This is because it seems an attractive option for the neo-Fregean to hold that logic is also semantically and epistemically based on (implicit) definitions.

That Hale and Wright possess the resources to endorse such a picture of logic becomes apparent at several occasions. In their work on implicit definition (Hale & Wright 2000), they also consider Gentzen's suggestion that logical constants can be defined by a stipulation of their introduction and elimination rules (Gentzen 1934). In later work, Wright (2007*a*) explicitly suggests endorsing inferential-role semantics for the (second-order) quantifiers, citing the above-mentioned paper on implicit definition. Although Wright's paper is mainly on the issue of ontological commitment, Wright mentions in a footnote that he hopes that this move also solves epistemological problems (Wright 2007*a*, footnote 9).

There are good reasons to apply the idea of implicit definition to both abstraction principles and logical operators. This would provide a uniform epistemic and semantic basis for logic and mathematics, and fits well together with Wright's characterization of logicism as "the thesis that logical knowledge and at least basic mathematical knowledge are, in some important sense, of a single epistemological kind" (see Wright 2007*a*, p.4). Note the similarity to the **(Same Source)** desideratum from the last chapter. As I said, I do not think that this claim is strong enough to capture what logicists such as Frege

[38] This is meant to exclude dialetheism etc.

really had in mind, since the rational intuition proposal would also meet this description. Logicism should rather be conceived of the **(Same Source)** claim plus the claim that logico-mathematical knowledge is *not* based on rational intuition. This is precisely what the neo-Fregeans would establish if the idea of implicit definition could be applied to logic and mathematics.

2.2.2 Neo-Fregean Platonism

What is the neo-Fregean argument for **(Neo-Fregean Platonism)**? The basic idea is this: (i) if a true atomic sentence contains a singular term t, then t will refer to an object; (ii) since the number terms of FA are singular terms, and FA is true (we can know that by **(Neo-Fregean Logicism)**), number terms refer to objects; (iii) because of **(Same Subject Matter)**, ordinary number talk is also (true) talk about these objects, and the ordinary number terms also refer to objects; (iv) some additional considerations show that these objects must be mind-independent abstract objects.

The steps the neo-Fregeans pay most attention to are (i) and (ii). Wright (1983) argues that these steps can be extracted from Frege's *Context Principle* — the claim that we should never ask for the meaning of a word in isolation, but only in the context of a proposition. According to Wright, one can interpret the Context Principle as:

> the thesis of the priority of syntactic over ontological categories. According to this thesis, the question whether a particular expression is a candidate to refer to an object is entirely a matter of the sort of syntactic role which it plays in whole sentences. If it plays that sort of role, then the truth of appropriate sentences in which it so features will be sufficient to confer on it an objectual reference. (Wright 1983, p.51)

The thesis consists of three claims, which MacBride (MacBride 2003, p. 108) calls *Syntactic Decisiveness*, *Referential Minimalism*, and *Linguistic Priority*. The first is:

(Syntactic Decisiveness) Expressions which syntactically behave like singular terms, are singular terms (i.e. terms which have the role to refer to objects).

Thus, so the thought goes, purely syntactic investigations can reveal that some terms are singular terms, and thus that they have *referential potential*.[39] Of course, this requires

[39]This is also Fine's interpretation of the Context Principle (Fine 2002, pp. 57f).

an account of what kind of syntactic behaviour is characteristic for singular terms. Hale (1994*b*, 1996) builds up on Dummett's idea (Dummett 1973) that one can argue for the claim that number terms are singular terms by examining their inferential role.

Examining the inferential role of the number terms in FA is not too difficult, because it is an artificially introduced formal system with well-defined properties. Presumably, such investigation can be carried out apriori.[40] After the referential potential of number terms has been revealed by appropriate criteria, so the thought goes, one can infer from the truth of a sentence which contains such terms that they must *refer*:

> **(Referential Minimalism)** If a true atomic sentence containing a singular term t is true, then t refers (MacBride 2003, p.108).

Since one's ontology may not only contain *objects* — for instance, we should allow for the possibility of there being objects and concepts — the following third claim is needed:

> **(Linguistic Priority)** If a singular term refers, it refers to an object.

This completes the argument for (i) and (ii). I discuss (iii) in 2.3.5. As to (iv), in their classical works, Hale and Wright say surprisingly little, except when it comes to the related Caesar problem. I briefly return to this issue when I discuss this problem in 2.4.1.

Secondly, the question arises of whether **(Neo-Fregean Logicism)** is really compatible with **(Neo-Fregean Platonism)**. In particular: is it possible to lay down **(HP)** as a definition if it commits us to the existence of infinitely many objects? Can the heavyweight ontological commitments of such abstraction principles be reconciled with the proposal that they are "just definitions"?[41] We have to examine the neo-Fregean conception of implicit definition in more detail.

[40] Of course, the reasoning could also be applied to ordinary number talk directly. Additional complications might arise in this case, because ordinary number talk does not take place in a formal system. In particular, one might wonder whether the required investigations are aposteriori. The presented version of neo-Fregeanism faces these worries when it comes to step (iii). I come back to these complications when I discuss hermeneutic neo-Fregeanism and the **(Same Subject Matter)** claim in 2.3.5.

[41] There is a related further problem: if property terms refer to properties, and second-order variables range over properties, how can second-order logic be logic? I cannot discuss this problem here. One possible answer is a position called "neutralism" — defended in (Wright 2007*a*) — according to which second-order quantifiers are not associated with domains of entities of any sort. Another option is to bite the bullet, arguing that a commitment to properties is just as unproblematic as a commitment to numbers. For this train of thought, see (Hale & Wright 2005).

2.3 Implicit definition

The neo-Fregean account of implicit definition — which is to account for knowledge of abstraction principles and maybe also for knowledge of the underlying logic — is discussed in (Hale & Wright 2000). In this paper, Hale and Wright defend the following thought (they dub it the *Traditional Connection*):

(Neo-Fregean Implicit Definition Thesis) "at least some important kinds of non-inferential apriori knowledge are founded in implicit definition" (Hale & Wright 2000, p. 177).

To implicitly define something is to stipulate that a sentential matrix containing a hitherto undefined term is to be true (Hale & Wright 2000, p. 117). By virtue of this act, so the thought goes, one fixes a pattern of use for the hitherto undefined term such that it is endowed with the unique meaning that renders the matrix true. Moreover, by virtue of this act, one can come to know the stipulation apriori.

I will ultimately endorse and defend this idea. However, on my own account, knowledge founded in implicit definitions turns out to be *inferential*. So Hale's and Wright's claim needs to be distinguished from the claim I will defend later, which is the following, weaker claim:

(Implicit Definition Thesis) At least some important kinds of apriori knowledge are founded in implicit definition.

2.3.1 Explicit vs. implicit definitions

The term "definition" can either be applied to acts of stipulating something or to the linguistic entities that are used as the defining devices. I will use the term in both ways, without explicitly mentioning which sense of "definition" is intended, since this will always be clear from the context. Traditionally, only a small class of sentential matrices are regarded as definitions proper — so-called *explicit definitions*:

(Traditional Conception of Definition) Only *explicit definitions* are definitions proper, i.e. definitions which are both *eliminable* and *conservative* (*non-creative*)

In this context, the following informal characterizations of these two criteria are sufficient:

(Eliminability) All statements in the new language — the language including the new term — can be translated into statements of the old language — the language without the term introduced by the definition — whereas the two statements are (provably / semantically) equivalent in the new theory.

And:

(Conservativeness) All statements of the old language which are consequences of the new theory — the old theory together with the definition — are also consequences of the old theory.

In a first-order language, definitions fulfilling these two criteria can be stated in "normal form" (Gupta 2009). This is the form paradigmatic "explicit definitions" take. It renders it immediate that they are eliminable. For example, a new one-place predicate P can be introduced by stipulating the biconditional "$\forall x\,(Px \leftrightarrow \phi\,(x))$" to be true, where ϕ is already understood, the only free variable in ϕ is x, and P does not occur in ϕ. For example: "For all x, x is a bachelor if and only if x is an unmarried man". To eliminate "Pa" (e.g. "Frank is a bachelor"), one just has to replace it by "$\phi\,(a)$" ("Frank is an unmarried man").

Explicit definitions impose no special epistemic or semantic obstacles whatsoever. Because of **(Eliminability)**, the meaning of the new terms is uniquely determined (assuming all the terms of the "old language" have determinate meanings), and the new language (including the defined term) can be seen as an expansion providing convenient shortcuts. These shortcuts are epistemically unproblematic, because the definition meets **(Conservativeness)**.

The **(Standard Conception of Definition)** has been endorsed by both Frege (1893) and Russell (Russell & Whitehead 1910-1913). However, it cannot be endorsed by the abstractionist, because relevant abstraction principles such as **(HP)** do not meet the criteria: "$\#F$" as defined by **(HP)** is not eliminable in contexts where the number operator occurs in the definition of the property F (Dummett 1991, p.138); and **(HP)** is not conservative in the above-defined sense either, because it (semantically and syntactically) entails the claim, expressible in second-order logic, that there are infinitely many objects.

Since **(HP)** is regarded as a definition, the neo-Fregean needs to reject both **(Eliminability)** and **(Conservativeness)**. There is, however, a second reason why the neo-Fregean might want to reject the **(Standard Conception of Definition)**. If the neo-Fregean wants to account for **(Same Source)** by applying the idea of knowledge by definition to the (basic) logical case — building on Gentzen's idea — he or she also needs to account for the stipulability of (the validity of) *inferential matrices*, and not only for the stipulability of sentential matrices. This yields the:

> **(Liberal Conception of Definition)** Ceteris paribus, we can define one or more terms by stipulating that sentential matrices containing one or more undefined terms are to be true (necessarily true), or by stipulating that inferential matrices containing one or more undefined terms are to be valid[42].

The ceteris paribus clause is crucial. Of course, there will be some restrictions. Not every sentential matrix or inferential matrix containing an undefined term will amount to a definition proper. However, so the thought goes, there are relevant cases in which there is *semantic success* — i.e. a unique meaning will be fixed such that the stipulation is true (or valid) — and *epistemic success* — i.e. the stipulation will enable the epistemic agent to acquire apriori knowledge of the truth of the stipulation. This possibility can be initially motivated as follows:

- Semantic success: because explicit definitions just provide shortcuts, they cannot possibly serve the introduction of new fundamental concepts, or provide means to refer to objects we are not able to refer to in the old language. We need to explain that possibility. Implicit definition would provide such a means in *some* relevant cases (Hale & Wright 2000, p. 18).

- Epistemic success: we need to explain the possibility of apriori knowledge. The idea that *some* apriori knowledge is grounded in definition is initially plausible, and has a long tradition (Hale & Wright 2000, p. 117). Moreover, it would provide a way to apriori knowledge that is not grounded in a dubious faculty of rational intuition (see e.g. Boghossian 1996).

[42]Think of validity as necessary truth-preservation here.

For the neo-Fregean, two cases are particularly relevant. Firstly, the case of **(HP)**. The neo-Fregean needs to argue that one can stipulate its truth, and thereby not only bring it about that "#" gets assigned the unique meaning that renders it true — by virtue of fixing a sufficient pattern of use for "#" — but also be able to acquire apriori knowledge of **(HP)**. Secondly, it transpired that the implicit definition theorist might also want to allow for the stipulation of the validity of introduction and elimination rules. As an example, consider the stipulation of *Modus Ponens (MP)* and *Conditional Proof (CP),* which is meant to define the conditional "$\rightarrow$":[43]

$$\text{(MP)}\ \frac{\Gamma,\phi\vdash\psi}{\Gamma\vdash\phi\rightarrow\psi} \quad \text{(CP)}\ \frac{\Gamma\vdash\phi \text{ } and \text{ } \Delta\vdash\phi\rightarrow\psi}{\Gamma,\Delta\vdash\psi}$$

By virtue of stipulating CP and MP to be valid, so the thought goes, one can fix a pattern of use such that "$\rightarrow$" is endowed with the unique meaning that renders these rules valid, and also come to know apriori that they are valid.

Two questions need to be answered in both the mathematical and the logical case:

1. How can it be brought about that — in the relevant cases — a unique meaning is fixed? In other words: in virtue of what is semantic success achieved?

2. How can it come about that — in the relevant cases — the epistemic agent can arrive at apriori knowledge of the truth or the validity of the stipulation in question? In other words: in virtue of what is epistemic success achieved?

I now turn to the (neo-Fregean) answers to these questions.

2.3.2 Three dimensions of achievement

We need to clarify the notions of semantic and epistemic success. The notion of semantic success is ambiguous. Using a well-known Fregean distinction, we can read "meaning" as either "sense" or "reference". It is thus useful to follow Ebert (2011) in distinguishing the following three dimensions of achievement:

(Effectiveness) An implicit definition is *effective* if and only if, by virtue of the stipulation, a sufficient pattern of use is fixed such that the hitherto undefined term is

[43]The rules are taken from (Bostock 1997, p.388).

endowed with a unique Fregean sense (concept), and is thereby understood by whoever carried out the stipulation (Ebert 2011, p.10).

(Success) An implicit definition is *successful* if and only if it is effective and the stipulation brings it about that the hitherto undefined term gets assigned a unique referent such that the stipulation is true / valid (Ebert 2011, p.9).

(Epistemic Productiveness) An implicit definition is *epistemically productive* if and only if, by virtue of making the stipulation, the epistemic agent can acquire apriori justification for the truth of the stipulated matrix or sentential pattern, and can even acquire apriori knowledge of it, in case the stipulation is successful (Ebert 2011, p.10).

The distinction between **(Effectiveness)** and **(Success)** can be further motivated by examples. For instance, Ebert (2011, pp.12f) argues that a stipulation of **(BLV)** is a case which cannot possibly be subject to **(Success)**, but which might suffice to generate understanding of (and to uniquely fix a Fregean sense for) the extension operator. On such a view, a stipulation of **(BLV)** is subject to **(Effectiveness)**, but not to **(Success)**.

However, there are implicit definitions which are not even subject to **(Effectiveness)**. As an example, take the stipulation that "If the moon exists, then it is F" as an attempt to fix the meaning of F. It is impossible to understand F on this basis (Hale & Wright 2000, pp. 134).

Another interesting example can be found in the rule case. To show that not every pair of introduction and elimination rules can be stipulated in order fix the meaning of a connective, Prior (1960) considers the following introduction and elimination rules, which are stipulated in order to fix the meaning of "τ" ("tonk"):

$$(\tau\text{-I})\ \frac{\phi}{\phi\tau\psi} \quad (\tau\text{-}E)\ \frac{\phi\tau\psi}{\psi}$$

The tonk rules allow for the derivation of any statement from any premise. Thus, the stipulation cannot be subject to **(Success)**. Ebert (2011), thinks that it is not even subject to **(Effectiveness)**.

2.3.3 Proposed conditions for semantic and epistemic success

The above cases motivate looking for conditions that a stipulation has to meet in order for being subject to the above-mentioned dimensions of success. Without such conditions, the neo-Fregean does not possess a complete account of implicit definition. In particular, the neo-Fregean will not be able to respond to unspecific doubts to the effect that **(HP)** is really subject to the **(Neo-Fregean Implicit Definition Thesis)**. In their major work on implicit definition (Hale & Wright 2000), Hale and Wright impose five such conditions:

Generality In order for a definition to be subject to **(Effectiveness)**, it must succeed in generating understanding. According to Hale and Wright (2000, pp.134f), a plausible necessary condition for understanding a term is that a version of Evan's Generality Constraint is met. That is: in order to understand t, one needs to understand all relevant contexts "$\phi[t]$", where the matrix $\phi[\_]$ is already understood. A stipulation that is supposed to fix the meaning of t needs to bring it about that t is understood in all relevant contexts.[44] Among other things, this criterion rules out stipulations that do not sufficiently contrain the meaning of the term to be defined, such as "If the moon exists, it is F".

Consistency Inconsistent stipulations cannot possibly be true, so consistency is a necessary condition for **(Success)** (Hale & Wright 2000, p.132). Note that whether a stipulation is regarded as being consistent may vary with the logical background system. For example, the condition will not rule out **(BLV)** if we restrict ourselves to predicative comprehension. However, Hale and Wright accept second-order logic with full comprehension, so they will rule out **(BLV)** for reasons of inconsistency.

Harmony Consider a stipulation of introduction and elimination rules. It is important that both kinds of rules work well together. One way to make this precise is to demand that the introduction and elimination rules need to be *in harmony*, i.e. that the result of applying the elimination rule is not any stronger than the conditions for applying the corresponding introduction rule, and vice versa (Hale & Wright 2000,

[44]The requirement can thus be seen as a weakened eliminability criterion. The current condition is that we need to *understand* the new terms in a sufficient range of sentential contexts, and not that we need to be able to *provide translations* to the old language.

p.136).[45] Stipulations which are in disharmony cannot be subject to **(Success)**, and presumably not even to **(Effectiveness)**. The tonk rules will be ruled out by this constraint.

Weak Conservativeness According to Hale and Wright, we still need *some* conservativeness criterion in order to rule out stipulations which say something substantial about the world and thus cannot be regarded as "epistemically innocent" (Hale & Wright 2000, p.133). Consider the stipulation "Jack the Ripper is the perpetrator of all these crimes", laid down to fix the meaning of "Jack the Ripper". There is only an appropriate referent to be assigned to "Jack the Ripper" if there is a single person who is the perpetrator of all these crimes, and this claim should not be part of a definition (Hale & Wright 2000, p.134). Thus, although the definition might be subject to **(Success)**, it cannot be subject to **(Epistemic Productiveness)**, at least not without the subject doing substantial further aposteriori epistemic work. However, Hale and Wright cannot use the notion of conservativeness presented above, for it would rule out **(HP)**. The notion they use is a bit weaker, so I call it "weak conservativeness". An abstraction principle is *weakly conservative* just in case it does not generate any new consequences, in the old language, about any old domain — i.e. any domain of objects we might recognize before we introduced the abstraction principle (Hale & Wright 2000, p. 133). Note that every weakly conservative definition is also consistent, so the consistency condition is — strictly speaking — redundant.

Non-arrogance There is another condition for **(Epistemic Productiveness)** which looks redundant, at least at first glance: the *non-arrogance constraint.* A stipulation is arrogant just in case its truth "cannot justifiably be affirmed without collateral (a posteriori) epistemic work" (Hale & Wright 2000, p. 128). Consider again the case of Jack the Ripper. It is arrogant according to Hale's and Wright's definition. However, such stipulations are already ruled out by the conservativeness criterion. So is the non-arrogance constraint redundant? It does not seem so. For although Hale and Wright define non-arrogance as applying to cases of where one needs to do further *aposteriori* epistemic work, they later also apply it to cases where they think

[45] It is a difficult question of how the notion of harmony should be spelled out in full detail. Hale and Wright do not say anything about how they think the notion should be explicated.

that further *apriori* epistemic work needs to be done in order to justifiably assert the proposed definition. For example: a direct stipulation of the truth of the Peano axioms. Hale and Wright explicitly rule out this stipulation for the reason that it is a *direct stipulation of existence*, and thus arrogant (Hale & Wright 2000, p.147; Hale & Wright 2007, section 2). Unfortunately, it remains unclear what exactly makes a stipulation arrogant. More work is required here. In particular, Hale's and Wright's characterization is trivial because it is trivial that we cannot obtain knowledge just by virtue of a stipulation if the stipulation requires further collateral epistemic work. Thus, the constraint does not provide any guidance for deciding which stipulations are subject to **(Epistemic Productiveness)** that goes beyond our pre-theoretic intuitions. It is certainly not an easy task to find a precise and independently motivated condition that rules out the Peano axioms but is met by **(HP)**, but it is a task that the neo-Fregeans cannot avoid. I will come back to this issue.

Determining a proper collection of conditions is a research project on its own. In fact, the set of conditions for **(Success)** cannot possibly be complete, because some of the principles meeting the conditions will be jointly unsatisfiable and thus cannot all be true. In particular, there are consistent and conservative but mutually unsatisfiable abstraction principles. I will discuss this problem further in 2.4. Let us assume, for the sake of the argument, that the neo-Fregeans have found a set of conditions whose obtaining guarantees that a stipulation is subject to **(Success)**, and that the relevant stipulations — including **(HP)**— turn out to meet these conditions.

2.3.4 The epistemology of implicit definition

It is still a pressing question how we should conceive of the epistemological process in individual cases. How exactly can an epistemic agent acquire apriori knowledge, by virtue of making a stipulation that meets all the conditions? It does not suffice to claim that this is *possible* as soon as the constraints are met. We need an *explanation* of how this can happen. Here is what Hale and Wright have to offer:

> How, just by stipulating that a certain sentence, '#f', it true (...) is it supposed to be possible to arrive at an a priori justified belief that #f? Well, the route seems relatively clear *provided* two points are granted: first that a stipulation

> of the truth of the particular '#f' is so much as properly possible— that the truth of the sentence is indeed something we can settle at will; and second that the stipulation somehow determines a meaning for 'f'. If both provisos are good, it will follow that the meaning bestowed on 'f' by the stipulation cannot be anything other than one which (...) results in the truth of the sentence in question. (...) Moreover, if the stipulation has the effect that 'f' and hence '#f' are fully *understood* (...) then nothing will stand in the way of intelligent disquotation: the knowledge that '#f' is true will extend to knowledge that #f. In other words: to know both that a meaning is indeed determined by an implicit definition, and what that meaning is, ought to suffice for a priori knowledge of the proposition thereby expressed. (Hale & Wright 2000, pp. 126f)

That a stipulation of the truth of a particular matrix is "so much as properly possible" and that a meaning is determined is meant to be ensured by the conditions above. According to Hale and Wright, an epistemic agent making the stipulation will then know apriori that it is true. On the basis of such knowledge, so the thought goes, one can then acquire knowledge of the stipulation itself by applying a disquotational step, given that the sentential matrix is understood.

This raises a couple of issues. Firstly, Hale's and Wright's route to apriori knowledge of the matrix whose truth is stipulated turns out to be *inferential*, since it requires making a disquotational step (for such a reading of this paragraph, see also Ebert 2005*b*, section VI; Ebert 2011, pp.22f). This is in tension with the **(Neo-Fregean Implicit Definition Thesis)**, which is about *non-inferential* apriori knowledge.

Maybe the claim is only meant to apply to knowledge of the *truth of* stipulations: nothing said so far rules out that *this* knowledge is non-inferential. However, secondly, it is entirely unclear how exactly a successful act of stipulating a sentential matrix generates knowledge of the truth of the stipulation in question. Hale and Wright contend that:

> a thinker who is party to the stipulative acceptance of a satisfactory implicit definition is in a position to recognize both that the sentences involved are true — precisely because stipulated to be so — and what they say. (Hale & Wright 2000, p. 138)

This suggests that the route to knowledge of the truth of a stipulation goes via a further line of reasoning, involving a reflection on the stipulation and its content. So maybe Hale and Wright cannot maintain the claim that such apriori knowledge is non-inferential and should only endorse the weaker **(Implicit Definition Thesis)**, which also allows for inferential apriori knowledge of the truth of stipulations. In what follows, I will only talk about this weaker claim for this reason.

In any case, the details of the knowledge generating process are left open. And, as I will show in 7.1.4, the details matter. I take this to be a major gap in the neo-Fregean proposal, which I close in chapter 7.

Let us ignore this gap for now, and discuss an argument purporting to show that — regardless of how the gap is filled — the neo-Fregean faces a dilemma. The dilemma is due to Shapiro and Ebert (2009).

In their paper, Shapiro and Ebert discuss the question of what the relation between the stipulating epistemic agent and the conditions for semantic success must be in order for the agent to acquire knowledge by virtue of making the stipulation (Ebert & Shapiro 2009, p.11). On the one hand, so the thought goes, one might hold that the conditions just have to be true, whether or not the epistemic agent has access to this fact. On the other hand, one could demand that the fact that these conditions are met must be reflectively available to the agent. Prima facie, these two options may seem to be exhaustive. If this is correct, then the neo-Fregean must either hold:

(Externalism) The conditions must be true.

Or:

(Internalism) That the conditions are met must be reflectively available to the agent. Thus, either the stipulating agent needs to show that the conditions hold, or that the conditions hold must be self-evident.

Given this dichotomy, Ebert and Shapiro (2009) argue that the neo-Fregeans face a dilemma. In particular, they argue that both options are problematic even in the central case of **(HP)**.

To see why **(Internalism)** is problematic, consider the consistency condition. Surely, the consistency of **(HP)** is not self-evident. And this is especially so in the light of inconsistent abstraction principles such as **(BLV)**. So we should demand that the agent needs to show that **(HP)** is consistent. However, in this context, this plausibly amounts to the requirement that the agent needs to prove that **(HP)** is consistent. And because of Gödel's second incompleteness theorem, a proof of the consistency of **(HP)** cannot be carried out in any weaker (and safer) system than the one that **(HP)** is meant to deliver. We face an epistemic regress problem (Ebert & Shapiro 2009, p. 426).

Is **(Externalism)** any better? Initially, it is hard to see how the mere truth of the conditions could ensure that the stipulating agent is epistemically responsible in any sense of epistemic responsibility required for the possession of knowledge of the truth of the respective stipulations. This worry affects externalistic construals of knowledge in general and has been contested. In any case, it requires a deeper investigation of the issue of externalism and internalism about knowledge. I come back to this issue in the next chapter.

There is a related, more specific worry. Suppose that **(HP)** meets all the conditions. Then a whole class of further abstraction principles allowing for easy proofs of very complex mathematical theorems will also meet all the conditions. For as we know from Richard Heck (1992), an abstraction principle can be designed to imply every statement whatsoever, by exploiting the "inconsistency" of the equivalence relation of co-extensionality. Consider the following scheme of abstraction principles:

$$\textbf{(HP+P)}\ \forall F \forall G \left(\#F = \#G \leftrightarrow \left(\begin{array}{c} \exists R\,(Bij\,(R,F,G)) \wedge \\ (\neg P \rightarrow (\forall x\,(Fx \leftrightarrow Gx))) \end{array} \right) \right)$$

If P is a consequence of **(HP)**, then this principle meets all the conditions for semantic success if **(HP)** does. Moreover, it is very easy to prove P on the basis of any such abstraction principle. This can be done "by reductio": if P was not true, then **(HP+P)** would be equivalent to **(BLV)**, but **(BLV)** can be shown to be false, exploiting Russell's paradox.

Insert any complex arithmetical theorem for P, such as Fermat's last theorem (FLT):[46]

$$\textbf{(HP+FLT)}\ \forall F \forall G \left(\#F = \#G \leftrightarrow \left(\begin{array}{c} \exists R\,(Bij\,(R,F,G)) \wedge \\ (\neg FLT \rightarrow (\forall x\,(Fx \leftrightarrow Gx))) \end{array} \right) \right)$$

[46] FLT says that there are no integers a, b, and c such that $a^n + b^n = c^n$, for any $n \geq 2$. It has been conjectured by Fermat in 1637. Wiles published a proof in 1995.

An epistemic agent stipulating **(HP+FLT)** will then make a stipulation meeting all the conditions for semantic success. Hence, by **(Externalism)**, the agent can come to know FLT just on the basis of stipulating **(HP+P)** and knowledge of Russell's paradox. Obviously, this should not be the case. Hence, **(Externalism)** needs to be rejected.

Both **(Internalism)** and **(Externalism)** are unacceptable for the neo-Fregean. The moral is that the proper requirement must be that the agent possesses some kind of justification in-between **(Internalism)** and **(Externalism)**.

Hale and Wright accept this argument. Their response is that one possesses a right for stipulation *by default*, which can be undermined by specific reasons to doubt that the stipulation is successful (MacBride 2003, pp. 147f, Hale & Wright 2009, p. 192). This yields the following position:

> **(Default Entitlement)** If we do not possess a sufficient reason to doubt that the conditions for **(Success)** hold, we possess an epistemic warrant for the obtaining of the conditions without having done any prior epistemic work.

The thought is that this suffices to warrantedly regard relevant stipulations as being successful, which suffices for knowledge of the truth of the stipulation in good cases (i.e. in cases where all the conditions are actually met), whilst excluding cases of easy knowledge, because in such cases there are sufficient reasons to doubt that the conditions hold.

One might interpret the above-mentioned non-arrogance condition as already containing this thought. A stipulation is non-arrogant, so the thought goes, if there is no sufficient reason committing the agent to further epistemic investigations. In the case of Jack the Ripper, for example, there is a sufficient reason to doubt that there is a single perpetrator, so that one cannot responsibly make the stipulation without having made sure that there is a single perpetrator. And maybe one can say something similar in the case of **(HP+FLT)**. In the case of **(HP)**, however, there is no sufficient reason to doubt that all relevant conditions are met. First and foremost, the reasoning that leads to Russell's paradox does not apply here, so we have no sufficient reason to doubt that **(HP)** is consistent, or so the thought goes.

So Hale and Wright already have the resources for two responses to the Shapiro-Ebert dilemma: they could either impose **(Default Entitlement)** as a further, individual condition for **(Epistemic Productiveness)**, or interpret the non-arrogance constraint as entailing **(Default Entitlement)** in relevant cases.

However, this proposal raises many questions. What exactly counts as a sufficient reason to doubt that a condition for **(Success)** is not met? For example: had Frege been justified in believing that **(BLV)** is true before he received Russell's letter, if he had regarded **(BLV)** as an implicit definition of the extension operator? And what considerations would provide a sufficient reason to doubt that **(HP)** is consistent? It is open questions like this that let Ebert and Shapiro conclude that the response is not satisfying (Ebert & Shapiro 2009, sections 6.3).

More generally: postulating warrants by default that are sufficient to underlie the epistemic workings of the **(Implicit Definition Thesis)** is one thing. Providing a clear epistemological background picture that shows how such warrants work and that they are sufficiently strong for these purposes is another. Clearly, more needs to be said here. First and foremost, we need an explication of the very notion of a warrant by default.

In sum, it transpires that there is substantial further work to be done on three fronts:

1. We have just seen that we need a clear explication of the notion of a warrant by default.

2. We need to find a set of precise and independently motivated conditions whose joint truth will ensure that a stipulation is subject to **(Effectiveness)** and **(Success)**.

3. We need to explicate how the process of knowledge-generation works in full detail. We need a story of (i) what conditions a stipulation needs to meet in order for it to be a candidate for **(Epistemic Productiveness)**; in particular (ii) what the epistemic relation between the agent making the stipulation and the conditions of semantic success needs to be; (iii) how the knowledge-generating process works in full detail and what the structure of justification is to be.

I will answer all three questions in the remainder of this thesis. As to question 1, in later work, Hale and Wright explicate the notion of *warrant by default* by Wright's notion of *entitlement* (Hale & Wright 2009, p.192).[47] The motivation, explication, and defense of an epistemic framework involving entitlements thus becomes very important for a neo-Fregean. Providing such a framework is the aim of Part II of this thesis. In chapter 7, I

[47]This move is anticipated and criticized by Ebert and Shapiro (2009, section 6.4).

will embed the neo-Fregean proposal into this epistemic framework, which yields a new, precise account of implicit definition. This account also provides answers to questions 2 and 3.

2.3.5 Hermeneutic reconstruction again

Assume that it has been established that (i) **(HP)** is subject to the **(Implicit Definition Thesis)**; (ii) that we can use logical reasoning to extend apriori knowledge; (iii) that the number operator can be interpreted as carrying ontological commitments to mind-independent abstract objects. The neo-Fregean then still has to establish **(Same Subject Matter)**, i.e. that he or she has really reconstruced knowledge of *arithmetic*. For this, it does not suffice just to note that FA relatively interprets second-order PA. In particular, it still needs to be argued that the meanings of the terms of FA are the same as the meanings of our ordinary arithmetical terms. First and foremost: that the referents of number terms of FA are the same objects as the referents of the number terms in ordinary mathematical discourse (i.e. mathematical, scientific, and everyday discourse of sufficiently competent speakers which involves number terms). Only if this can be established will the neo-Fregean account meet the demands of **(Weak Hermeneutic Reconstruction)** and the **(Arithmetical Knowledge)** constraint (for a definition of these constraints, see 1.2.2).

The same holds at the level of definitions, and for **(HP)** in particular. Even if **(HP)** can count as a definition proper, the question is whether it is a definition that merely lets us latch onto an entirely new subject matter that behaves like numbers in relevant ways (this claim would amount to **(Revolutionary Reconstruction)**), whether it is a means to talk about the numbers we all know and love (this claim would would amount to **(Weak Hermeneutic Reconstruction)**), or whether it is an explanation or explication of the ordinary *concept* of number (this claim would amount to **(Strong Hermeneutic Reconstruction)**).

Hale and Wright often make a weak hermeneutic claim. It is implicit in their aim to vindicate our knowledge of the existence of *numbers*, where "numbers" is to be understood in a pre-theoretical sense, or to vindicate arithmetic, where arithmetic is the theory about the numbers we all know and love (Wright 1999, p.322). However, they also think that **(HP)** explains the nature of numbers (Wright 1999, p. 320). And they also say that it is a central aspect of their proposal that "the concept of (cardinal) number can be explained (...) by Hume's Principle" (Hale & Wright 2001*a*, p.1). This suggests that they might have in mind a strong hermeneutic claim as well (see also MacBride 2003, footnote 10). Be that as it may. How do Hale and Wright argue for hermeneutic reconstruction?

Reference Supervenes on Use MacBride (2003) reconstructs a neo-Fregean argument for **(Weak Hermeneutic Reconstruction)**. Building on a remark in Wright's work (Wright 1999, p. 322), MacBride contends that the neo-Fregean argues for **(Weak Hermeneutic Reconstruction)** on the basis of the following principle (MacBride 2003, p.109):

> **(Reference Supervenes on Use)** Whenever two different terms (in two different theories) share the same pattern of use, the terms have the same referent.

What would establish that the terms of FA share a pattern of use with relevant ordinary mathematical terms? The obvious thought is that a technical reconstruction theorem such as the claim that (second-order) PA is relatively interpreted by FA, suffices for a claim to the effect that the number terms of FA share a pattern of use with the number terms of (second-order) PA, and we can assume that second-order PA is about the numbers we all know and love. However, there are problems with this approach:

- Set theory interprets arithmetic but numbers are not sets. What is worse: there are many different interpretations of numbers as sets. And they cannot all have the same reference.
- FA relatively interprets theories that it does not reconstruct in the relevant sense, such as the theories of real numbers (Walsh 2010, p. 16).

Frege's Constraint and applicability Wright (1999) argues that one cannot decide the question of sameness of meaning just by considering pure theories, for similar reasons as those sketched above. Rather, so Wright:

> Any doubt on the point [of whether the terms of both reconstruction and reconstructed theory share the same pattern of use] has to concern whether the definition of the arithmetical primitives which Frege offers, based on Hume's

Principle and the logical notions, are adequate to the ordinary applications of arithmetic. (Wright 1999, p. 322)

On this basis, Hale and Wright impose an additional condition that abstraction principles have to meet in order to be candidates for hermeneutic reconstruction — *Frege's Constraint*. Wright (2000, section 2), extracts it from *GG*, §159:

> **(Frege's Constraint)** The epistemic and semantic foundations of reconstructed mathematical theories must render their applications *immediate* by building their applications into their foundations.

This is still pretty vague, but in the case of **(HP)**, one can see how **(Frege's Constraint)** might be met. For **(HP)** makes it immediate that numbers are objects assigned to sortal concepts, whereas the same objects are assigned just in case the concepts are equinumerous. According to Frege, this captures the fundamental role of numbers in ordinary discourse (*GL* §46).

A complication needs to be discussed here.[48] The question arises whether **(Frege's Constraint)** is just a constraint on **(HP)**, or on **(HP)** and the Frege definitions. I am inclined to say that the claim should be that stipulating **(HP)** alone enables us to refer to the numbers we all know and love, in the sense that it should not be the case that what **(HP)** refers to is indeterminate until we have laid down additional definitions. I am thus inclined to say that the constraint is primarily a constraint for the relevant axioms (e.g. **(HP)**). However, it is of course also important that the right definitions are made. Only if we define the relevant terms properly, will we actually come to refer to the right objects. For example, if we define "0" as "$\#_x\,(x \neq x)$", the term "0" will refer to the number 0 we all know and love.[49] If we define "0" as "$\#_x\,(x = x)$", then "0" will not refer to a natural number, and relevant hermeneutic claims will fail.

---

[48] I am indebted to Carrie Jenkins and Andrew McGonigal for pressing this issue.

[49] Andrew McGonigal has pointed out to me that there are deviant definitions which are extensionally equivalent. For example, one might define "0" as "$\#_x\,(x \neq x \wedge x = x)$". We can prove in FA that "$\#_x\,(x \neq x) = \#_x\,(x \neq x \wedge x = x)$", so both terms refer to the same object. I think that I have to (and can) take the point. There are several co-extensive definitions for 0. In a weak hermeneutic project, getting the right referents is all that matters. However, someone concerned with a weak hermeneutic project can still take the Frege definitions as a canonical, natural, and sufficient way to get the right objects. When it comes to strong hermeneutic reconstruction, however, deviant definitions are ruled out. The deviant definition of "0" above certainly does not capture the ordinary concept of 0. The Frege definitions seem to be the best candidates for definitions capturing our ordinary arithmetical concepts, and I assume that in what follows.

How does the (true) claim that **(HP)** meets **(Frege's Constraint)** deliver the hermeneutic claim the neo-Fregean wants to establish? The Wright quote suggests extending MacBride's line of reasoning above: to use the **(Reference Supervenes on Use)** principle to establish **(Weak Hermeneutic Reconstruction)**. The idea is that we do not just take into account the use facts of pure arithmetical theories, but also the use facts regarding applications of these theories. In particular, we ask: are the patterns of use of *applied* Frege Arithmetic identical to the use facts of ordinary arithmetical terms in *applied contexts* — i.e. in the sciences and in everyday reasoning? Wright thinks that it is plausible that **(HP)** does the job because it meets **(Frege's Constraint)** (Wright 1999, p.322), and I agree. Of course, more needs to be said here. A full argument requires a precisification of what the relevant contexts are, and how exactly the sameness of patterns of use is to be determined. This merits further research, but is beyond the scope of this thesis.

A second line of argument rests on *metaphysical* premises. The claim is that, if **(HP)** meets **(Frege's Constraint)**, then **(HP)** will sufficiently explain the *nature of* cardinal numbers. Neo-Fregeans clearly endorse this claim about **(HP)** (Wright 2000, pp. 319f). For example, the neo-Fregeans agree with Frege (GL §46) that numbers essentially belong to concepts. If the nature of the objects introduced by **(HP)** is sufficiently like the nature of the numbers we all know and love, this presumably entails that relevant instances of "#" refer to these numbers.

Thirdly, maybe it can be argued that, since **(HP)** meets **(Frege's Constraint)**, it provides a *conceptual analysis* of the concept of number. Heck (2000), for example, argues that a principle very similar to **(HP)** — a principle he calls HPJ — is a fundamental conceptual truth about numbers, because of relations between the principle with the practical relevance of number talk.[50]

I am optimistic with respect to all three strategies of endorsing **(Frege's Constraint)** to establish a hermeneutic claim. However, all three arguments are in need of clarification, and merit further research.

[50] However, Heck does not think that HP has the same status as HPJ.

Two worries Two general worries need to be discussed here.[51] Firstly, numbers are not only employed as cardinalities assigned to concepts, but also in counting. To what extend can **(HP)** account for this application of numbers? A fortiori, if this application is equally important, does **(HP)**'s focus on the cardinality aspect not threaten the claim that **(HP)** really meets **(Frege's Constraint)**, let alone any hermeneutic claim? The response should be that the cardinality aspect is more fundamental to the ordinary concept of number.[52]

One might object that in the process of acquiring the concept of number, counting comes before assigning cardinalities, and it must be the counting role which is conceptually prior (Heck 2000). However, I agree with Heck in that I do not think that such facts show what is conceptually prior. Young children can count without grasping the concept of number (Heck 2000, p. 197). However, one cannot possess the concept of number without understanding what it is for two concepts to be equinumerous (Heck 2000, pp. 199ff).[53] Note that I am only concerned with the ordinary concept of number that a fully competent user of number talk — in a sense that includes the reader of this thesis — possesses, and that I have assumed that there is only one such concept, so it is not a good reply to worry that there might be different concepts of number, and the made claims only apply to some of them.

The second worry concerns the logical form of ordinary arithmetic, and our epistemic access to facts about it. Clearly, even the weak hermeneutic claim will only go through if the surface grammar of ordinary number talk can be taken at face value, i.e. if ordinary number talk is object talk and ordinary number terms are singular terms. One might worry that it has not been excluded that linguistic investigations reveal that number terms are not really singular terms. This also threatens the apriority of the claim that number terms are singular terms, and thus also the apriority of the claim that the reconstructed theory meets the demands of **(Weak Hermeneutic Reconstruction)**.

I do not think, however, that the neo-Fregeans need to worry about this. Firstly,

[51] I am indebted to Carrie Jenkins and Andrew McGonigal for pressing these issues.

[52] Note that one can explain the practise of counting by appealing to equinumerosity and the Frege definitions. Suppose there are four forks on the table, and they are counted. The first fork is assigned the number 1, the second fork is assigned the number 2, and so on. As soon as one has assigned the number 4 to the last fork, one knows that the number of the forks on the table is identical to the number of numbers from 1 to 4 — namely 4 — because both concepts are in one-one-correspondence (we have assigned exactly one number to each fork on the table).

[53] However, Heck (2000, p.199) also argues that one can possess a concept of equinumerosity without possessing the concept of one-one-correspondence, although the two concepts are extensionally equivalent. Discussing this complication is beyond the scope of this thesis.

although I think it is an epistemic possibility that further linguistic investigation reveals that number terms are not singular terms, I think this claim is clearly false. Secondly, the mere empirical defeasibility of the claim that number terms are singular terms does not yet show that the claim cannot be justified apriori, if apriority is understood in a sense that allows for empirical defeasibility. In 1.2, I have made the plausible assumption that it is transparent to sufficiently competent subjects what the logical form of number talk is, in a sense that entails that claims about logical form are apriori justifiable. Here, apriority should be understood in a sense that allows for empirical defeasibility. Thirdly, even if it turns out that the hermeneutic claim cannot be justified apriori, this will not mean that the hermeneutic claim is false. It is true. Of course, it would be ideal if Hero could also come to know apriori that his reconstructed theory is about the numbers he talked about before, but this is not part of the hermeneutic claim, and it is also not a part of the **(Arithmetical Knowledge)** constraint.

2.4 Three objections to neo-Fregeanism

In addition to the question of how exactly the argument for hermeneutic reconstruction is to work, and the unclarities about its exact epistemological workings, neo-Fregeanism faces a bunch of specific objections. In this section, I briefly sketch the three most pressing ones, namely:

- The Caesar problem.
- The Bad Company objection.
- (Global epistemic) rejectionism.

I will leave it open whether the neo-Fregean responses are fully convincing. In the final chapter of this thesis, I come back to the objections and discuss them in a more general setting.

2.4.1 The Caesar problem

Frege dismissed **(HP)** as a definition of number because of the Caesar problem. To repeat: in *Grundlagen*, Frege complains that **(HP)** cannot be a proper definition of number because it only decides identity statements where both terms are of the form "$\#F$", but does not provide any guidance for deciding *mixed* identity statements such as (C) "The number of planets = Julius Caesar".

Obviously, this also poses a challenge to the neo-Fregeans. Before I can discuss Hale's and Wright's response, however, the problem needs to be disambiguated. I already mentioned that the Caesar problem admits of more than one reading. In fact, at least four readings of the problem can be identified:

- *Epistemic*: according to the epistemic reading, **(HP)** does not enable us to *acquire knowledge* of whether (C) is true or false. The question remains *evidence-transcendent* if all we can use to decide such questions is **(HP)**, or so the thought goes. Why is this a problem? There is a fact about the matter, and we should be able to know it.[54] In particular, one might think that one should be able to know all facts about numbers on the basis of the proposed foundations of arithmetic, and that **(HP)** cannot be a complete definition because it does not enable us to decide all questions regarding numbers. However, in the light of incompleteness, this is a very strong claim. A better way of putting the objection is that we have a strong intuition to the effect that the number of planets cannot possibly be identical to Julius Caesar (Heck 1997*b*, p.276). Thus, there is some knowledge of numbers that we cannot obtain on the basis of the proposed epistemological foundation. We thus might have to appeal to rational intuition, and this is unacceptable (Heck 2005, section 4).

- *Semantic*: **(HP)** does not fix the truth conditions of all identity statements involving number terms. Thus, it appears that **(HP)** cannot fully fix the meaning (reference) of "#". This raises issues with semantic indeterminacy. For example, it has been argued that this entails that the concept of cardinal number (defined by **(Numbers)** in 1.3.2) does not have a unique extension and is a *pseudo concept* (Schirn 2003, p.211).

- *Cognitive*: suppose that Evan's *Generality Constraint* is a condition for understanding a term, i.e. that it is a necessary condition for understanding a term that one understands a range of relevant contexts containing it. Moreover, suppose that statements like (C) count as relevant contexts. Under certain assumptions, it follows that **(HP)** cannot effect an understanding of "#". For example: if understanding (C)

[54] This way of putting the Caesar problem has been suggested to me by Robert Williams.

requires knowing its truth-conditions. This is because **(HP)** does not fix the truth-conditions of (C), or so the thought goes (Hale & Wright 2001*b*, pp. 341ff).

- *Metaphysical*: the definition does not fully determine the *nature of* numbers. If it did, both the epistemic and the semantic problem would not arise.

These four problems are related. The Caesar problems have been discussed extensively, and I cannot repeat the discussion here. After briefly discussing two allegedly easy solutions of the problem, I present the official neo-Fregean solution that Hale and Wright propose in their (2001*b*), which I think is promising.

Firstly, one might want to respond that no number can be identical to Caesar because numbers exist necessarily (assuming they exist at all) and Caesar does not, or because numbers are abstract objects (assuming they exist at all), and Caesar is a concrete object (for such lines of reasoning, see e.g. Hale & Wright 2001*b*, p. 366; Rosen 1993). Although this is no doubt correct, the problem also arises in cases where both kinds of objects are abstract, and exist necessarily. For example: is 0 ($= \#_x[x \neq x]$) identical to $\emptyset$? How can **(HP)** help with deciding these questions? One constraint to any solution of the Caesar problem is that the solution generalizes to such cases.

Secondly, one might want to respond that abstract-person identities are meaningless because they constitute category mistakes. Thus, so the thought goes, statements such as (C) are irrelevant, because there simply is no thought to grasp, or no fact to know. However, it is far from clear that statements like (C) are meaningless. To the contrary: I think we clearly understand (C), and there clearly is a fact about the matter. Moreover, the response raises the issue of separating bad cases from good cases. Is the statement $\#_x[x \neq x] = \emptyset$ meaningless as well? I assume that all mixed identity statements are meaningful, and turn to Hale's and Wright's official solution of the Caesar problem.

Hale and Wright (2001*b*) build on Wright's proposal in *Frege's Conception* (Wright 1983). According to Wright (1983), the root of the problem is that in order to introduce a new sortal concept — a concept of a sui generis kind of object — one does not only have to fix a *criterion of identity*, but also a *criterion of application*. Whereas the former only enables us to decide identities in which both sides are number terms, the latter enables us to decide whether any given thing is a number.[55]

The basic idea in *Frege's Conception*, which Hale and Wright adopt in their (2001*b*), is that abstraction principles fix both criteria in relevant cases, because, given some plausible metaphysical background assumptions, criteria of application are determined by criteria of identity (Hale & Wright 2001*b*, p.369). The idea is based on two thoughts. Firstly, although every sortal concept has its own specific criterion of identity, two different sorts of things can *share a criterion of identity*. That two sorts of things F and G (with the specific criteria of identity eq_F and eq_G) share a criterion of identity means that the following is a necessary conceptual truth (Hale & Wright 2001*b*, p. 391):

$$(*)\ (\forall xy)\,(x eq_F y \leftrightarrow x eq_G y)$$

F and G will share a criterion of identity if F can be subsumed under G. For example, since all tigers are animals, tiger shares a criterion of identity with animal. Now, secondly, there are maximally inclusive sorts of things: *categories*. Categories are associated with a criterion of identity that all the sortals in this category share. Thus, if two sorts of things do not share a criterion of identity, they cannot belong to the same category.

It can be difficult to determine whether two sorts of things share a criterion of identity. However, it is very plausible that numbers and people do not share such a criterion (Hale & Wright 2001*b*, p. 393). Thus, they cannot belong to the same category. This yields a solution to the original Caesar problem, which takes the form of a dilemma. Either categories never overlap, or they sometimes overlap. In the former case, numbers cannot be persons, so the original Caesar problem is solved. In the latter case, the problem becomes a problem for singular terms in general, and thus the problem cannot be used as a specific objection against using abstraction principles to introduce new sorts of objects. For no matter how new kinds of things were introduced, cross-categorial identities would be indeterminate (Hale & Wright 2001*b*, p. 396).

Hale's and Wright's solution is general enough to apply to various generalizations and complications of the Caesar problem. However, it does not provide a lot of guidance in particular cases. For example, what has been said does not help us when it comes to

[55]This suggests an epistemic reading of "criterion of identity" and "criterion of application"; but note that just as there is an epistemic and a semantic version of the Caesar problem, there is an epistemic and a semantic reading of these criteria. We could either conceive of them as criteria providing the means to decide certain questions, or as criteria that are in some sense fundamental to sortal concepts.

deciding whether 0 is identical to $\emptyset$. For it does not help us with the question of whether sets and numbers belong to the same category. To answer this question, we need to be able to answer the question of whether numbers and sets share a criterion of identity. This problem generalizes. We need a procedure to decide whether the abstracts introduced by different abstraction principles belong to the same category or not. The neo-Fregean faces the following task:

> **(Categorization of Abstracts)** The neo-Fregean needs to find a principled and metaphysically motivated partition into categories of all kinds of objects that can be introduced by abstraction principles.

This task is very difficult, at least if it is a constraint that the partition meets pretheoretic intuitions. There are cases in which it is plausible that two sorts of abstracts, introduced by different abstraction principles, belong to the same category or even are identical. For example: the numbers introduced by Hume's Principle and the numbers introduced by Heck's *Finite Hume's Principle.*[56] However, there are also cases in which it is plausible that two sorts of abstracts belong to different categories. For example, numbers and sets are too different to be in the same category. All conditions for identifying intra-abstract identities proposed so far fail to accomodate one of these cases.

As an example, consider the simple view that there are as many categories as equivalence relations. This view cannot accomodate the intuition that the "natural number segments" of the numbers introduced by Hume's Principle and Finite Hume's Principle are the same (see also Cook & Ebert 2005, 2). On the other hand, views that accomodate this intuition will identify kinds of abstracts that should not be identified. As an example, consider the view that all the abstracts introduced by an abstraction principle P are identical to some of the abstracts introduced by another abstraction principle Q if Q relatively interprets P. This will identify the numbers introduced by **(HP)** and Finite Hume's Principle, but it will also imply that numbers are sets (where sets are introduced by a sufficiently strong abstraction principle for sets). Even if one thinks that Hale's and Wright's solution is acceptable in principle, these issues certainly require further work.

[56]Finite Hume's Principle is the abstraction principle that says that the number of the Fs is identical to the number of the Gs if F and G are infinite, or F and G can be put into one-one correspondence. This equivalence relation is not identical to simple one-one correspondence. A similar abstraction principle was introduced and discussed by Heck (1997*a*).

2.4.2 The Bad Company objection

The *Bad Company objection* can be stated as follows: there are many abstraction principles — statements having the same form as **(HP)** — which cannot have the special epistemic and semantic status that **(HP)** is supposed to have. This casts doubt on the neo-Fregean justification of **(HP)**, at least when this justification is to be based on the fact that **(HP)** is an abstraction principle.

For example, Frege's **(BLV)** — a second-order abstraction principle just like **(HP)** — is inconsistent. Thus, no appropriate function from concepts to objects can be assigned to the extension operator, and **(BLV)** cannot possibly be true. The stipulation cannot be semantically successful, and it cannot be known on the basis of stipulation. So the neo-Fregean needs to be able to make it plausible that relevant abstraction principles such as **(HP)** are different. In general, the neo-Fregean needs to separate the good abstraction principles from the bad. This needs to be done in a principled, non ad-hoc way, which turns out to be a substantial task. For example, it is not enough to just demand consistency. For there are consistent, but mutually inconsistent abstraction principles (Boolos 1990, Wright 1999). I discuss these problems in some more detail below.

First, it makes sense to distinguish two programmes, and two corresponding Bad Company objections. This will bring to light the full scope of the problem.

Fine's programme Although Hale and Wright are somewhat opposed to such an interpretation (Hale & Wright 2009, §4), neo-Fregeanism can be conceived of as some kind of maximalism. Consider the following meta-ontological view:

(Maximalism) Whatever can exist does exist.

Applied to abstraction principles, the idea becomes the following:

(Maximalism about Abstraction) Every abstraction principle that can be true is true.

This provides a setting for the Bad Company problem. What the objection shows, so the thought goes, is that not *every* abstraction principle can be true. Thus, the abstractionist needs to sort out a maximal collection of abstraction principles which can be jointly true. Every member of this collection can then be regarded as true, in accord with **(Maximalism about Abstraction)**.

To sort out such a maximal collection is Fine's agenda in his (2002), so let me dub this task "Fine's programme":

> **(Fine's Programme)** Find a maximal collection of abstraction principles which can be jointly true.

This is a difficult *mathematical* task. One does not only have to describe a collection of abstraction principles which could be true, but one has to describe a collection of abstraction principles which could be *true together* with all other true abstraction principles, and one also has to argue that the collection one has identified is maximal. It is not immediately clear how such a criterion should look like, and even whether there is a single maximal set of abstraction principles.

Moreover, it is important for the neo-Fregean to show that the collection will include enough abstraction principles to deliver a sufficient amount of mathematics (i.e. most of today's mathematics). Whether this is so remains an open (mathematical) question, and only carrying out **(Fine's Programme)** will deliver an answer.

There are two branches of Fine's project: the neo-Fregean branch, and Fine's own branch. Here, I focus on the latter. So which abstraction principles are the good ones, and which are the bad ones?

Let us begin with the famous infamous **(BLV)**. In the neo-Fregean framework, it leads to Russell's Paradox. It is worth noting, however, that it only leads to inconsistency given a sufficient amount of second-order comprehension. This is because the derivation rests on the existence of the *Russell Property*, i.e. the property of being a set which does not have itself as a member: this property includes both the notion of set, and quantifies over properties — so it is actually impredicative in two different ways.[57]

Can we avoid the inconsistency by weakening second-order comprehension appropriately? The answer is yes (Heck 1996), but weakening comprehension in a way that blocks the derivation of the inconsistency also blocks Frege's proof of the Peano axioms from

[57]With the help of the Russell Property, we can define the "Russell Set", i.e. the set of all sets which are not members of themselves. And now inconsistency looms: the Russell set is a member of itself if and only if it is not a member of itself.

(HP) (Linnebo 2004).[58]

So **(BLV)** remains a bad companion. The neo-Fregean needs to restrict the collection of good abstraction principles, and a first obvious criterion is consistency. However, this is not enough. It turns out that there are pairs of consistent, but mutually unsatisfiable abstraction principles. And mutually unsatisfiable principles can not both be *true*.

There are many examples of such pairs. We already encountered Heck's trick to construe abstraction principles that imply every statement whatsoever, by exploiting the equivalence relation of co-extensionality. Consider the following scheme (where ϕ does not include any free occurrences of F or G):

(AP+Phi) $\forall FG\,(\Sigma F = \Sigma G \leftrightarrow (\phi \vee \forall x\,(Fx \leftrightarrow Gx)))$

It is easy to see that **(AP+Phi)** implies ϕ. This is because the abstraction principle is equivalent to **(BLV)** in case ϕ does not hold.

How does that bear on the Bad Company problem? Well, pretty straightforwardly. Just choose two satisfiable but not jointly satisfiable second-order sentences ϕ_1 and ϕ_2, and use Heck's technique to obtain two consistent but mutually unsatisfiable abstraction principles. As an example, consider the second-order versions of "the domain is finite" and "the domain is infinite".[59]

Neo-Fregeans might want to rule out such *paradox exploitative* abstraction principles just because they are paradox exploitative. Wright suggested this move (Wright 1999, Appendix 1). However, there are examples which do not exploit paradoxes, so nothing rests on Heck's technique.

For example: Frege's Theorem implies that **(HP)** is only satisfiable in infinite domains. However, there are consistent abstraction principles which are only satisfiable on finite domains. Two well-known examples are Boolos' *Parity Principle* (Boolos 1990, pp. 214f) and Wright's *Nuisance Principle* (Wright 1999, p. 318). Let me use the Nuisance Principle as an example. It is the following second-order abstraction principle (where $Fin_x\,(\phi[x])$ stands for the second-order sentence that Dedekind-infinitely many objects fall under $\phi[x]$):

[58] I am aware of the fact that there are different approaches to abstraction which do not impose restrictions on abstraction principles, but restrict what properties there are (Linnebo 2007). I cannot discuss such approaches here.

[59] That the domain is (Dedekind-)infinite can be expressed as the statement saying there is a bijection from the domain to a proper subset of it. For a discussion of the expressibility of notions of infinity in second-order logic, consult (Shapiro 1991, 5.1).

(NP)

$$\forall F \forall G \left(\begin{array}{r} Nuis\,(F) = Nuis\,(G) \leftrightarrow \\ \left(Fin_x\left([Fx \wedge \neg Gx]_x\right) \vee Fin_x\left([Gx \wedge \neg Fx]_x\right)\right) \end{array} \right)$$

Model-theoretic reasoning shows that **(NP)** only has finite models (1999, pp. 318f). Thus, **(HP)** and **(NP)** cannot possibly be true together. One of these principles must be false. Thus, the neo-Fregean needs to say more about which abstraction principles are the good ones.

It has been proposed that principles like **(NP)** should be ruled out on the ground that they limit the size of the domain. Surely, an abstraction principle should only introduce a new sort of abstract entity, and never imply anything about the objects already introduced or understood — at least no statements that are expressible in the language without the new abstraction operator. Any principle that implies that there are only finitely many objects violates this constraint.

There is a technical notion capturing this idea: Field's notion of conservativeness (Field 1989). To express the notion, we need to restrict quantifiers to the old domain — the domain minus the objects the newly introduced abstracts refer to. To do this, we can define a restriction of second-order formulas (Weir 2003, p. 21). For any first or second order formula ϕ, and a predicate $R\,(x)$, the *restricted formula* ϕ^R is the result of altering quantified statements as follows ($\phi^R \equiv \phi$ if ϕ does not contain any quantifiers):

- $\forall x\,(\psi\,(x))^R$ is $\forall x\,\left(R\,(x) \rightarrow \psi\,(x)^R\right)$
- $\exists x\,(\psi\,(x))^R$ is $\exists x\,\left(R\,(x) \wedge \psi\,(x)^R\right)$
- $\forall X\,(\psi\,(X))^R$ becomes $\forall X\,\left((\forall x\,(Xx \rightarrow Rx)) \rightarrow \psi\,(X)^R\right)$
- Thus $\forall x\,(\psi\,(x))^R$ becomes $\exists X\,\left((\forall x\,(Xx \rightarrow Rx)) \wedge \psi\,(X)^R\right)$

We then define a notion of conservativeness as follows:

(Conservativeness)

Let T be any theory in a language L_T and AP be the abstraction principle for an abstraction operator Σ which is formulated in the language L_{T+AP}, which is L_T plus the abstraction operator. Let $Abstr(x) \equiv \exists y\,(x = \Sigma y)$ if AP is a first-order abstraction principle, and $Abstr(x) \equiv \neg\exists F\,(x = \Sigma F)$ in the second-order case. Then AP is a *conservative abstraction* over T if and only if for every formula ϕ it holds that

$$\left((T+A)^{\neg Abstr(x)} \models \phi^{\neg Abstr(x)}\right) \Rightarrow (T \models \phi)$$

Note that "$\models$" stands for *semantic consequence*. Of course, we might also define conservativeness using a syntactic notion of consequence. However, for various reasons, the semantic notion is preferable (Weir 2003, pp. 22f; Cook 2009, p. 352) .

There is another notion which is turns out to be equivalent to conservativeness in the case of abstraction principles — the notion of an abstraction principle being unbounded. An abstraction principle is unbounded just in case it meets the following condition: if it has a model with a domain of size κ, it has a model with a domain of size $\epsilon > \kappa$ as well. Weir shows that all unbounded abstraction principle are conservative (Weir 2003, p. 23). Since **(HP)** has models in all infinite cardinalities (assuming ZFC, see Boolos 1987), **(HP)** is unbounded, hence conservative. Principles like **(NP)** and the parity principle, however, are not unbounded.

Does this complete the task? No: Weir (2003) shows that there are conservative but mutually unsatisfiable abstraction principles. One reason to see this is to note that one can choose formulas for Heck's **(AP+Phi)** that imply that the universe is of a certain size *type*, but which does not impose an upper bound on the size of the domain. Since there are incompatible size types, there are unbounded (hence conservative), but mutually insatisfiable abstraction principles. Also, note that there are such principles which are not paradox-exploitative: Weir's "Distraction Principles" (Weir 2003, p.17): again, nothing substantial depends on Heck's technique.

I focus on Heck's technique for simplicity. For ϕ, we use "the domain is of the size of a successor cardinal" and "the domain is of the size of a limit cardinal" (both are expressible in second-order logic). No cardinal can be a successor cardinal and a limit cardinal, but for any successor cardinal or limit cardinal, there is a bigger one. Hence, the respective versions of **(AP+Phi)** can be shown to be unbounded and conservative. They are nevertheless incompatible.

Weir suggests a solution on behalf of the neo-Fregeans, based on a new technical notion. One can intuitively arrive at it through reflecting on what the they might want in the end.

A natural proposal after what has been said is that the neo-Fregean looks for a maximal collection of jointly satisfiable, conservative abstraction principles. Now what if you simply demand that? The suggestion is that the collection of good abstraction principles just is the collection of those abstraction principles which are consistent with every conservative abstraction principle. Weir (2003) calls such abstraction principles *irenic*.

Indeed, it turns out that this solution works for the size-type restricting principles and includes **(HP)**. Moreover, it turns out that there is another, equivalent notion which expresses a constraint in terms of the cardinality of models. Say an abstraction principle is *stable* if and only if for some cardinal κ, it is satisfiable in models of all and only cardinalities of size equal to or greater than κ. Using ZFC, Weir shows that an abstraction principle is stable if and only if it is irenic (Weir 2003, pp. 32f). Using ZFC, **(HP)** can be shown to be stable, since it then can be shown to have models of every infinite cardinality. Thus, using ZFC, we can show that **(HP)** irenic. Moreover, the size-type restricting principles are of course not irenic. Finally, every set of irenic principles can be shown to be satisfiable and hence to be consistent as well (Cook 2009, pp. 354f). This is the status quo of at least one branch of the neo-Fregean version of **(Fine's Programme)**.

How much mathematics can we obtain with irenic / stable abstraction principles? Weir (2003) idenitifies stable sets of abstraction principles which deliver sets of numbers, sets of sets of numbers, etc. In fact, one can obtain most of mathematics with stable set theory. However, one cannot obtain full-fledged set theory, but only a "slice of the cumulative hierarchy" (Weir 2003, p. 26). For instance, it cannot be shown that for every object, there is a unit class which is distinct from this object. Moreover, Weir's theory does not allow for sets of urelements other than numbers. There are weak forms of set-theories with urelements which are stable. Uzquiano gives the example of the theory ZCU2, which is Zermelo set-theory with "countable replacement". This theory "might suffice to recapture real analysis, functional analysis, complex analysis and most of ordinary mathematics" (Uzquiano 2009, p. 14), so there is room for optimism. If suitable abstraction principles can be found, a lot of mathematics can be reconstructed.

Still, there are many complications with the irenicity / stability criterion. There are arguments to the effect that it cannot be a necessary condition (Cook 2009; Linnebo & Uzquiano 2009; Weir 2003, p.35). And what is worse, whether an abstraction principle meets the criterion depends on one's meta-theory. This leads to the problem that, depending on one's initial choice of a meta-theory, one will accept different abstraction principles as good (Weir 2003, section 7).

This points towards a more general problem: whatever the outcome of the current mathematical investigations will be, they cannot in principle do full justice to the programme of distinguishing good abstraction principles from the bad, because the neo-Fregean programme is more than just a mathematical programme. Let me explain.

The neo-Fregean programme The neo-Fregean claims that (good) abstraction principles have a special epistemic and semantic status, as opposed to some "special mathematical status". We should distinguish **(Fine's Programme)** from the following programme:

> **(Neo-Fregean Programme)** Find a collection of abstraction principles which fix the meaning of the respective abstraction operators, and which can be known apriori on the basis of meta-linguistic stipulation.

At least prima facie, it is a different matter entirely to spell out the criteria an abstraction principle has to meet in order to be a member of this "neo-Fregean collection" of good abstraction principles. Thus, we obtain two different projects of sorting out abstraction principles. Whereas **(Fine's Programme)** is first and foremost a mathematical project, the **(Neo-Fregean Programme)** is a philosophical project. To both projects, there corresponds a specific Bad Company objection.

All bad companions for **(Fine's Programme)** will be bad companions for the **(Neo-Fregean Programme)**, because knowledge requires truth, and Fine's programme sorts out all true abstraction principles. However, there may be bad companions specific to the **(Neo-Fregean Programme)**. In particular, there may be abstraction principles which are good from a mathematical point of view, but bad from a philosophical point of view.

The criteria that Hale and Wright suggest for the **(Neo-Fregean Programme)** are just the conditions for the "Traditional Connection" discussed in 2.3.3. We can see how they are motivated by different kinds of bad companions. Consider the non-arrogance constraint, for example. It is meant to rule out stipulations (and abstraction principles) that are epistemically irresponsible. Such abstraction principles might still be good from a mathematical point of view.

As an example, reconsider **(HP+FLT)** from 2.3.4. We know that FLT is true. So **(HP+FLT)** has exactly the same relevant mathematical properties as **(HP)**: if **(HP)** is a good abstraction principle in the sense of **(Fine's Programme)**, so is **(HP+FLT)**. However, a stipulation of **(HP+FLT)** seems to be irresponsible. We cannot prove FLT that easily, or so the thought goes.[60] So we have an example of a bad companion for the **(Neo-Fregean Programme)**, which is not a bad companion for Fine's Programme.

How do the programmes relate? The question arises of how both programmes relate. Presumably, the most desirable result for a neo-Fregean would be that the criteria of Fine's project can be regarded as a systematization of the criteria of the traditional connection for abstraction principles.[61]

Let me explain. Say an abstraction principle is "tenable" just in case it is good in the sense of **(Fine's Programme)**, and say an abstraction principle is "stipulable" just in case it is good in the sense of the **(Neo-Fregean Programme)**. I assume that the neo-Fregean would like to get as much knowledge by abstraction as possible. The maximum is the collection of all the tenable abstractions. Thus, the optimal result for the neo-Fregean would be that the tenable abstractions are the knowable ones.

Let us assume that tenability is irenicity. It would be a nice result if the single criterion of irenicity sufficed for stipulability. Unfortunately, in the light of counter-examples such as irresponsible but irenic stipulations, this result appears impossible.

However, one might think that something similar but weaker can be achieved, namely that we can show that possessing a warrant for stability is sufficient for stipulability. I will come back to this question in 7.2, and it will become apparent that even this weaker connection is problematic.

2.4.3 Epistemic rejectionism

Epistemic rejectionism is the third big objection to the neo-Fregean programme. The term "rejectionism" has been coined by MacBride (2003), but the objection itself has a long history. The upshot is that one cannot come to know **(HP)** as easily as Hale and Wright envisage, because a stipulation of **(HP)** makes substantial demands on the world,

[60] Another version of the objection would not be epistemic, but semantic: that we should not build it into our concept of number that FLT is true.

[61] I owe this idea to a discussion with Robert Williams.

and one is epistemically irresponsible if one did not make sure that these demands can be met prior to making the stipulation.[62] Here are two examples from the literature:

- Satisfiability: Boolos complains that, before we can legitimately regard **(HP)** as being true by virtue of stipulation, we *first* need to *make sure* that HP is *satisfiable*, for "what kind of guarantee do we have, why should we believe, that there is a function that maps concepts to objects in the way that the denotation of the octothorpe [#] does if HP is true?" (Boolos 1997, p. 306).

- Conditionalization: Field (1984*b*, p.661) argues that **(HP)** cannot be a definition proper because it has existential commitments. What can be regarded as a definition proper, so the thought goes, is the following *conditionalized* version of **(HP)**: *if* numbers exist, then **(HP)**. However, this definition does not allow for a derivation of the Peano axioms.

Both objections are intimately related. By conditionalizing the stipulation, so the thought goes, one removes the need to make sure that **(HP)** is satisfiable. However, one cannot infer the existence of numbers from the stipulation anymore. In order to do that, one first has to ensure that numbers exist, which amounts to ensuring that **(HP)** is satisfiable.

Hale and Wright discuss these objections at length and at several occasions (Wright 1990, 1999, Hale & Wright 2000, 2009). I cannot go through the whole dialectic here — the responses are notoriously hard to assess — but just carve out what I take to be the most important points. Two responses are particularly important.

"Ought implies can" Suppose a non-defective epistemic agent (call him "Hero") does not yet have the concept of number and is about to stipulate **(HP)** to define "#". According to the rejectionist, Hero needs to make sure that "#" refers and that there are numbers before he can warrantedly regard the stipulation as true, or before Hero can infer **(HP)** from a conditionalized stipulation.

However, how could Hero decide the question of whether numbers exist or not? It was **(HP)** which was supposed to provide sufficient conditions for numbers being identical (and hence for numbers to exist). It seems that the rejectionist presupposes that there are

[62] There is also a "semantic version" of rejectionism (Ebert 2005*a*, section 4.1). I can only focus on the epistemic version here.

independent means to decide the question, and the neo-Fregean denies this (Wright 1999, pp. 311f;Hale & Wright 2000, p.143;MacBride 2003, p.124).

MacBride (2003, p.125) thinks that the rejectionist should not retreat. For example, so the thought goes, one might think that indispensability considerations provide independent means to decide the question of whether numbers exist or not. This branch of the dialectic seems to end up in a standoff.

Content recarving The second response employs Frege's notion of content recarving, i.e. the notorious idea, outlined in *Grundlagen* §64, that one can recarve the content of an equivalence relation as an identity, and thereby obtain a new sortal concept.

Applied to the neo-Fregean case, the idea is that the states of affairs expressed by instances of the right-hand side (RHS) of **(HP)** are exactly the same as the states of affairs expressed by the corresponding to instances of the left-hand side (LHS) of **(HP)**, only "carved up" (expressed) in different ways (Hale 1997). In other words: there is no substantial gap between both sides of the biconditional, and thus there cannot be any substantial further epistemic obligations when stipulating **(HP)**.

This involves two claims: a metaphysical claim, and an epistemological claim.

The metaphysical claim is that, ceteris paribus, states of affairs underlying equivalence relations can also be expressed as identities. The ceteris paribus clause is crucial. Recarving is not possible in cases like **(BLV)**. The idea must be that content can be recarved in all cases where this is as much as possible. The epistemological claim is that this shows that, as long as one does not have any reason to believe that a particular case is a bad case, one can warrantedly assume that content is recarved, and thus that the stipulation is true, without being irresponsible.

There is much more to say, and in 7.2 I will discuss this idea further. Here, I close with some notes on the metaphysical claim. For prima facie, there seems to be a gap between the states of affairs expressed by instances of both sides of **(HP)**. In particular, it appears that the truth of instances of the RHS does not require the existence of numbers. In other words: the state of affairs that instances of the RHS express do not seem to be identical to state of affairs that instances of the LHS express.

If this was the case, then one would indeed need antecedent warrants that justify the transition from instances of the RHS to instances of the LHS. But that the states of affairs are different is precisely what the neo-Fregeans deny.

Of course, this imposes some constraints on the neo-Fregean conception of states of affairs. Since instances of the LHS can only be true if the number terms refer — remember that the neo-Fregean needs this linguistic priority thesis for their Platonism — instances of the RHS are already committed to numbers. Thus, it cannot be required that state of affairs are transparent in the sense that their ontological commitments are transparent. For example: Hero might grasp the concept of one-one correspondence, and know that a certain concept stands in one-one correspondence with another, without it thereby being transparent to Hero that the obtaining of this state of affairs also involves the identity of two numbers (and hence the existence of a number). Indeed, Hale and Wright reject a notion of states of affairs that meets such a ***transparency principle*** (Hale & Wright 2009, p.189).

MacBride argues that this entails that the neo-Fregean has to hold that states of affairs are structured by language, as opposed to an external world, and that this commits Hale and Wright to some kind of anti-realism (MacBride 2003, pp.126f). Hale and Wright maintain the view that neo-Fregeanism is a kind of Platonism, and hence some kind of realism, but they also admit that the relationship between the (meta-)metaphysics of content recarving and anti-realism requires further research (Hale & Wright 2009, p.209).

2.5 Intermediate conclusion

Neo-Fregeanism is a very attractive programme, because it promises to meet all the desiderata laid down in the last chapter:

- Neo-Fregeanism is designed to meet **(Arithmetical Platonism)**, although there is the worry that the neo-Fregean response to the rejectionist worry might push the neo-Fregean towards anti-realism.

- Given the **(Implicit Definition Thesis)** withstands critical scrutiny, the neo-Fregean can accounts for **(Arithmetical Knowledge)**, although the neo-Fregean use of **(Frege's Constraint)** merits explication. Clearly, the neo-Fregean would also account for **(Arithmetical Foundationalism)** and the **(Apriority Constraint)**.

- The neo-Fregeans have the resources to account for **(Same Source)**, by applying the **(Implicit Definition Thesis)** uniformly to mathematics and logic. However, it is not clear how it is to be applied to the logical case, so more work needs to be done here.

- Presumably, the neo-Fregeans use abstraction principles that meet **(Frege's Constraint)**, so one can be optimistic about the **(Applicability Constraint)** as well. If the application of mathematical theories is immediate in the neo-Fregean foundations, it will be possible to account for their applicability.

The major obstacle is that it is unclear how exactly the **(Implicit Definition Thesis)** is to work. It is unclear what exactly the conditions for the semantic success of implicit definitions are, and what the structure of justification is to be. In any case, we need to be able to account for appropriate warrants for the conditions, whatever they are. A dilemma by Shapiro and Ebert shows that these warrants must be of a special kind: they need to enable us to be epistemically responsible stipulators, but it cannot be a requirement that we are able to *prove* them. A good candidate for warrants for these conditions might be Wright's "entitlements". However, this notion is embedded in a more general internalist epistemic framework, which needs motivation and investigation. Moreover, one needs to say how exactly the neo-Fregean proposal can be embedded in such a framework.

To fill these gaps is the ultimate aim of this thesis. In Part II, I discuss a general epistemological framework which includes entitlements, and I come back to neo-Fregeanism in Part III. In the last chapter of the thesis, I will not only close the mentioned epistemological gaps, but my strategy to close these gaps will also shed light on the three major objections to neo-Fregeanism.

Part II

An epistemological framework

3 Internalism

In this chapter, I begin with sketching the epistemological framework in which I shall eventually embed my neo-Fregean account of implicit definition. The framework rests on three tenets — internalism, non-evidential warrant at the basic level, and transmission-failure — which are to be discussed in the three chapters of this part of the thesis. I begin with internalism.

3.1 Motivating the distinction

In contemporary epistemology, there is a fundamental divide between so-called *internalistic* and *externalistic* conceptions of knowledge and justification. Whereas internalists emphasize the contribution epistemic agents make to the epistemic value of their doxastic attitudes, externalists focus on the contribution the *external world* makes to them. The origin of the debate lies in the conceptual analysis of knowledge (and justification). However, in recent years, the divide has become increasingly important within discussions of scepticism. I begin by examining how both discussions give rise to the divide. After that, I draw some distinctions and argue for a particular version of internalism.

3.1.1 Analysing knowledge and justification

Are there non-trivial conditions $C_1, C_2, ...$ such that, necessarily, S knows that p if and only if $C_1, C_2, ...$? Surely, that S knows that p entails that S holds a *true belief* that p. Many believe that for a belief to be knowledgeable, it must also be justified, in a sense of justification which entails the possession of evidence (Gettier 1963, p.121). However, it is common ground that Gettier (1963) has shown that these three conditions are not sufficient for knowledge.

Gettier cases are cases in which a subject S acquires a justified true belief that p, but in which we have the intuition that S does not know p. As an example, suppose Smith applies for a job that Jones applies for as well. During the time they are waiting for the final interview, Smith sees Jones counting the 10 coins in his pocket. Smith acquires a justified belief that Jones has 10 coins in his pocket. Moreover, suppose that Smith has good evidence to the effect that Jones will get the job (he might have overheard a relevant conversation). From this information, Smith infers that the person who gets the job has

10 coins in his pocket. Smith thereby acquires a justified belief. However, it turns out that Smith gets the job. Moreover, by sheer coincidence, Smith also has 10 coins in his pocket. Clearly, Smith has a justified, true belief that the person who gets the job has 10 coins in his pocket. However, intuitively, this belief cannot count as knowledge, because it is true by sheer luck.

How does that bear on the internalism vs. externalism debate? According to Goldman (1967), Gettier cases show that it cannot *only* be the internal states of epistemic agents which render true belief justified (and hence knowledgeable). Rather, one has to add the condition that the belief is caused in an appropriate way.[63] Clearly, so the thought goes, in the example above Smith's belief is not caused in the appropriate way.

Unfortunately, causal theorists face problems with epistemic luck as well. Suppose that Hero travels around the countryside, happens to see a barn, and correctly forms the belief that there is a barn in front of him. The belief forming process is perfectly normal. Hero's belief is caused in an appropriate way. According to the causal account, this belief counts as knowledge. However, suppose it turns out that Hero just entered barn facade county, which almost entirely consists of fake barns, and that Hero looks at the only real barn in the area by sheer coincidence. In this case, our intuitive verdict is that Hero does not know that there is a barn in front of him, and this is in tension with what the causal account predicts (Goldman 1967, p.773).

To be sure: there are externalist responses to such cases (see e.g. Goldman 1979; Nozick 1988), but I cannot go into the dialectic here. In any case, many theories of knowledge have been proposed — externalistic, internalistic, and mixed — each of them trying to handle new alleged counter-examples to the analysis of knowledge. Recently, Williamson (2000) argued that all this suggests that it is a mistake to think that the notion of knowledge can be analysed into a set of necessary and sufficient conditions at all.

Goldman (2009, p. 309) points out that, although the origin of the internalism vs. externalism debate lies in the question of how to analyse knowledge, the debate quickly focused on the nature of (epistemic) justification in general: the question arose of whether the notion should be analysed in externalistic or internalistic terms (or both). For example, Goldman (1979) suggested that our notion of justified belief should be analysed as having

[63]Obviously, the phrase "caused in the appropriate way" is in need to clarification. However, for the purposes of this chapter, the informal characterization suffices.

purely externalist application conditions, namely that the belief has to be formed by a de facto reliable belief-forming method. This account of justification is known as *process reliabilism*. On the other end of the spectrum, there are purely internalist accounts. For example, it has been argued that justificatory status *supervenes on* mental states (Conee & Feldman 2001). Chisholm (1977) even argued one needs to possess *reflective access* to the grounds of one's justification in order to count as justified.

3.1.2 Scepticism

Another epistemological debate in which a distinction between internalistic and externalistic notions of justification has become particularly relevant is the debate on scepticism. *Closure scepticism* is a type of scepticism that purports to show that we cannot acquire (or possess) justification for ordinary external-world beliefs, such as the belief that one has two hands. One version of it rests on the following two premises:

(Closure) If p and $p \to q$ are justifiable, then q is justifiable.

And:

(Impossibility) There are scenarios whose non-obtaining is entailed by ordinary external world beliefs, but whose non-obtaining is impossible to justify.

(Closure) is very plausible. Assume that p and $p \to q$ are justifiable. This means that it is possible to acquire justified beliefs that p and $p \to q$. On this basis, one can justifiably infer q by Modus Ponens. Thus, it is possible to acquire a justified belief that q.

In order to establish **(Impossibility)**, closure sceptics often make use of so-called *Cartesian scenarios* — metaphysically or even physically possible setups of the world in which a subject S would have exactly the same experiential seemings, but in which these seemings are massively misleading. As an example, consider the *brain-in-a-vat scenario* (BIV scenario), in which S's brain is removed from its body by mad scientists, and then envatted to henceworth be fed with coherent inputs emulating a normal environment.

Closure sceptics argue that it is impossible to acquire a justified belief that the BIV scenario does *not* obtain, because there is no possible evidence that could be used to distinguish the BIV scenario from the normal environment. After all, if we were in the BIV scenario, everything would still *seem* normal.

(Closure) and **(Impossibility)** entail that one cannot acquire a justified belief that one has two hands. Suppose, for reductio, that one can do this. One's having two hands entails that one is not a brain-in-a-vat, for it is a conceptual truth that brains in vats do not have hands. Thus, by **(Closure)**, it is justifiable that the brain-in-a-vat hypothesis does not obtain. This contradicts **(Impossibility)**.

In general: let "$J(p)$" stand for "p is justifiable", and let O stand for any ordinary external world belief. By **(Impossibility)**, there is a scenario S such that $O \to S$, but $\neg J(S)$. However, by **(Closure)** we obtain $J(O) \to J(S)$. Thus, by *Modus Tollens*, $\neg J(O)$.

One way in which the internalism vs. externalism debate arises here is through the suspicion that the argument for **(Impossibility)** rests on contentious *internalist* assumptions. To see this, consider the reliabilist picture again. On this picture, all that is required for justification is that the used belief-forming method is in fact a reliable guide to truth. Now, if S is in the good case — i.e. the world is as it seems to be — then nothing precludes S from reliably forming beliefs about the external world. In particular, so the thought goes, S can also reliably infer that he or she is not a brain-in-a-vat. Thus, the sceptic should not claim that we cannot justify the non-obtaining of the BIV scenario: **(Impossibility)** does not hold. Process reliabilism clearly undermines the sceptical argument above.

Internalists reply that the externalist's responses to scepticism are intellectually dissatisfying. For example: that the externalist can only claim that we acquire justification if he already assumes that we are in the good case, and that this is problematic because a hypothetical sceptic would question this assumption. Moreover, internalists argue that externalist responses miss the point, because the most interesting sceptical challenges are directed against internalistic notions of justification in the first place (Bergmann 2000, p. 164). In particular, so the thought goes, we want to be able to have reflective access to our knowledge, and this possibility is what sceptics purport to undermine (Pritchard 2005, Wright 2004*b*, 2007*b*, 2008). I think that the internalist's complaints are well-motivated and I will come back to these issues below. First, however, I need to draw some distinctions.

3.2 Epistemic warrant and the internalism vs. externalism debate

What has been said above shows that we need to distinguish at least two types of internalisms and corresponding externalisms: those which make a claim *about* our *ordinary notions* of knowledge and justification, and those which make a claim about what is at stake when it comes to scepticism. To render these distinctions clearer, it is useful to introduce a concept of (an epistemic) *warrant.*

3.2.1 The notion of warrant

A ***warrant for p*** is any state that renders a doxastic attitude towards p epistemically valuable.[64] That a subject ***S possesses a warrant for p*** means that this state obtains (for S), but does not entail that S has the doxastic attitude in question. The notion of possessing a warrant is much wider than the notion of possessing justification, understood as the notion of possessing evidence. For example, under certain conditions, even the sheer truth of a proposition might count as a warrant for it. Of course, what precisely will count as a warrant will depend on what epistemic value is. Before I discuss this, some notes and distinctions are in order.

I speak of doxastic attitudes in general — and do not restrict myself to belief — because we will also encounter cases in which the relevant attitude is not belief. However, in this chapter the focus will lie solely on belief, and in what follows, belief is the relevant attitude unless I explicitly say something to the contrary.

I say that a warrant W for *p* is ***doxastic*** just in case W is a warrant for *p* that renders an actually possessed doxastic attitude epistemically valuable, and this attitude is based on W. In case the attitude is belief, I say that S possesses a ***warrant for believing P***. Otherwise, I say that W is a ***propositional*** warrant for *p*, and, in case belief is the relevant attitude, I speak of a subject having a warrant ***to believe p***.

If S does not believe that *p*, then S cannot possess a (doxastic) warrant *for* believing *p*, but might still possess a (propositional) warrant *to* believe *p*. Moreover, there is the possibility that a subject S believes that *p*, but on some other basis or ground than W, in which case W also does not count as a (doxastic) warrant for believing *p*.

Here is an example for the first case: suppose that Hero goes for a walk in Leeds, having

[64] Although in contemporary epistemology the terms "warrant" and "justification" are often used interchangeably, the latter term has some connotations which might lead to confusion.

the experience as of a unicorn walking around Hyde Park. There is indeed a unicorn walking around, and Hero's perceptual faculties are functioning properly. Presumably, this suffices for a warrant to believe that there are unicorns. However, Hero does not form the belief that there are unicorns because he suspects he is hallucinating. Thus, the warrant can only be a propositional one. However, if Hero had formed the belief that there are unicorns on the basis of his experience, then the (reliable) perception would be a doxastic warrant for Hero's belief.[65]

Here is an example for the second case: suppose that Hero in fact holds the belief that there are unicorns, but not on the basis of his experience as of a unicorn. Rather, Hero believes that there are unicorns because they exist in possible worlds different from the actual world. Although Hero now does possess a warrant for believing that there are unicorns (assuming he has good reasons for believing in genuine modal realism), his experience as of a unicorn still only counts as a propositional warrant.

A second relevant distinction is the distinction between *prima facie* and *ultima facie* ("all things considered") warrant. This distinction can be motivated by the observation that a proposition or attitude is always warranted against a complex cognitive background. Some types of warrants can be *defeated by* other warrants, which means that although a subject possesses a warrant of this type, the subject cannot count as warranted overall because of the presence of these other warrants. For example, one's justification for a belief can be defeated by new evidence undermining the antecedent justification (e.g. evidence to the effect that the original evidence was misleading). *Prima facie* warrants are warrants which would render a doxastic attitude epistemically valuable, if we ignored any undermining warrants, but which can be defeated in the sense that they might not render possessing a doxastic attitude epistemically valuable, considering all other warrants the subject possesses. Warrants that are not so defeated I call *ultima facie*, or "all things considered" warrants. Ultima facie warrants are also prima facie warrants.

As an example, consider again the case in which Hero sees a unicorn. It is plausible that Hero's perceptual warrant is defeated by Hero's background information. Everyone knows there are no unicorns, and this certainly undermines the perceptual warrant for the presence of a unicorn. All things considered, the perceptual warrant cannot render the

[65]Of course, this warrant is probably defeated by other evidence Hero possesses. See below.

belief that there are unicorns epistemically valuable, so it is merely a prima facie warrant.

Finally, note that my usage of the term "warrant" is quite different from the common usage of "warrant" as "whatever has to be combined with true belief in order to yield knowledge". I call warrants which are sufficient for knowledge that p — when combined with true belief — *full* warrants. What full warrant consists in is a difficult question which I cannot answer.

3.2.2 Warrant pluralism

I endorse the following claim about warrant and epistemic value:[66]

> **(Warrant Pluralism)** There is a variety of types of warrants, including different types of internalistic and externalistic warrants. Although all these warrants render doxastic attitudes epistemically valuable in some way or other, they do so in different ways, and serve different purposes.

Whether there are different types of warrant rests on the possibility of doxastic attitudes being epistemically valuable in different ways. Thus, **(Warrant Pluralism)** entails:

> **(Multiplicity of Epistemic Value)** There are different ways in which a doxastic attitude can be epistemically valuable.

Note, however, that this claim is weaker than the following claim, which is not entailed by **(Warrant Pluralism)**:

> **(Epistemic Value Pluralism)** There is more than one *fundamental* epistemic value.

There might be a single epistemic value, which can be served in different ways. For example, Alston (2005, chapter 3) argues that there is a variety of very different "epistemic desiderata" that all serve, directly or indirectly, one single principal aim of cognition. Simply put, this aim is gathering true beliefs as opposed to false ones (Alston 2005, p. 30).

[66]Later, it will become important that certain facts about epistemic value are available to relevant subjects, including the readers of this thesis. For the purposes of this project, I assume that these facts about epistemic value can be warranted apriori. In particular, I assume that they can be revealed by apriori philosophical reflection. Maybe they can even be *known* apriori, but this is not crucial for my purposes. Of course, in order to form warranted beliefs about warrant and epistemic value, one needs to possess the concepts of warrant and epistemic value. However, since the readers of this thesis will possess these concepts, nothing hangs on this fact.

Alston's "desiderata" correspond to my "warrants". Thus, Alston accepts **(Warrant Pluralism)** without accepting **(Epistemic Value Pluralism)**. I will remain neutral on the issue of **(Epistemic Value Pluralism)**.

What types of warrant does Alston recognize? Here are three examples:

- The truth of a belief (Alston 2005, p.40).
- That a belief is formed through a reliable process, for such beliefs are normally true (Alston 2005, p.43).
- That the evidence for a belief is reflectively accessible. This *indirectly* serves the truth aim by enabling a subject to discriminate true and false beliefs (Alston 2005, p.43)

Note that the first two warrants are externalistic, whereas the third is internalistic, according to the distinction drawn at the beginning of this chapter.

Another warrant pluralist is Wright (2008), who acknowledges a variety of "epistemic norms", some of which are connected to the "teleology of belief", and some of which are "constitutive of managing a system of beliefs" (Wright 2008, p. 501). Epistemic norms include "truth, knowledge, justification, coherence, and the multi-faceted notion of rationality" (Wright 2008, p. 502). This clearly entails **(Warrant Pluralism)**: every state entailing that a norm of belief is met will count as a warrant, and the way these states render beliefs epistemically valuable will be different for the different norms. For example: the truth of a belief and the presence of justification for a belief are very different things. Wright (2008, p. 505) is unsure about whether his view also entails **(Epistemic Value Pluralism)**.

It becomes apparent that there are different types of warrants serving different purposes. And this bears on the internalism vs. externalism debate. In particular, I agree with Wright in that we should be "receptive to the possibility that externalist conceptions may promise best for some norms, and internalist conceptions for others" (Wright 2008, p. 501).

3.2.3 The question of internalism vs. externalism

The notion of warrant can be used to make explicit what is at stake in the two debates sketched above.

Question 1: analysing knowledge and justification We can approach the question regarding the analysis of the notion of knowledge in the following way: we examine what types of *full* warrant there are, and categorize them into internalistic and externalistic warrants. The externalist (/internalist) can be said to have won the debate if it turns out that all types of full warrant are externalistic (/internalistic) in character.

However, especially in the light of **(Warrant Pluralism)**, it seems unrealistic that all full warrants will be purely externalistic or internalistic in character. Both characters of warrant will be relevant. There is a lot to be said in favour of externalism:

- Sometimes externalism is the only viable option. For example, when it comes to animal knowledge (Sosa 2007).
- Gettier cases show that ordinary external-world knowledge requires that certain external conditions are met, in addition to the truth of the proposition in question.
- In perceptual cases, unreflective reliable belief-formation seems to be sufficient for knowledge. Ceteris paribus, a reliable perceiver of barns can acquire knowledge of the fact that there is a barn nearby just on the basis of a perception of a barn.

However, there are also many instances of knowledge which seem to involve internalistic warrants. For example: knowledge of one's own internal states; mathematical knowledge which rests on having carried out and understood a mathematical proof in full detail.

Thus, a more interesting question about the nature of full warrant is whether one type of warrant is the exception rather than the rule, i.e. whether the majority of full warrants are of externalistic (or internalistic) character. The same issues arise for the *ordinary* notion of ***justification***. For example, Goldman (2009) writes:

> Factors that (help to) fix justificational status are generally called justifiers, or J-factors. So the central question is whether justifiers, or J-factors, have an internalist or externalist character. (...) One configuration of the terms of engagement is existential: externalism wins if there is at least one externalist type of J-factor. Internalism wins only if all J-factors are internalist. A second possible configuration is majoritarian. That side wins that has a majority of types of J-factors of the kind it promotes. (Goldman 2009, p. 310)

Goldman's j-factors are what I call warrant. We can explicate Goldman's first option by the following two contradictory positions:

(Exclusive Internalism) All warrants underlying our ordinary notion of justification are internalistic.

And:

(Existential Externalism) There is at least one externalistic warrant that suffices for being justified in the ordinary sense.

I think that **(Existential Externalism)** is clearly correct. We just saw that there are some cases in which we can ascribe knowledge — and hence also justification — just on the basis of externalistic factors. Thus, a more interesting question concerning the nature of justification is Goldman's *majoritarian configuration*. It can be represented by the following two contradictory positions:

(Majoritarian Internalism) The majority of warrants underlying our ordinary notion of justification are internalistic.

And:

(Majoritarian Externalism) The majority of warrants underlying our ordinary notion of justification are externalistic.

Goldman (2009) argues for **(Majoritarian Externalism)**. I remain neutral on this issue.

Question 2: scepticism Most relevant for my purposes is the second debate — the debate on scepticism. We can precisify the issue of externalism vs. internalism as the following question: which types of warrant are the targets of interesting sceptical arguments? In particular: are such arguments directed against the (possibility of the) possession of internalistic or externalistic types of warrant (or both)? If it turns out that some interesting sceptical arguments are directed against internalistic kinds of warrant, it cannot be satisfactory to have only externalist types of warrant in one's epistemological toolbox.

One might think, in any case, that the most interesting sceptical challenges are directed against the possibility of possessing justification and knowledge *in the ordinary sense.* If this was the case, then our answers to the first and the second question about externalism vs. internalism would coincide: for whatever the kinds of warrant underlying our ordinary notions of knowledge and justification are, they would also be the kinds of warrant the possession of which the sceptical arguments purports to undermine. However, it might just as well turn out that the best or most interesting sceptical arguments make use of notions of warrant that are not best regarded as underlying our ordinary notions of knowledge and justification. In this case, the two questions come apart.

It is thus useful to explicitly distinguish the first issue — the issue of analysing our ordinary notions — from the second issue — the question of which type of warrant is most relevant when it comes to scepticism — by distinguishing **(Exclusive Internalism)** and **(Majoritarian Internalism)** from the following claim, which is entailed by the claim that some interesting sceptical challenges about relevant regions of thought R cannot be discharged by invoking externalist notions of warrant:

> **(Relevance Internalism)** For relevant regions of thought R, every good epistemology for R needs to explain how we can acquire full internalistic warrants for ordinary R-truths — internalistic warrants sufficient for knowledge.[67]

For the purposes of this thesis, the following instance of **(Relevance Internalism)** is particularly relevant, which is equivalent to **(Arithmetical Knowledge)** from 1.2, with an added internalistic requirement:[68]

> **(Relevance Internalism for Arithmetic)** Every satisfactory epistemology for arithmetic needs to explain the possibility of possessing full internalistic apriori warrants for arithmetical truths.

I defend **(Relevance Internalism)** and **(Relevance Internalism for Arithmetic)**

[67] This entails that one needs to account for the possibility of possessing knowledge (in the ordinary sense). So an analysis of the notion of knowledge is not entirely irrelevant here. However, accounting for the possibility of acquiring knowledge of ordinary R-beliefs might not be enough: we need to account for the possibility of acquiring knowledge and for the fact that some of the warrants underlying this knowledge meet internalistic desiderata.

[68] It is equivalent, because arithmetic is a body of truths, which can be believed. If we have established that we can possess warrants sufficient for knowledge, we have established that we can possess knowledge, and vice versa.

below, but I remain neutral about the ordinary notions of justification and knowledge. These questions are very interesting, but not directly relevant to my purposes.

3.3 Relevance Internalism about the external world

3.3.1 Accessible warrant

So far, I have worked with an informal, intuitive characterization of the notions of internalistic and externalistic warrant. In what follows, the following internalistic property of warrant will be relevant:

> **(Accessibility)** A warrant W is *(reflectively) accessible* if and only if the following condition holds: whenever a subject S possesses W, S can — at least in principle — determine that S possesses W, by means of apriori reasoning and self-knowledge[69].[70][71]

Some notes are in order. First, one does not need to possess the theoretical concept of a warrant in order possess an accessible warrant. However, one needs to possess the concepts to be able to possess the thought that one possesses some specific warrant W. Moreover, that one possesses an accessible warrant for *p* does not entail that one is certain that one possesses *p*, or that one is certain that *p*.

Clearly, that *p* is reflectively accessible in the sense defined above entails that one possesses a propositional warrant for *p*. So, if S possesses an accessible warrant for *p*, S does not only possess a warrant for *p*, but also possesses a propositional warrant to believe that W is a warrant for *p*. The following principle holds:

> **(Existential Iteration)** Whenever S possesses W for *p*, and W is accessible, then there is a propositional warrant W', such that S possesses W' for the proposition that S possesses W.

Note: a propositional warrant. It is not required that S actually believes that he or she possesses W for *p*. S merely needs to be in a position to find out that he or she possesses W for *p*. It is a plausible part of the picture that W' will be some kind of internalistic

[69]Self-knowledge is meant to include knowledge by introspection.

[70]This characterization of "reflectively accessible" in the context of internalism is similar to Wright's (see Wright 2004*b*, p.209).

[71]Formulated as above, the criterion does not imply that a subject can also determine — by means of apriori reflection and self-knowledge — that it does *not* possess some warrant of the relevant type.

warrant, maybe even an accessible one. What is required for this is the assumption that, whenever a fact is accessible by (a suitable class of) introspective belief-forming methods and apriori reasoning, it is also accessible by such means that one possesses a warrant for it. This would entail:

(Iteration of Accessibility) Whenever S possesses W for p, and W is accessible, then there is an *accessible* propositional warrant W', such that S possesses W' for the fact that he or she possesses W.

So accessible warrants would be subject to a *WW principle* in the sense that possessing an accessible warrant for p entails the possession of the same type of warrant for the claim that one possesses the warrant. That is: if "$W(p)$" stands for "S possesses an accessible warrant for p", it holds that: $W(p) \rightarrow W(W(p))$. I think that this property is desirable for the internalist. However, that it holds would require an argument for the assumption that self-knowledge and apriori reasoning are reflectively justifiable. I think that the considerations in the final chapter of my thesis provide steps towards this claim, but I cannot say more about it here.

Let "$J(p)$" stand for "S possesses justification for p in the ordinary sense" and "$K(p)$" stands for "S knows that p". It is important to note that the accessibility of J and K alone does not entail any JJ or *KK principle* of the form: (JJ) "$J(p) \rightarrow J(J(p))$" or (KK) "$K(p) \rightarrow K(K(p))$". Such principles require the additional assumption that the warrant provided by apriori reflection and self-knowledge (W) is sufficient for knowledge or sufficient for an ascription of our ordinary notion of justification. W might be too weak. For example: even if a certain bit of knowledge is accessible, the type of warrant that we obtain for the possession of knowledge might not be a *full* warrant. In such a situation we do obtain $K(p) \rightarrow W(K(p))$, but we do not obtain (KK), because $W < K$ (i.e. W is not sufficient for knowledge when combined with true belief).

The notion of internalistic warrant relevant in this thesis is the notion of an accessible warrant. This motivates the following definitions:

(Internalistic Warrant) S possesses an *internalistic warrant* to believe p if and only if S possesses a reflectively accessible warrant W to believe p.

Secondly, although we already defined the notion of a *full* warrant, let me make the combination with internalistic warrant explicit as follows:

(Full internalistic warrant) S possesses a *full internalistic warrant* to believe p if and only if S possesses an internalistic warrant to believe p which is sufficient for knowledge when combined with true belief.

My last definition is:

(Internalistic Knowledge) *Internalistic knowledge* is fully internalistically warranted true belief.

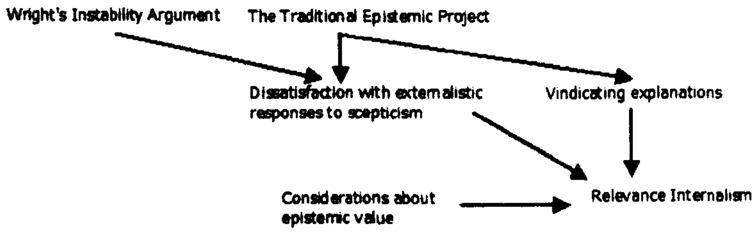
In what follows, **(Relevance Internalism)** should be understood along the lines of **(Full Internalistic warrant)**: the aim is to explain how we can acquire full internalistic warrants in certain areas of cognitive enquiry.

3.3.2 Arguing for Relevance Internalism

(Relevance Internalism) for full internalistic warrant can be motivated by three, interconnected, themes:

- Independent considerations about the epistemic value of full internalistic warrant. For example: that possessing such warrant is conducive to managing a system of beliefs.
- The willingness to engage in what has been called the *Traditional Epistemic Project*.
- Considerations about scepticism. In particular: that relevant forms of sceptical arguments are directed against internalistic (accessible) warrant, and that externalist responses to relevant sceptical arguments are unsatisfactory.

I will say more about the first line of thought in the next chapter (5.1.2), and focus on the latter two points here. My plan is as follows. First, I motivate the Traditional Epistemic Project, and argue that it is inextricably linked with **(Relevance Internalism)**. I then motivate the claim that externalist responses to scepticism are unsatisfactory, assuming we want to engage in the Traditional Epistemic Project. Finally, I provide an independent reason for this claim, using Wright's *Instability Argument* to the effect that externalists cannot provide a dialectically stable response to a simple closure-based sceptical challenge. The dialectic can be displayed as follows:



3.3.3 The Traditional Epistemic Project

My main motivation for endorsing **(Relevance Internalism)** is that I want to engage in the:

> **(Traditional Epistemic Project)**
>
> The *Traditional Epistemic Project* is the project — famously initiated by Descartes — of vindicating from scratch and from the armchair our right to claim knowledge of most of the knowledge we pre-theoretically take ourselves to possess, bracketing all antecedently held beliefs about the external world .

The project can be pursued by telling an epistemological *Hero story*: a story of a non-defective epistemic agent — Hero — who goes through some canonical lines of reasoning, and realizes that he possesses warrants sufficient for knowledge for most of the beliefs he antecedently took to be knowledge. The project is motivated by the desire to give a ***vindicating explanation*** of all our external-world knowledge, i.e. to give an answer the question whether and how "I really know *any* of the things which I take myself to know about the world" (Leite 2005, p.514).

Answering this question is subject to a constraint. It can be brought about by the following considerations. When we are challenged to explain whether and how we know a particular proposition p, we cannot appeal to p in our explanation, because of the "pragmatics of assertion or explanation" (Leite 2005, p. 514). Consequently, if we want to explain whether and how we know *anything* about the world, our explanation cannot make (inelimable) appeal to *any* external-world considerations. In other words, a vindicating explanation for all our external-world knowledge requires that, at the beginning, all antecedently held beliefs about the world are bracketed.

An attentive reader might discover some tension in this demand: in order to establish that we possess knowledge of p, so the thought goes, we certainly need to claim that p is true because it is a necessary condition for knowing that p, and many p will be claims about the external world. So is it trivially impossible to provide vindicating explanations for external-world knowledge?

No. For one, it is not so clear that one always needs to cite p as a premise in order to establish that one knows that p. We will see examples for such arguments below. Secondly, the whole project should not be understood in a way which renders it impossible for this reason. If I have established that an antecedently held belief is warranted in a way that suffices for knowledge, when combined with true belief, then I can take this belief to be knowledge. What vindicating explanations really forbid is ineliminable appeal to considerations about the external world, when it comes to establishing that we possess such warrants. We can make this explicit by imposing the:

> **(Cartesian Constraint)** Our arguments for the possession of *full warrants* for relevant propositions must not *ineliminably* appeal to any considerations about the (external) world.[72]

However, at first glance, this constraint still seems impossible to meet for full warrants which are subject to external conditions, and it is very plausible that at least some relevant full warrants are subject to external conditions because of Gettier cases.

In general, it appears that the **(Cartesian Constraint)** is incompatible with externalism in any sense entailing that it is a necessary condition for the possession of knowledge that an external condition in addition to truth holds. For suppose there is such a condition. Then how are we to argue that we possess full warrants without citing these external conditions as premises? Suppose, for example, that the reliability of perceptual belief-forming methods is a necessary condition for acquiring perceptual knowledge. We would not be able to establish that we possess such knowledge without appealing to external-world conditions to the effect that my perceptual faculties are reliable. Any argument for a particular item of perceptual knowledge will involve a reliability claim as a premise.

Does this show that the **(Cartesian Constraint)** is incompatible with such externalism? This would be a bad result, for such externalism might well be correct.

[72]A slightly different criterion with the same name has been suggested by Leite (2005).

Fortunately, what has been said does not yet show that there is such incompatibility. For the **(Cartesian Constraint)** allows for an appeal to external-world conditions, as long as it is *eliminable*. So we need to argue that there is a notion of "making eliminable appeal to p", meeting the following two criteria:

- It is possible to eliminate appeal to external-world conditions in our vindicating explanations.
- Explanations making only eliminable appeal to external-world conditions are suitable for pursuing the **(Traditional Epistemic Project)**.

Leite (2005, p. 516) argues that this cannot be done. From the point of view of someone pursuing the **(Traditional Epistemic Project)**, so the thought goes, an appeal to an external-world condition (e.g. reliability) in a particular vindicating explanation can only count as eliminated if we can already claim to possess a full warrant for the condition, because it is part of the collection of propositions knowledge of which we wanted to vindicate. However, how are we to account for this warrant? It seems as if we will run into circularity at some point. It seems as if we cannot eliminate appeal to some external-world conditions if what is at stake is *all* our external-world knowledge.

Leite (2005, p. 518) suggests that we should engage in a lighter epistemic project, in which one gives up the claim that appeal to external-world conditions must be eliminable, and in which one proceeds piecemeal, always assuming some external-world conditions. However, this amounts to giving up the **(Traditional Epistemic Project)**, so it is not an option.

One might want to respond that all this does not show that we cannot account for the possession of types of warrants that are not subject to external conditions. Maybe we should not aspire to give vindicating explanations for all our *knowledge* (*full warrants*), but just give vindicating explanations for the possession of types of warrants that are not subject to external conditions, or so the thought goes. However, anyone engaging in such a project can be accused of changing the subject. The initial motivation for the **(Traditional Epistemic Project)** was vindicating claims to *knowledge* (Leite 2005, p. 517).

So what are we to do? I think that the key to an answer is to observe that Leite's requirements for when an appeal counts as eliminated are too strict. Leite demands that an appeal to an external-world condition can only count as eliminated if we already established that we possess a *full warrant* for it. However, what is really important when we cite an external-world condition is that we do not beg the question. And if we can establish that we are in a position to claim that this condition is met, and this claim does not rest on any considerations about the external world, we do not beg the question. What is important is that we can assure ourselves, *from the inside*, that we possess a warrant for it. And *this* warrant does not have to be sufficient for knowledge in the ordinary sense, or be subject to external conditions, contrary to what Leite assumes.

In other words, my suggestion is that it is acceptable to cite external-world propositions as premises in our arguments for the possession of full warrants as long as we possess some *purely internalistic* warrant for these premises. Note that this is not the cheap internalist response that we give up accounting for knowledge meeting externalist conditions. We still do that. However, we allow that the premises of our arguments for the possession of such knowledge are warranted internalistically (and in a way that might not be sufficient for knowledge in the ordinary sense).

Suppose that we have to claim that our perceptual faculties are reliable (*Rel*), in order to claim that we possess a full warrant for the claim that we have two hands. I contend that our appeal to *Rel* is acceptable as long as we can establish, just on the basis of apriori reasoning and self-knowledge, that we possess a warrant for *Rel*. In this case, we can assure ourselves, from the inside, that we can (warrantedly) claim *Rel*, and we cannot be accused of begging the question. Although we have to cite an external-world condition, we can eliminate our appeal to it, by arguing that we are in a position to cite it, without making appeal to any external-world conditions.[73]

In general, the thought is that our vindicating explanations must *ultimately* — at some level — be grounded in internalistic warrants — warrants accessible by apriori reasoning and self-knowledge. Maybe what is fundamentally at issue between the externalist and the internalist is precisely whether justification should be, and can be ultimately grounded in this way.[74] This picture implies that Leite is wrong and the success of the **(Traditional Epistemic Project)** is independent of the analysis of the ordinary notion of knowledge

[73]Of course, this raises the question how exactly *Rel* is warranted. I discuss such questions in the next chapter.

[74]For example, my response assumes that there are genuinely epistemic types of warrants (and reasons) that are not subject to external conditions. Externalists might doubt that this is so.

in the sense that it will not pose a special problem if the warrants underlying knowledge in the ordinary sense are subject to external conditions.

Note that if we meet **(Relevance Internalism)** in a certain area of cognitive enquiry, we will also be able to meet the **(Cartesian Constraint)** in this area. For **(Relevance Internalism)** requires that our knowledge is available on the basis of apriori reasoning and self-knowledge, and this entails that we can argue that we know relevant propositions, without making ineliminable appeal to external-world considerations. Moreover, if we meet the **(Cartesian Constraint)** in a certain area of cognitive enquiry, then we also meet the requirements of **(Relevance Internalism)**. For the existence of arguments for knowledge meeting the **(Cartesian Constraint)** show that the relevant knowledge claims are available just on the basis of apriori reasoning and self-knowledge. The constraints are necessarily equivalent.

3.3.4 Scepticism

Now reconsider the closure-based sceptical challenge from above. According to process reliabilist, it is simply not true that the non-obtaining of sceptical scenarios cannot be justified. In particular, so the thought goes, we can acquire justification for ordinary external-world beliefs, and for the non-obtaining of sceptical hypotheses, as long as the external cognitive environment allows for reliably forming these beliefs.

In order to further examine this response, I put the closure-based argument in a more general setting. Let "W" be a template for a certain form of warrantability for which a relevant closure principle holds, let "O" stand for an ordinary statement about the external world, and let "SH" stand for a proposition expressing a typical sceptical hypothesis such as the BIV scenario. I examine the following argument template:

(CB1)	$\neg W(\neg SH)$	*Sceptical Premise*
(CB2)	$W(O) \rightarrow W(\neg SH)$	*Closure Principle*
(CB3)	$\neg W(O)$	**(CB1)**, **(CB2)**, *Modus Tollens*

Again, the idea is that, for some reasons connected to our notion of warrant W and the subjective indistinguishability of the good case and SH, the non-obtaining of the sceptical hypothesis cannot be warranted, and that this undermines our warrants for the ordinary propositions because of a closure principle for W. Since O was chosen arbitrarily, the argument — if sound — would establish that we do not possess W-warrants for any ordinary external-world proposition.

We assumed closure. So the anti-sceptic needs to attack **(CB1)**. How exactly the anti-sceptic will attack this premise will depend on the type of warrant in question. If the notion of warrant is a simple externalist notion, such as reliability, this *seems* relatively easy. The reliabilist can simply deny to assert **(CB1)**, because this would be tantamount to the asserting that we are in a sceptical scenario, or so the thought goes. After all, in every non-sceptical scenario, we can clearly be warranted in believing that no sceptical scenario obtains, for such beliefs can be reliably formed.

However, since this is all the externalist has to say, this response clearly violates the **(Cartesian Constraint)**. The reliabilist cannot claim that our beliefs are actually reliably formed, without presupposing that we are in the good case, and he has nothing to say about how this appeal can be eliminated.[75] Hence, the reliabilist cannot claim that we possess a warrant for ordinary external world beliefs, without (ineliminably) presupposing that we are in the good case. This problem generalizes to all externalist notions.

We do not have to appeal to the **(Cartesian Constraint)** in order to see that there is something wrong with purely externalistic responses. Pritchard (2005, chapter 4) argues that the "heart of" the sceptical challenge is that we need to provide reasons of why our (internal) evidence favours one hypothesis about the external world over another:

> for isn't the concession that the [internal] 'evidence' one has in favour of one's everyday beliefs doesn't favour those beliefs over belief in known sceptical alternatives simply the concession that one doesn't really have any evidence of substance in favour of one's everyday beliefs? (Pritchard 2005, p.112)

One way to make this precise is to consider a hypothetical sceptic who believes that he is a BIV, and who reasons as follows: "I have an experience as of my two hands" — "Therefore, there is a computer designed by an evil scientist just emulating my two hands". We want to be able to say why an experience as of two hands favours the ordinary external-world

[75] Of course, if the reliabilist would be able to argue apriori that we are in the good case, without appealing to any external-world considerations, things might look different. However, in this case the response becomes an internalist one, in the sense of **(Relevance Internalism)**. The response would not be purely externalistic. A more interesting case would be one where apriori knowledge is explained externalistically. I do not know what to say about such responses, but such a response seems to me to be subject to the same difficulty as the purely reliabilist response, because, at some point, appeal to external-world considerations is ineliminable.

belief over the analogue of the hypothetical sceptic.

An argument by Wright (2007*b*, 2008) builds on similar intuitions. Wright argues that the externalist response to scepticism is in some important sense dialectically unstable, and that only recourse to accessible warrants can avoid such instability.

3.3.5 Wright's Instability Argument

Wright (2004*b*, 2007*b*, 2008) agrees with the advocate of the **(Traditional Epistemic Project)** in that he thinks that reliabilist responses to the closure-based challenge are not very interesting. Clearly, it is not enough for the anti-sceptic just to note that the possession of externalistic warrant — and knowledge — is *possible*. Rather, so the thought goes, we want to be in a *position to claim* that we possess such knowledge:

> To be sure, if the sceptical argument is taken to be to the effect that knowledge of the material world is impossible, then it must founder if a reliabilist conception of knowledge is sound; for even the most skilful monger of paradoxes cannot show that we are not as a matter of fact so situated in a material world that our cognitive faculties reliably generate mostly true beliefs about it. But the residual dissatisfaction with the externalist suggestion as a response to scepticism is that it merely points to a congenial possibility: nothing has been offered to put us in position to claim that it, rather than one of the many contrasting uncongenial sceptical scenarios, actually obtains. (Wright 2007*b*, p.7)

According to Wright (2004*b*, 2007*b*, 2008), a harder (and more interesting) sceptical challenge arises. We need to argue that we can be in a position to *rationally claim* that we are warranted in believing ordinary external-world propositions. And, according to Wright, rational claimability is an internalistic notion.

However, two questions arise:

1. Why should the externalist agree that we need a notion of being able to claim a warrant, in addition to whatever notion of warrant we use?

2. Why should the externalist agree that this notion is to be construed internalistically?

As to the first question: the externalist has to rebut the sceptical argument — it is not enough just to remain silent. The externalist wants to be in a position to explain what is wrong with it. Assuming closure, the externalist needs to argue that something is wrong with the first premise **(CB1)**. This could be done by arguing that the sceptic *is not in a position to claim* this premise because claiming it is tantamount to the assertion that we are in the bad case and the sceptic has provided no considerations at all of why we should be in the bad case.

As to the second question, Wright (2007*b*, 2008) argues that it is impossible to construe this notion externalistically, on pain of dialectical instability. According to Wright, there are two constraints on what would count as an effective attack against the sceptic, and these constraints make it impossible to effectively respond to the sceptical argument if "W" is construed along *simple* externalist lines (e.g. along the lines of reliability). Let me explain.

For one, assume that the anti-sceptic has established that the sceptic cannot properly claim **(CB1)**. The anti-sceptic should not be happy with just that. The anti-sceptic also wants to be in a position to claim that a hypothetical sceptic who actually believes that he is in a sceptical scenario does not possess warrants for his deviant beliefs. In other words, the anti-sceptic wants to be able to claim the non-sceptical analogue of **(CB1)** which Wright (2008) dubs "**":

(**) $\neg W(SH)$

Note that this is similar to Pritchard's intuition that we want to be able to claim that our evidence favours the ordinary beliefs, and not analogues of hypothetical sceptics. In general, the anti-sceptic wants to arrive at an *asymmetric* situation. This yields our first constraint:

> **(Asymmetry)** The anti-sceptic needs to be able to make positive claims about his epistemic situation which are in a better standing than the corresponding negative claims made by a (hypothetical) sceptic.

This becomes problematic for the externalist if we add a second, plausible constraint to the effect that the debate between the (hypothetical) sceptic and the anti-sceptic is to take place on neutral ground, i.e. that the responses to the opponents arguments do not assume any propositions which already entail that the subject is in either the good case or the bad case:

(Neutrality) In the course of rebutting the opponent's arguments, neither the sceptic nor the anti-sceptic are allowed to make any assumptions that are not neutral with respect to the actual situation of the subject in the external world.

Now consider the pair **(CB1)/(**)** $= \neg W(\neg SH)/\neg W(SH)$. Asymmetry demands that the anti-sceptic can establish that the former cannot be claimed without depriving himself of the possibility of being able to claim the latter. **(Neutrality)**, however, makes it hard to see how the argument against the claimability of the former does not also apply to the latter, if "W" is a simple externalist notion like reliability.

Consider the reliabilist notion of warrant. The reliabilist's idea to undermine the sceptic's claim to **(CB1)** was that claiming it presupposes that we are not in the good case, and that there is no way the sceptic can argue, on neutral ground, that we do not actually form beliefs in a reliable way. By the sceptic's own lights, the good case and the bad case are subjectively indistinguishable. So the sceptic would need to appeal to considerations about the external world. Thus, the sceptic's ability to claim **(CB1)** is undermined by **(Neutrality)**.

So far, so good. The problem for the anti-sceptic is that it is now hard to see how the anti-sceptic's ability to claim **(**)** is not similarly undermined by **(Neutrality)**. To see this consider again a hypothetical sceptic who believes that he is a BIV reasoning as follows:

1. I have, right now, the experience as of my two hands.

2. Therefore, the big computer creates my experience as of a world in which I have two hands.

3. Therefore, I am a BIV.

If the hypothetical sceptic was a BIV, the sceptic would obtain a warrant for **(**)** by this line of reasoning, since his beliefs would be reliably formed, and closure holds. Thus, the anti-sceptic's ability to claim **(**)** is undermined by a line of reasoning exactly analogous to his own line of reasoning against **(CB1)**. In particular: claiming **(**)** presupposes that we are not BIVs, and there is no way the anti-sceptic can argue, on neutral ground, that the hypothetical anti-sceptic does not actually form his beliefs in a reliable way. For the BIV case and the external world case are, by assumption, subjectively indistinguishable. Thus, the pair $\neg W(\neg SH)/\neg W(SH)$ cannot be where where asymmetry is found.

The externalist might try to find other ways to distinguish both cases. Note that the reasoning above still relies on an intuitive notion of claimability. The externalist might look for a way out by construing the notion in an externalistic way. For example, the externalist anti-sceptic might construe claimability as (externalistic) warrantability, and make the following claim (where both occurences of "W" stand for the externalistic notion of warrant in question):

(*) $\neg W(\neg W(\neg SH))$

The problem is that, given some plausible assumptions — (*) entails(**), which we have seen cannot be claimed without a violation of **(Neutrality)**.[76] Thus, (*) cannot be claimed by the anti-sceptic and the dialectical situation is still symmetric.[77]

This is a serious problem for the stubborn externalist. To be sure, it is not the case that the externalist is forced into scepticism. The externalist might simply uphold his or her beliefs. Rather, it means that there is no unproblematic way to *reject* even a very basic sceptical argument without violating either **(Asymmetry)** or **(Neutrality)**. Call a response to the sceptic which meets both demands a *stable* response. The problem is that the externalist cannot find such a stable response, and this seems to be deeply unsatisfactory.[78]

[76] $W(\neg SH)$ entails $\neg SH$, if "W" is construed externalistically. Moreover, because "$\neg W \neg$" is closed under logical consequence (this follows from the fact that "W" is so closed), $\neg W(\neg W(\neg SH))$ entails $\neg W(\neg\neg SH)$, which entails (**), assuming we have available the classical principle of double negation elimination.

[77] A second way to see the dialectic instability is to realize that (*) leads to a *Moorean instability* in the sense that, for some P, the externalist anti-sceptic has to assert both P and $\neg W(P)$. This was pointed out to me by Robert Williams.

Think about how the externalist anti-sceptic will argue for (*): the argument will be based on the thought that we might be in the good case, and in the good case we can become warranted in believing $\neg SH$. However, exactly the same considerations can be applied the "other way round": we might be in the bad case, and in the bad case we can become warranted in believing SH. In other words, if we can argue for (*), then we can also argue for: $\neg W(\neg W(SH))$. Now set $P = \neg W(SH)$. The anti-sceptic has to accept both P and $\neg W(P)$.

[78] Of course, in order to avoid Wright's argument, the externalist might still reject closure by going for a sensitivity-based account. Moreover, the argument as stated only goes through in classical logic because we need double negation in order to obtain (**) from (*).

Any externalist? The argument above assumed closure and focused on process reliabilism, which provides the externalist with room for maneuvre. Assuming that closure for externalistic warrant is not negotiable, we still need to argue that similar considerations apply to all other externalist notions of warrant that obey closure. Wright (2008) shows that the argument also applies to safety accounts. This strengthens the argument. Although it is still incomplete, I think what has been said at least shifts the burden of proof to the side of the externalist.

Williamson (2000) rejects **(Neutrality)**. His main strategy against scepticism is to note that:

> Nothing (...) should convince someone who has given up ordinary beliefs that they did not in fact constitute knowledge, for nothing said here should convince her that they are true. The trick is never to give them up." (Williamson 2000, p. 27)

According to Williamson, the best strategy is never to be neutral. As long as we stick to our ordinary beliefs, we have plenty to say against the sceptic. On the other hand, Williamson seems to concede Wright's point that, as soon as we occupy the neutral standpoint, we cannot effectively respond to the sceptical arguments if we are externalists (otherwise we would not need a "trick"). Thus, Wright and Williamson seem to agree that externalism, closure, and the two constraints are incompatible.

I think that **(Neutrality)** captures what is so unsatisfying about externalist responses: they always assume something about the world, and ultimately leave us with the feeling that they might be false. Only endorsing an internalistic notion of warrant (or claimability) will enable us to meet **(Neutrality)**. This is because the conditions for the possession of internalistic warrant will not depend on the state of the external world. The hope is that we will be able to argue against scepticism from a neutral standpoint on this basis. This provides further motivation for **(Relevance Internalism)** in the external-world case.

The notion of an accessible warrant seems to be particularly suitable for Wright's purposes. Arguing that we can possess an accessible warrant for the non-obtaining of sceptical scenarios is tantamount to arguing that we can stably affirm that we possess such a warrant, for, by definition, an accessible warrant is available by apriori reasoning and self-knowledge, and thus promises to preserve **(Neutrality)**.

Wright's idea is just that, although he uses a different terminology. Wright (2004*b*, 2004*a*) employs an internalistic notion of *claimable warrant* in order to be able to stably affirm that we are in a position to rationally claim that we possess a warrant for the non-obtaining of sceptical scenarios. We will see in the next chapter how this works. I now examine Wright's notion of claimable warrant, and argue that possessing a claimable warrant in Wright's sense is equivalent to possessing an accessible warrant. We thus obtain a unified internalist notion of warrant, which I will employ in my further investigations.

3.4 Wright's notion of being in a position to claim a warrant

Wright (Wright 2004*b*,*a*, 2007*b*, 2008) draws a distinction between *possessing* a warrant and *being in a position to (rationally) claim*[79] a warrant. Moreover, he argues that it is our right to rationally claim warrants that is under attack by the most interesting sceptical arguments:

> I want to contrast the idea of possessing a warrant for P with another idea, viz. that of a thinker's being in position to claim possession of a warrant for P. And by this, I do intend something with internalist resonances. I want to understand the claimability of a warrant to be what is at issue when, for example, a philosopher feels that one has not been given everything one needs to address scepticism about the external world, say, merely by impressive arguments — if any such there be — that knowledge can be constituted by reliably generated true belief. (Wright 2007*b*, p. 30)

Thus, the notion of claimable warrant is also the notion that Wright thinks can be used to provide a stable response to the closure-based sceptical challenge (2004*b*). How should this notion be understood?

That an agent is in a position claim a warrant for p could be interpreted to mean that the agent is in a position to claim that he possesses some warrant or other. However, this usage of the notion is not very illuminating. Wright endorses the notion in contexts like "Hero can claim that he is justified in believing p" or "Hero is in a position to claim that he knows that p". So we should construe the notion as one of being in a position *that* one

[79] Wright uses "being in a position to claim" and "being in a position to rationally claim" interchangeably. So do I. The latter may be used to indicate more clearly that Wright has an internalist notion in mind.

possesses a *particular* warrant W for p. So what does it mean that an agent is in a position to claim that he possesses W for p?

A simple answer is that the conditions for properly asserting that one possesses W must be met. However, this just raises the further question of what these conditions are.

In any case, note that being in a position to claim a warrant is tantamount to the possession of a higher-order warrant: for however the notion is analysed, being in a position to claim that q will make a doxastic attitude towards q epistemically valuable in some way or other, so being able to rationally claim q entails the possession of a warrant for q. Thus, being in a position to claim W for p entails possessing a propositional warrant for the claim that one possesses a warrant for p. Note: a propositional warrant. To be in a position to claim a warrant neither entails that one believes that one possesses the warrant, nor that one actually claims the warrant.

Two questions arise with respect to the question of how to construe the higher-order warrant. Firstly: is it factive in the sense that being in a position to claim a warrant W entails that one possesses W? And secondly, can we not construe the higher-order warrants — the warrant for the claim that one possesses a warrant — externalistically? As to the first question, I think that the notion should not be construed as a factive notion. For example, it has become apparent that truth might be a warrant, and, clearly, Wright thinks that one can be in a position to claim that a statement is true without the statement being true. However, in most cases where I use the notion it is assumed that the subject actually possesses the warrant in question. As to the second question, given what has been said above, it should be clear that Wright must hold:

(Equivalence) For any warrant W a subject S possesses for p, W is rationally claimable by S if and only if its possession is reflectively accessible by S, i.e. available by apriori reasoning and self-knowledge.

For this is crucial for his claims about internalism and scepticism. As we saw above, Wright thinks that an externalistic construal of claimability precludes us from giving a satisfying account of scepticism. Using this notion of rational claimability, we can define the following property of warrants:

(Wright Accessibility) A warrant W is *Wright-accessible* if and only if whenever S possesses W, S is in a position to rationally claim *that* he or she possesses W.

By **(Equivalence)**, it follows that **(Wright Accessibility)** is equivalent to the earlier **(Accessibility)**. This equivalence motivates introducing a single, unified notion of internalistic warrant and knowledge. We could just as well have defined the notion of internalistic warrant as follows:

(Internalistic Warrant*) S possesses an *internalistic warrant* to believe p if and only if S possesses a warrant W to believe p and S is in a position to claim that he or she possesses W.[80]

And:

(Full internalistic warrant*) S possesses a *full internalistic warrant* to believe p if and only if S possesses an internalistic warrant to believe p which is sufficient for knowledge when combined with true belief.

Using Wright's notion, *internalistic knowledge* should be defined as follows:

(Internalistic Knowledge*) S possesses *internalistic knowledge* of p if S possesses knowledge of p and S is in a position to rationally claim this knowledge.[81]

The notion of **(Internalistic Knowledge*)** is equivalent to the above-defined notion of **(Internalistic Knowledge)**, given some plausible assumptions (which I shall assume in what follows):[82]

- "**(Internalistic Knowledge*)→(Internalistic Knowledge)**" is is trivial. For to be in a position to rationally claim that one knows that p entails that one is in a position to rationally claim that one possesses a full warrant for p, which entails that one's full warrant is reflectively accessible.

- "**(Internalistic Knowledge)→(Internalistic Knowledge*)**" requires three assumptions. First, assume that the following two conditions are sufficient for being in

[80] Because of **(Equivalence)**, we could just as well have used an accessibility criterion as the crucial property for internalistic warrant.

[81] Moreover, because of **(Equivalence)**, we could just as well have defined internalistic knowledge as knowledge whose possession is reflectively accessible.

[82] Pritchard (2005, section 3.4) argues for a similar claim, namely that it is plausible that knowledge meeting an internalistic (accessibilist) justification condition and knowledge which can be properly asserted coincide. The former corresponds to the first definition, the latter to the second definition.

a position to rationally claim that p: (i) one believes that p; (ii) one is in a position to rationally claim a full warrant for p. Second, assume that believing p suffices for being in a position to claim that one believes p. Third, assume that being in a position to rationally claim a full warrant entails being in a position to claim that it is a full warrant. Given the three (plausible) assumptions, the agent who possesses a fully internalistically warranted true belief is in a position to claim that all conditions for knowledge are met, and hence in a position to claim this knowledge.

Above, I defended **(Relevance Internalism)**, where internalistic warrant was understood along the lines of **(Accessibility)**. We can define a version of **(Relevance Internalism)** using Wright's notion of claimable warrant:

(Wright Internalism for R) Every satisfactory epistemology for R has to explain how we can possess *claimable* knowledge of ordinary R-truths.

In the remainder of this chapter, I will argue for:

(Wright Internalism about Arithmetic) Every satisfactory epistemology for arithmetic needs to explain how we can possess *claimable* apriori knowledge of ordinary arithmetical truths.

3.5 Wright Internalism about Arithmetic

I first examine an argument for **(Wright Internalism about Arithmetic)** that I find wanting. Although it might provide some motivation for the claim, it cannot be used to establish it.

3.5.1 The argument from mathematical practice

One might think that actual mathematical practice is in favour of internalism, because proof is considered to be the gold standard of justification in the discipline. The best explanation of this fact, so the thought goes, is that mathematicians aim at being in a position to claim that they possess knowledge and justification. And since mathematicians aim at claimable knowledge, our epistemology should account for it.

In fact, it might look as if mathematicians engaged in arithmetic can live up to this aim, because the following principle is initially plausible:

(Claimable Knowledge by Proof) If one believes that p on the basis of a correct mathematical proof from arithmetical axioms, then one is in a position to claim apriori knowledge of p.

Unfortunately, if "being in a position to claim knowledge" is Wright's internalistic notion presented above, then **(Claimable Knowledge by Proof)** is far from obvious, and probably false in many cases. It presupposes that the axioms can also be claimed to be known apriori, because one cannot claim to know p apriori on the basis of a proof from some other statement q unless one is also in a position to claim apriori knowledge of q. However, if proof is our best candidate for the means by which claimable knowledge is generated in mathematics, then the story is incomplete. For at least some axioms cannot be established by any further proof. Thus, proof cannot be the gold standard for the justification of *all* mathematical beliefs. The crucial question is whether the gold standard for the axioms, whatever it is, is sufficient for claimable apriori knowledge in Wright's sense.

It might turn out that it is not. For example, it might turn out that mathematicians do not possess claimable apriori knowledge of the axioms, but just some externalistic type of apriori knowledge. In this case, the lower epistemic status of the axioms (whatever it is exactly) will leach upwards to the theorems one proves from them. Although the warrant we obtain for the theorems would be *inferential* and *apriori*, it would not be *internalistic* in the required sense. And it is not obvious at all that apriori knowledge of axioms must be internalistic. One might think that apriori knowledge of necessary truths must be internalistic. However, there are externalistic accounts of such knowledge. For example, Jenkins (2008*b*) argues for an externalistic epistemology of mathematical axioms, based on the idea that mathematical knowledge can be based on reflection on our concepts.

An externalistic picture is perfectly compatible with the fact that mathematical justification involves proof. For example, one might regard proofs as the most reliable belief-forming method, and thus the best way to expand mathematical knowledge, externalistically conceived. That mathematicians regard proof as the gold standard of justification can neither be used to show that internalistic knowledge is actually possessed, nor that mathematicians actually aim at internalistic knowledge. This makes it much more difficult to construct an argument to the effect that every satisfactory epistemology of mathematics *should* account for internalistic knowledge, on the basis of actual mathematical practice.

Of course, the kind of reflection we discover in actual mathematical practice certainly suffices to *initially motivate* the claim that an internalistic treatment of the whole discipline is required. But to *establish* this claim on this basis is a different matter entirely.

3.5.2 The Argument from Analogy

I believe that Wright's considerations for **(Relevance Internalism)** can also be applied to arithmetic, assuming the semantic component of **(Arithmetical Platonism)**, i.e. that the role of number terms is to refer to mind-independent abstract objects. I shall call the ensuing argument the *Argument from Analogy*. This argument is not only interesting because it is an argument for internalism, but also because it sheds light on a connection between closure-based external-world scepticism and similar challenges in mathematics. Such connections have not been discussed to a great extent. To my knowledge, the first to uncover such connections was Pedersen (2006), who discusses a Moorean argument for arithmetic.

It transpired that at least simple externalist notions of warrant cannot be endorsed to stably rebut the simple closure-based sceptical challenge in the external-world case. I now construct an analogous argument for the arithmetical case, assuming that the role of number terms is to refer to mind-independent abstract objects. The upshot is that we can construct a closure-based challenge for arithmetic, and mirror Wright's reasoning above.

In order to construe such a challenge, we need to find analogues for O and SH above. Analogues for O are easily found. We just pick an ordinary arithmetical belief, such as "$1 + 1 = 2$" ($O[Math]$). What is the analogue for the sceptical hypothesis SH? The scenario needs to be such that our ordinary arithmetical beliefs were false although our inner cognitive situation remains the same.

If the role of number terms is to refer to numbers — mind-independent abstract objects — it appears plausible that there is such an hypothesis. For then we can make sense of a situation in which numbers do not exist, but in which our inner cognitive situation remains the same. That we can make sense of such a situation, so the thought goes, is also one of the intuitions underlying Benaceraff's and Field's challenges from 1.1. Thus, I suggest the following sceptical hypothesis:

(SHM) There is no mind-independent realm of abstract mathematical objects.

Clearly, the negation of **(SHM)** follows from $O[Math]$, given that $O[Math]$ is interpreted as being about a mind-independent realm of abstract mathematical objects. We thus obtain the following closure-based sceptical challenge for mathematics:

(SHM)	$\neg W(\neg SHM)$	*Sceptical Premise*
(CBM2)	$W(O[Math]) \rightarrow W(\neg SHM)$	*Closure Principle*
(CBM3)	$\neg W(O[Math])$	**(CBM1)**, **(CBM1)**, Modus Tollens

A hypothetical sceptic, so the thought goes, might argue for **(SHM)** in a similar way as for the corresponding premise in the external-world case: we cannot subjectively distinguish the good case (the case in which there is a realm of mind-independent abstract objects) from the bad case (the case in which everything appears normal but there is no realm of mind-independent abstract objects), so we cannot be warranted in believing that we are in the good case, for some interesting notion of warrant. And now we just mirror Wright's argument above. The externalist might want to counter this move by invoking an externalist notion. But this leads to dialectic instability, as long as we make the sceptical assumption (shared by the externalist) that we cannot subjectively distinguish the good case from the bad case. In sum: assuming that the role of number terms is to refer to numbers, we obtain an Instability Argument against simple externalism, which shows that we can only effectively rebut the sceptical argument using an internalistic notion of warrant. This strongly motivates **(Relevance Internalism)** for the arithmetical case, and, equivalently, **(Wright Internalism for Arithmethic)**. For arguing that we can possess accessible warrants for the existence of a mind-independent realm of abstract objects would provide us with a stable response to the closure-based sceptical challenge.

An objection I envisage is that there is an important difference between the mathematical case and the external-world case, because numbers are supposed to be pure abstracts. Thus, if numbers exist, they exist necessarily. This creates a disanalogy to the external-world case, so the thought goes, because in the external-world case, both the good case and the bad case were possible, which allowed for a standoff.

Why can there be no standoff in the number case? The thought must be that it must be apriori detectable whether numbers exist or not. And if this is the case, we have an asymmetric situation. However, either this apriori justification is internalistic or it is externalistic. If it is externalistic, then the Instability Argument is not avoided. If it is internalistic, then we do not need the Instability Argument, because we already established that justification in mathematics is internalistic justification.

Thus, if we do not already assume internalism, we have to allow for the sceptic arguing for the claim that we are not warranted in believing that numbers exist necessarily, and that our inner situation does not allow to distinguish the case in which numbers exist necessarily from the case in which numbers necessarily don't exist.

3.5.3 Arithmetic and the Traditional Epistemic Project

Be that as it may, my primary motivation for **(Relevance Internalism about Arithmetic)** rests on the desire to engage in the **(Traditional Epistemic Project)**. I think that the **(Traditional Epistemic Project)** should be pursued wherever possible. Thus, it should also be pursued in the logico-mathematical case, and in the arithmetical case in particular. Moreover, the possession of accessible warrant generates additional epistemic value. Clearly, we should account for as much epistemic value as possible.

Of course, the question arises of how the **(Cartesian Constraint)** is to be interpreted in the logico-mathematical case. The constraint prevents us from making ineliminable appeal to external-world propositions in our second-level justifications, but what are these propositions in the logico-mathematical case? I think that the best way to interpret the constraint is that we have to avoid making ineliminable appeal to logico-mathematical basic principles in our second-level justifications in these areas. For example: we should not make ineliminable appeal to a validity claim for a rule R in the course of arguing for the claim that we are justified in believing that R is valid, and we should not make ineliminable appeal to an arithmetical claim in the course of arguing for the claim that we are justified in believing that arithmetical axioms are true. It will become apparent that we can meet this constraint because we can establish logico-mathematical basic principles on the basis of self-knowledge of our own meaning-fixing commands.

3.6 Intermediate conclusion

Let me sum up: I sketched two ways in which debates between internalists and externalists might emerge in epistemology. After that, I introduced a general concept of warrant and rephrased both debates using this concept. I introduced **(Relevance Internalism)**, and argued for two instances of it. I argued for **(Wright Internalism about R)**, for R="the external-world" and R= "Arithmetic".

These are very strong demands. Many have argued that such demands lead to scepticism. I now begin to argue how scepticism can be avoided. I engage in the **(Traditional Epistemic Project)**. In the next two chapters, I focus on the external-world case and the logical case, to point out certain features of my epistemological framework. In the last part of this thesis, I focus on the logico-mathematical case.

4 Scepticism and non-evidential warrant

In the last chapter, I committed myself to the **(Traditional Epistemic Project)**, and argued that we need to explain how we can acquire *rationally claimable knowledge* of propositions we ordinarily think we possess knowledge of. The **(Traditional Epistemic Project)** is most naturally combined with a certain form of *foundationalism* — that relevant areas of knowledge rest on a small class of basic beliefs and belief-forming methods, which can be used to acquire all the knowledge in the area by a finite, step-by-step process. For any foundationalist, the question arises of what the basic beliefs and belief-forming methods are, and under what conditions they generate knowledge. I adopt conservativism, i.e. the claim that it is a precondition for the generation of evidential warrant that the subject possesses certain supporting warrants. This makes it hard to see how the internalist can avoid certain radical forms of *scepticism*. I argue that, in order to avoid scepticism, every access internalist needs to invoke a notion of an internalistic warrant by default, i.e. a notion of an internalistic warrant one can possess without having done any prior evidential work. I then render this response to scepticism explicit by endorsing Wright's notion of entitlement of cognitive project.

4.1 Cognitive projects, belief-forming methods, and foundationalism

4.1.1 Cognitive projects and their belief-forming methods

I begin by drawing some relevant distinctions. An *area of cognitive enquiry* is a piece of reality one can acquire knowledge of. A *cognitive project* is an epistemic agent's project of acquiring knowledge in a certain area of cognitive enquiry.[83] Three cognitive projects are relevant in what follows:

- The *external-world project* — the project of discovering ordinary external-world facts, i.e. facts about medium-sized objects in our usual cognitive environment. For example: that there is a table in front of me, that I have two hands, etc. We are all engaged in this project, and the possibility of successfully executing this project is essential to our lives.

[83] It is not clear to me whether Wright would accept this formulation, because it sometimes seems as if knowledge is not the most important notion for him, but rather "full justification" (see e.g. Wright 1991, p.88).

- The *arithmetical project* — the project of discovering the arithmetical facts. For example: solving the question of how many primes there are between 100 and 200, what the product of 123 and 234 is, and whether Fermat's Last Theorem is true. This project is not only interesting on its own (i.e. for those with an interest in pure mathematics), but also because it is importantly entangled with everyday tasks such as determining whether one carries enough money for one's purchase, and more sophisticated projects of finding out about the external-world, such as astrophysics.

- The *logical project* — the project of discovering the logical facts, i.e. the facts of what, in general, deductively follows from what. For example: that a certain basic deductive rule of inference such as Modus Ponens (MP) is valid; that a certain longer deductive argument is valid; tautologies. Just as the mathematical project, the logical project is not only pursued for its own sake, by those with an interest in formal logic. The logical project is important for everyone. The possibility of successfully engaging in the logical project is a precondition of the complete success of all projects in which deductive inference plays a role in extending knowledge.

These cognitive projects can be canonically pursued by the successive application of a finite number of suitable basic belief-forming methods. A *belief-forming method* (BFM) is any procedure or method a subject can carry out to form new beliefs.[84] For example:

- The perceptual BFM of forming the belief that p whilst undergoing a visual experience as of p.

- The deductive BFM of forming the belief that q on the basis of the belief that p and the belief that $p \to q$ (for any p, q), or — in other words — the method of drawing inferences in accordance with MP.

BFMs can be executed consciously and unconsciously. When I speak of belief-formation below, I always mean *conscious* belief-formation. I call a belief resulting from applying a BFM in a certain situation the *target belief* of the BFM. BFMs also have characteristic *sources*: states on which the belief-forming process is based, and which bear on the content of the belief to be formed. For example:

[84] This piece of terminology is due to Enoch and Schechter (2008).

- The experience *as of* p is a source for the BFM of forming a belief that p on the basis of an experience *as of* p.
- The beliefs featuring as premises of deductive inferences are sources of these inferences.

An *inferential* BFM (or: *inference*) is a BFM whose sources are only beliefs. For example: applying MP. A *non-inferential* BFM is any BFM that is not inferential. For example: forming the belief that p on the basis of a visual experience as of p.

Of course, the proper aim of a BFM is not only the formation of a target belief, but the formation of a *fully warranted* target belief. For a cognitive project can only succeed if the relevant target beliefs can become warranted in a way that renders them knowledegable.

Every cognitive project has *characteristic* BFMs — BFMs which are such that the project in question would be radically impaired if the BFM in question would not normally deliver fully warranted beliefs. In this sense, forming simple external-world beliefs on the basis of visual experiences is a characteristic BFM of the external-world project. Drawing inferences in accordance with MP is not only a characteristic BFM of the logical project, but also of the mathematical project. Moreover, a BFM is *basic* if and only if it is not reducible to a chain of other BFMs. Presumably, all basic BFMs of a project are characteristic of the project.

I say that a BFM M *confers a warrant W on S's belief that p* if and only if S acquires W for the belief that p in virtue of properly executing M. In case M is based on a source S, any warrant conferred by M will called an *evidential warrant*, and S may be called (M-)*evidence* for p.

A special case is warrant-transmission. An inference from $P_1, ..., P_n$ to Q *transmits a warrant W for a subject S* just in case S acquires a warrant for the belief that Q by virtue of drawing the inference from W-warranted $P_1, ..., P_n$. An inference from $P_1, ..., P_n$ to Q *transmits a warrant of type T for a subject S* just in case S acquires a warrant of type T for the belief that Q by virtue of drawing the inference from T-warranted $P_1, ..., P_n$. A transmitted warrant may be called an *inferential warrant*. A warrant that is only conferred, but not transmitted, may be called a *non-inferential warrant*.

The concept of transmission of warrant is first introduced in (Wright 2003). It differs substantially from the well-known concept of *closure* of warrant, which we already encountered in 3.1.2. If a warrant is transmissible over a certain inference, a subject can *learn of* the conclusion by drawing the inference (Wright 2003, p. 57). For a warrant of type T being closed over a certain consequence relation, it is not required that the warrant one possesses for the conclusion can be generated *by virtue of* drawing the inference. It is merely required that *there is* some warrant of type T for the conclusion. The transmissibility of warrant over an inference entails closure, but not vice versa:

> Closure will hold but transmission may fail in question-begging cases—cases where there is warrant for the premises in the first place only because the <u>conclusion</u> is antecedently warranted. (Wright 2003, p. 57, author's emphasis)

We will encounter such cases below. Note that the concept of conferring warrant and the concept of transmission of warrant are subject to a two-fold relativization. The relativization to types of warrant is motivated by the **(Warrant Pluralism)** principle from 3.2.2. It is plausible that the conditions for transmission or conferral of warrant are sensitive to the type of warrant in question. For example, it is plausible that basic perceptual BFMs confer simple externalistic warrants in every non-sceptical scenario, whereas we will see below that additional conditions have to obtain for such BFMs to confer reflectively accessible warrant.

The relativization to subjects is required because abilities and circumstances matter. For example, simple deductive inferences might always transmit inferential warrant for unsophisticated subjects, but some kind of reflective access might be required for such inferences to transmit inferential warrant for more sophisticated subjects (Boghossian 2001, p.25; Wright 2001, p.70).

Another important feature of warrant-conferral and warrant-transmission is that BFMs cannot only be used to acquire *first* warrants for a newly formed belief, but also to *upgrade* or further support one's warrants for a pre-existing belief. For example, Hero might not have left the flat yet, but have acquired a warranted belief that it is cloudy on the basis of testimony. When S leaves the flat a little bit later, S might look up at the sky and see that it is indeed cloudy. No new belief is formed here. However, the doxastic warrant Hero acquired on the basis of testimony is enhanced by a further, perceptual warrant.

4.1.2 Foundationalism

I return to the question of how to pursue the **(Traditional Epistemic Project)** when it comes to the three projects specified above. How can we explain that we can acquire full internalistic warrants for the beliefs belonging to them? My explanation rests on the following tenet:

> **(Foundationalism)** For all true propositions p that belong to a relevant cognitive project X and that are candidates for justification in the first place, a non-defective epistemic agent can obtain a full warrant for p on the basis of a finite step-by-step application of a small class of warrant-conferring or warrant-transmitting basic X-BFMs.

This idealization reduces the explanation to the question of how basic beliefs can be fully internalistically warranted and how basic-belief forming-methods confer or transmit full internalistic warrant. The restriction to propositions that are candidates for justification in the first place is meant to exclude cases like undecidable propositions.

Moreover, note that I am not interested in actual justification, but just in *canonical* justification. For example, I ignore the possibility of acquiring knowledge by testimony because we can — in principle — do without it.

I do not have a general argument to the effect that the belief-forming processes underlying the projects above are systematizable in accordance with **(Foundationalism)**. I think that this claim can be made plausible by examples.

4.1.3 Two basic belief-forming methods

Two basic BFMs are relevant in this chapter.

Visual perception A lot of ordinary external-world knowledge in the sense above can be canonically based on perceptual experience. This makes it relatively easy to determine candidates for the basic BFMs of the external-world project: basic *perceptual* BFMs. For example: the BFM of forming the belief that P whilst undergoing a visual experience as of P. Using this BFM, an epistemic agent will be able to form a large variety of beliefs about the external world, assuming the agent possesses relevant concepts. However, there are two ways to conceive of this BFM. In particular, we can conceive of the belief-forming process as a one-step or a two-step process. According to the first model, an external-world belief is formed directly on the basis of a visual experience. According to the second model, a visual experience first gives rise to a belief about the existence of a certain experiential seeming, from which an external-world belief is then *inferred*. In the latter case, two basic BFMs are required: a non-inferential introspective BFM of forming a belief that one has a visual experience as of p, whose source is an experiential seeming, and another, inferential BFM that allows us to infer p from "I have a visual experience as of p". For example: "I am, right now, undergoing a visual experience as of my two hands. Therefore, I have two hands". All examples involving perceptual experience below should be understood as being about visual perception.

Modus Ponens *Modus Ponens* (MP) is the inferential BFM of forming the belief that q on the basis of the beliefs that p and that $p \to q$. I do not want to engage in complications with ordinary language and thus regard "$\to$" as the material conditional, as opposed to the English "if". MP is plausibly regarded as a basic deductive (or basic logical) BFM.

4.2 Conservativism

That basic perceptual BFMs generate or transmit warrant is subject to certain conditions. Let me explain.

4.2.1 Conservativism about Perception

Wright (2007*b*) argues that, in order for an epistemic agent to acquire an evidential warrant by virtue of properly executing a basic perceptual BFM, the agent already needs to be in a *conducive informational context*. This means that the agent already needs to be warranted to *accept* a range of propositions ensuring the good standing of the BFM, where *acceptance* is an attitude excluding both doubt and agnosticism (Wright 2007*b*, p. 27).

Here is why. According to Wright (2007*b*), one of Quine's important insights is that evidence is information-dependent in the sense that evidential achievement takes place in a certain "informational context" — a context of background acceptances determining what counts as evidence for what in the first place.

This is very plausible. Clearly, experiences can support all sorts of (different) propositions in the context of different background acceptances. The darkness of a room can indicate nighttime, but also the closed state of the window shutters. In case I just entered the building in full daylight I will not regard (and *should* not regard) the darkness as evidence for the former, but for the latter. If I know that the windows in this house do not have shutters, I will take (and *should* take) the darkness as evidence for the former, and not the latter. Depending on my geographical location and the time of the year, the sound of gunfire can indicate nearby fighting, but it could just as well indicate that Oktoberfest begins. What I take the sound to indicate (and what I *should* take the sound to indicate) will depend on my background beliefs about my current situation in the wider world.

This dependence generalizes. Even my warrant for drawing the inference from "I have, right now, an experience as of a table in front of me" to "There is a table in front of me" depends on my current background information. If I am certain that there is no table in this room, then I will rather take the experience to show that I hallucinate — maybe I infer that the glass of water I just drank contains psychoactive substances. Similarly, I might possess evidence to the effect that I have recently been envatted. In this case, I will rather infer that the computer simulates a vat-world in which I vat-face a vat-table. This shows that the evidence relation itself — what is evidence for what for an epistemic agent — depends on the agent's current background acceptances. And this suggests that in order for perceptual BFMs to generate evidential warrant for external-world propositions, we need to warrantedly accept propositions that make up an informational context in which we can regard an experience as of p as evidence for p.

Note that this is compatible with externalism. An externalist might hold that one needs to possess *externalistic* warrants for relevant background acceptances. In fact, our wide notion of warrant enables us to even incorporate reliabilism into this picture. For one might contend that our warrant to accept all the relevant background information just consists in the reliability of the BFM.[85]

Call the propositions which need to be in an agent's informational context in order for a BFM M to deliver an internalistic warrant for its target beliefs the *presuppositions of M*. According to Wright, the presuppositions of basic perceptual BFMs include:

- "the conditions articulating the general co-operativeness of the prevailing cognitive environment" (Wright 2004*a*, p. 164)

[85]This might imply that the conservative position and what has been called *liberalism* (Pryor 2004) collapse when it comes to crude externalistic warrant. I do not think that this is a bad result.

- "the proper functioning of the relevant cognitive capacities" and
- "[the good standing of] the very concepts involved" (Wright 2004*b*, p. 189)

An example for the first type of presupposition is the non-obtaining of the possibility that the agent in question is a BIV. I will use this presupposition as my primary example for a presupposition of perceptual BFMs.

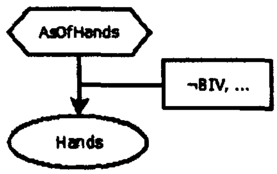
It is an interesting question what exactly a complete set of presuppositions is, and in what sense it depends on the sophistication of the agent. For example, if an agent does not have the concept of a brain-in-a-vat, we cannot demand that the agent needs to possess a warrant to accept the presupposition without precluding the agent from acquiring perceptual warrants, because the agent cannot possibly possess such a warrant to accept it without possessing the relevant concepts, or so the thought goes. Moreover, it seems plausible that presuppositions expand in the sense that a relevant collection of consequences of a presupposition are also presuppositions. For example: if it is a presupposition that the agent is not a brain-in-a-vat, then it is also a presupposition that the agent is not a brain-in-a-vat wearing a hat. I do not know how the relevant notion of consequence should be spelled out, and Wright does not say anything about it, but I will briefly return to this issue later in this thesis. Finally, note that the warrants to accept the presuppositions are *propositional* warrants. It is very implausible that an agent needs to have *entertained* all presuppositions. The emerging position can be summarized as follows:

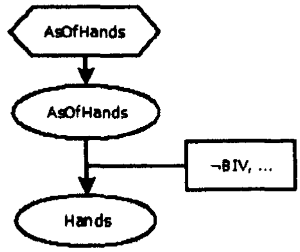
> **(Conservativism about Perception)** It is a necessary condition for perceptual BFMs to confer evidential warrants on their target beliefs, that the agent in question already possesses (propositional) warrants to accept all of their specific presuppositions.

Corresponding to the inferential version of liberalism about perception, there is also an inferential version of conservativism about perception:

> **(Inferential Conservativism about Perception)** It is a necessary condition for inferences from "I have an experience *as of p*" to *p* to transmit evidential warrant, that the agent in question already possesses warrants to accept all of their specific presuppositions.

Let propositions in boxes stand for propositions that are propositionally warranted, and let "$\neg BIV$" stand for the proposition that the agent is not a brain-in-a-vat. We can display the proposed structure of justification as follows:





Conservativism about perception *Conservativism about perception (inferential)*

The inferential version includes the application of an additional, non-inferential BFM, generating the belief that there is an experiential seeming as of the agent's two hands. Wright does not say anything to the effect of whether he believes that this BFM also has presuppositions, but it is very plausible that it does. For example: the coherence of the concepts used to describe the experience, and the reliability of introspection.

I accept both versions of conservativism about perception, and I will assume them throughout this thesis. In addition to the (Quinean) motivation above, there is a whole variety of arguments for conservativism that I cannot discuss here. Wright (2007*b*) presents five such arguments. I will come back to one of these arguments in the next chapter: that only conservativism provides us with a good diagnosis of what is wrong with Moorean arguments (and other question-begging arguments).

4.2.2 Deduction

Consider a particular instance of MP. Considerations similar to those above suggest that in order for it to transmit evidential warrant to its conclusion, a sufficiently sophisticated agent already needs to be in a proper informational context, including the proposition that the MP step is *valid*.[86] The validity of a deductive inference is among its presuppositions. Wright (2001, 2004*a*, 2007*b*) also seems to hold this position.[87] We can summarize it as follows:

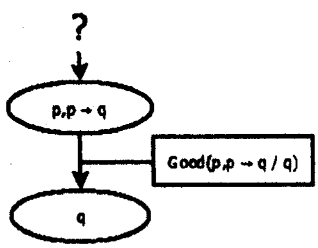
[86]The reader can think of validity as necessary truth-preservation here.

[87]In any case, he holds conservativism for internalistic warrant. The considerations from the next chapter onwards just require this less general conservative claim.

(Conservativism about Deduction) It is a necessary condition for a logical inference to transmit evidential warrant to its target belief, that the agent already possesses a warrant to accept the validity of the inferential step in question.

As I said, one motivating argument for this position is analogous to the argument in the perceptual case. Surely, we cannot *claim* the conclusion of an inference to be warranted without also *claiming* that the inferential step is warranted — i.e. without being able to claim that there is some warrant for the goodness of the inferential step — and this suggests that it must be a condition for an inference to transmit warrant that we already possess a warrant for validity. Of course, the question arises of what "validity" means here. In what follows, I assume a version of **(Conservativism about Deduction)** where "validity" is read as "necessary truth-preservation".[88]

Again, the position is compatible with externalism. An externalist might hold that the warrant in question just consists in the de facto validity of the inferential step. We can depict the envisaged structure of justification for the case of MP as follows, where "Good" stands for what one takes to be the correct statement of the goodness of the rule (e.g. necessary truth-preservation):



Conservativism about deduction

4.3 Scepticism, circularity, and non-evidential warrant

Everyone endorsing **(Relevance Internalism)** and conservativism needs to appeal to a notion of an internalistic warrant by default — an internalistic warrant one does not have to earn by doing any evidential work — for the presuppositions of basic BFMs.

[88]One might think that it is too much to demand a warrant for anything stronger than truth-preservation. For a particular inference from true premises to serve the aim of extending true belief, nothing more than truth-preservation is required, or so the thought goes. This motivates the following weaker position:

> **(Weak Conservativism about Deduction*)** It is a necessary condition for a logical inference to transmit internalistic warrant to its target beliefs, that the agent already possesses a warrant to accept that the inferential step in question is truth-preserving.

This issue is tricky, and merits further research. One reason why it is difficult to endorse this version of conservativism about deduction is that it becomes problematic in contexts including suppositions.

I first present two arguments for the claim that every conservative needs to endorse a notion of warrant by default. The first argument shows that without such a notion, we face (first-order) scepticism.[89] The second argument establishes that, in the course of arguing that we possess warrants for the target beliefs of basic perceptual and logical BFMs, we need to appeal to the possession of a non-evidential notion of warrant, because otherwise the argument will be premise circular.

After that, I argue that everyone endorsing **(Relevance Internalism)** needs to hold that the possession of this non-evidential notion of warrant is claimable on the basis of apriori reasoning and self-reflection, i.e. that it is an internalistic notion of warrant, because otherwise the **(Cartesian Constraint)** is violated.

4.3.1 A sceptical challenge

For the conservative, the question arises of how to avoid scepticism: for in relevant cases it is hard to see how a subject S can ever *enter* the informational contexts required for the generation of evidential warrant.

Consider the case of perception. How could S acquire a first warrant for the non-obtaining of the brain-in-a-vat hypothesis? Prima facie, there seem to be two options: firstly, the warrant could be acquired by perception (or some line of reasoning based on perception). Secondly, it could be acquired by some line of apriori reasoning. However, both options are problematic. The first option leads to plain epistemic circularity: if the conservative diagnosis is correct, we already needed a warrant for the non-obtaining of sceptical scenarios in order to be able to acquire a warrant for the non-obtaining of sceptical scenarios by perception. The second option, on the other hand, would require initial motivation as well as a sustained defence. What is at issue is that there are certain lines of apriori reasoning — such as inferences to the best explanation or considerations about the apriori likelihood of being in the good case which can serve the refutation of *all* relevant sceptical scenarios. The burden of proof lies on the side of the anti-sceptic.

The situation in the basic logical case is similar. How could S acquire a first warrant for the validity of MP? Assuming that rational intuition is not an option, we seem to be forced to justify MP inferentially. However, we can safely assume here — without loss of

[89] This argument is very similar to Wright's presentation of "Humean scepticism" in (Wright 2004*b*).

generality — that every inferential justification of the validity of MP will make use of MP.[90] However, since making epistemically effective use of MP already requires an antecedent warrant for its validity, we are stuck again: given our assumptions, the evidential route seems blocked.

Thus, in both cases there does not seem to be an evidential route to first warrants for presuppositions. Now, given this predicament, how can the conservative avoid scepticism? The crucial observation made in (Wright 2004*b*) is that a further assumption is needed in order to move from the tentative conclusion that there is no evidential warrant for the relevant presuppositions to the sceptical conclusion that there is no warrant for them at all, namely that evidential warrants are the only plausible candidates for warrants for presuppositions. Wright (2004*b*) argues that this assumption has to be rejected because the relevant presuppositions can be warranted *non-evidentially*. If this is correct, the sceptical argument as stated is unsound. It only shows that we cannot possess *evidential* warrants for the presuppositions.

4.3.2 Circularity at the second-level

There is a related argument that applies to the second-level of justification. We begin by noting that it is desirable to be able to provide second-order justifications — arguments to the effect that we are justified in believing certain propositions. In the context of **(Foundationalism)**, this requires arguments to the effect that basic BFMs confer justification upon their target beliefs. I consider the perceptual and the deductive case in turn, and argue that the possibility of providing such justification rests on the possibility to claim that we possess non-evidential warrants at the basic level.

Perception Consider the case of Hero having formed the belief that he has two hands (Hands), on the basis of the corresponding visual perception. How could Hero justify that he acquired a warrant for Hands? It is plausible enough that Hero's argument will proceed from two premises: that Hero has the relevant perceptual seeming, and that Hero possesses a warrant for the presuppositions of visual perception. Now, it is unproblematic for Hero to claim that he has the relevant perceptual seeming. This will be an instance of self-knowledge. However, the question arises on what basis Hero can claim a warrant for

[90] In any case, there will be some basic rule of inference for which the circularity problem arises.

the presuppositions. Let us use the non-obtaining of the BIV hypothesis as an example ($\neg BIV$). On what basis can Hero claim an evidential warrant for it? That is: on what basis can Hero claim $W(\neg BIV)$, where "$W(p)$" stands for "I possess an evidential warrant to accept p"?

The warrant for $\neg BIV$ could either be acquired by some apriori line of reasoning, or on the basis of experience. We can ignore the former possibility for the same reasons as above. Because of **(Foundationalism)**, this means that Hero needs to argue for $W(\neg BIV)$ by arguing that he acquired this warrant by a belief-forming process that included basic perceptual BFMs. However, in order to do so he needs to be able to claim, among other things, that he already possesses warrants for their presuppositions, including $\neg BIV$. Thus, Hero needs to cite $W(\neg BIV)$ as a premise in his argument for $W(\neg BIV)$. The argument is premise circular. The envisaged higher-order justification is useless: it cannot possibly generate a warrant for $W(\neg BIV)$.

Now, one might save Hero's argument for $W(\neg BIV)$ by claiming that the notion of warrant whose possession is claimed by the conclusion of the argument — call it W_C — is a different notion than the one whose possession is claimed by the premise W_P. In particular, $W_P(p)$ must not entail $W_C(p)$. And prima facie, there are two possibilities of what W_P could be, if W_C is evidential warrant: it could either be an externalistic warrant or an internalistic warrant one can possess without having done any prior evidential work. For example: $W_P(p)$ might be the simplest type of externalistic warrant — that p is true. No evidential work is required for one to possess this warrant, so the argument would not be circular anymore.

Deduction The same problem arises in the case of deduction. On what basis can Hero claim an evidential warrant for the conclusion of any inference, i.e. on what basis can Hero argue for $W(q)$ if he has carried out an MP step with warranted premises p and $p \to q$, where "$W(p)$" stands for "I possess an evidential warrant to accept p"?

Clearly, in order to be able to argue for $W(q)$, Hero needs to be in a position to claim two things. Firstly: that he possesses (evidential) warrants for the premises. And secondly: that he possesses a warrant for the claim that the inference is good in some sense (e.g. necessarily truth-preserving). Let G stand for the proposition that the inference is good in the relevant sense. On what basis can Hero claim a warrant for G?

Suppose that the warrant that Hero needs to claim here is an evidential warrant, i.e. a warrant that has been acquired (or can be acquired) by some other BFMs. Prima facie, there are three possibilities of how Hero's warrant for G could be acquired:

- By means of empirical BFMs.
- By using rational intuition.
- By means of deductive reasoning.

The first and the second option are implausible from the start. In any case, the burden of proofs lies on the side of someone who wants to endorse these options. The third option, however, leads to a circularity problem again.

In particular, we can safely assume here that the argument for G will be rule-circular, i.e. it will make use of at least one MP-step. Now, in order to argue that he acquires W for G by virtue of going through this argument, Hero needs to claim, among other things, that he already possesses warrants for the presuppositions of all the argument's inferential steps, including the MP steps. Thus, Hero needs to cite $W(G)$ as a premise in his argument for $W(G)$. The argument is premise circular. The envisaged higher-order justification is useless: it cannot possibly generate a warrant for $W(G)$.

Just as in the perceptual case, one might try to save Hero's argument for $W(G)$ by claiming that the notion of warrant whose possession is claimed by the conclusion of the argument — call it W_C — is a different notion than the one whose possession is claimed by the premise — call it W_P. In particular, $W_P(p)$ must not entail $W_C(p)$. So Prima facie, there are two possibilities of what W_P could be, if W_C is evidential warrant: it could either be an externalistic or an internalistic warrant one can possess without having done any prior evidential work. For example: $W_P(G)$ might be the simplest type of externalistic warrant — that G is true (i.e. that the rule is de facto necessarily truth-preserving). No evidential work is required for one to acquire this "warrant", so the argument is not circular anymore.

4.3.3 Internalistic warrant by default

However, saving the second-order arguments from premise circularity by just appealing to externalistic warrants is unsatisfactory. For this is tantamount to making ineliminable appeal to external-world conditions in our second-order arguments, which is in tension with the **(Cartesian Constraint)**. Rather, at some point we need to appeal to non-evidential warrants whose possession is available just on the basis of apriori reasoning and introspection. We need to be able to appeal to *internalistic* warrants by default.

Consider the perceptual case. If we have to claim, at some point, that $W(\neg BIV)$, and "W" denotes an externalistic notion of warrant, then the claim involves a substantial claim about the external-world. According to the externalist, this is all we can do. In particular, we cannot eliminate such appeal to external-world propositions (here, "eliminate" is of course to be understood as in 3.3.3). Thus, the **(Cartesian Constraint)** is violated.

Consider the logical case. If, at some point, we need to claim that $W(G)$, and "W" denotes an externalistic notion of warrant, then the claim involves a substantial claim about the external-world. Because of the **(Cartesian Constraint)**, we need to eliminate this claim. However, this means that we either need to be able to establish $W(G)$ by some apriori line of reasoning, without making ineliminable appeal to external-world considerations (which amounts to the possession of an evidential internalistic warrant), or we need to possess a non-evidential internalistic warrant for $W(G)$. As to the first option, it is hard to see how there could be such an argument that does not lead to circularity at some point. Consider, for example, the radical externalist proposal that $W(G)$ is idential to G. In this case, we need to justify G before we can run the original second-level argument, i.e. justify our argument for G. Appeal to G would not be ineliminable, and the **(Cartesian Constraint)** would be violated. The second option is just conceding my main point: that we need to appeal to an internalistic notion of warrant by default.

Ergo: in both cases we need to appeal to a non-evidential notion of internalistic warrant (or: internalistic warrant by default). Suppose there is a non-evidential type of warrant E that can be claimed just on the basis of apriori reasoning and self-reflection. Then we can solve the problem of higher-order justification by claiming E for the presuppositions of relevant BFMs. For example: Hero could claim $W(Hands)$ on the basis of the claims that he has the relevant experience and $E(\neg BIV)$. The question of how to *acquire* a warrant for $\neg BIV$ does not arise, since E is non-evidential. We just have to claim that the conditions for E are met, but, by assumption, these will be available on apriori reasoning and self-reflection.

Appealing to such a notion will not only solve the problem of second-order justification outlined in 4.3.2, but also solve the simple sceptical problem outlined in 4.3.1. I now turn to an explication of such a notion which I will endorse in this thesis — Wright's notion of entitlement of cognitive project.

4.4 Entitlement of cognitive project

Providing an epistemology in accordance with **(Relevance Internalism)** requires arguing for the following two claims:

- That there is a non-evidential, reflectively available type of warrant which is such that possession of this type of warrant for the presuppositions of a BFM enables one to acquire fully internalistically warranted beliefs.[91]

- That we (can) actually possess such warrants for the presuppositions of relevant basic BFMs.[92]

Wright (2004*b*) attempts to establish both claims for his notion of *(epistemic) entitlement*. Entitlements are construed as non-evidential but genuinely epistemic types of warrants *to trust* in a proposition.[93] The conditions for a proposition to be an entitlement are construed in such a way that they are available on apriori reasoning and introspection, and that the conditions are met for the presuppositions of basic perceptual and deductive BFMs. Most important for the two cases at hand is the notion of *entitlement of cognitive project*. Let me explain.

Call a proposition P a *Wright-presupposition* of a cognitive project if and only if "to doubt P (in advance) would rationally commit one to doubting the significance or competence of the project" (Wright 2004*b*, p.191). It is very plausible that the presuppositions of basic BFMs of cognitive projects — in the sense of presupposition defined in 4.2 — are also Wright-presuppositions of these projects. For these presuppositions are the conditions ensuring the good standing of the relevant BFMs, including their reliability. Doubting

[91]That we explain how we can acquire such warrants is required by **(Relevance Internalism)**. Note that conservativism only imposes a *necessary* conditions for warrant-generation, so arguing that there are suitable warrants only rebuts scepticism, but does not yet deliver an explanation.

[92]At least for those presuppositions to which the circularity considerations above apply. There might some presuppositions that can be justified, and for which we do not have to endorse warrants by default.

[93]Maybe also to trust in a *rule*. Wright is not entirely clear here. I will only make use of entitlements to propositions in this thesis, so I can ignore this case.

that such a condition obtains rationally commits one to doubting that these BFMs can be used to extend knowledge. So doubting that they are met commits us to doubting the significance of the project. For example:

- Doubting that I am *not* a BIV commits me to doubting that (basic) perceptual BFMs can be used to extend ordinary external-world knowledge, and hence commits me to doubting the significance of the external-world project, since the project cannot be seriously pursued without such methods. So that I am not a BIV is both a presupposition of basic perceptual BFMs and a Wright-presupposition of the external-world project.

- Doubting that MP is valid commits me to doubting the significance of extending knowledge by using MP.[94] So it commits me to doubting the significance of any project in which MP is an essential means to extend knowledge. Because logic is used everywhere, that a certain basic logical rule is valid is both a presupposition of (projects of extending knowledge using) the respective BFMs, and also a Wright presupposition of (substantial parts of) almost every cognitive project.

Now, it appears that *every* cognitive project has (Wright-)presuppositions which cannot be justified without invoking further (Wright-)presuppositions which are in no better epistemic standing than the (Wright-)presupposition to be justified. Consider the case of forming ordinary external-world beliefs on the basis of visual perception. Assuming there is no apriori argument for the non-obtaining of Cartesian scenarios, we will have to use other perceptual BFMs to justify them. But the project of justifying the non-obtaining of Cartesian scenarios again has (Wright-)presuppositions which are in no better epistemic standing: in fact, we saw above (4.3.1) that it is plausible that any project of justifying the non-obtaining of Cartesian scenarios has the *same* (Wright-)presuppositions as the project of forming ordinary external-world beliefs. Thus, no epistemic progress is possible with respect to such (Wright-)presuppositions. This point generalizes. We saw that the same holds for the logical case. It is to be expected that this generalizes to all cognitive projects.

[94]One might think that it does not yet commit me to doubting the significance of extending knowledge by using MP, because I might still have a warrant for the claim that MP is truth-preserving. As a response, one might say that validity is the best explanation for logical rules being truth-preserving, and doubting that MP is valid also commits me to doubting that MP is truth-preserving for this reason. I owe this point to Robert Williams.

We already saw that this yields a pressing sceptical challenge to which we have to respond. Wright's suggestion is that the key to a response lies in the fact that this challenge tells us something important about the nature of (internalistic) warrant and the structure of justification, which, once we got it right, avoids scepticism. For:

> If there is no such thing as a process of warrant acquisition for each of whose specific presuppositions warrant has already been earned, it should not be reckoned to be part of the proper concept of an acquired warrant that it somehow aspire to this—incoherent—ideal. Rather, we should view each and every cognitive project as irreducibly involving elements of adventure—I have, as it were, to take a risk on the reliability of my senses, the conduciveness of the circumstances, etc., much as I take a risk on the continuing reliability of the steering, and the stability of the road surface every time I ride my bicycle. (...) warrant is acquired whenever investigation is undertaken in a fully responsible manner, and what the [sceptical] paradox shows is that full epistemic responsibility cannot, per impossibile, involve an investigation of every presupposition whose falsity would defeat the claim to have acquired a warrant. (Wright 2004*b*, pp. 190f)

One way to extract a general anti-sceptical strategy from this paragraph is as follows: we can read Wright as conceiving of the sceptical conclusion as being paradoxical — in the sense that it is a highly undesired conclusion which is arrived at on the basis of prima facie plausible assumptions — and taking this as a reason to reject the (prima facie plausible) assumptions about our requirements with regard to the warrantability of presuppositions.[95] In particular, "it should not be reckoned to be part of the proper concept of an acquired warrant" that a warrant for such presuppositions has to be *evidential.*

There are at least two ways of understanding the phrase "not to reckon X to be part of the proper concept of an acquired warrant". We can read it as a *meta-epistemological demand* to *revise* or *construct* our concept of an acquired warrant in such a way that the possession of such a warrant does not require X, and we can read it as an *epistemological insight* about our actual, ordinary concept of an acquired warrant: namely that the posses-

[95] Another case is the so-called naïve conception of truth. Maybe it is another "incoherent ideal" which leads undesired conclusions — in this case: contradictions. And maybe the right reaction is just to reject some of the prima facie plausible assumption, e.g. the so-called naïve T-rules.

sion of such a warrant does not require X (see also Jenkins 2007, p. 30). Which option is the better one? Since I am not sure whether there is anything like a pre-theoretic concept of warrant — and this also makes it somewhat problematic to conceive of the situation as a genuine paradox — I shall read the passage as a suggestion of how to construct one's notion of warrant in the first place. Jenkins (2007, p. 30) deems this way of reading the passage to be problematic, because there is a risk of "changing the subject". However, since I do not believe that there is anything like a pre-theoretic concept of warrant, I do not think that there is any prior subject we could change. The crucial question, I believe, is what we should say about the pre-theoretic notion of *knowledge*, because the aim of the **(Traditional Epistemic Project)** is not just the vindication of some claimable warrant, but of ***full*** warrant and claimable knowledge. I come back to this question. It will turn out that it is not too important how exactly the notion of knowledge looks like, but whether there is a notion of warrant that allows us to rationally claim knowledge on the basis of entitled presuppositions (or, in other words: a notion of warrant that allows us to possess internalistic warrants for knowledge claims).

The argument suggests that we should include a non-evidential notion of warrant in our epistemological framework, which applies to Wright-presuppositions for which the evidential route is closed, i.e. where presuppositions of valuable projects cannot be justified without invoking further presuppositions which have an epistemic status no better than those of the original presuppositions. These warrants are Wright's *entitlements of cognitive project.* They are construed as follows:

> If a cognitive project is indispensable, or anyway sufficiently valuable to us—in particular, if its failure would at least be no worse than the costs of not executing it, and its success would be better—and if the attempt to vindicate (some of) its presuppositions would raise presuppositions of its own of no more secure an antecedent status, and so on ad infinitum, then *we are entitled to*—may help ourselves to, take for granted—the original presuppositions without specific evidence in their favour. (...)
>
> wherever we need to carry through a type of project, or anyway cannot lose and may gain by doing so, and where we cannot satisfy ourselves that the presuppositions of a successful execution are met except at the cost of making further

presuppositions whose status is no more secure, *we should—are rationally entitled to—just go ahead and trust* that the former are met. (Wright 2004*b*, p. 192, own emphasis)

Trust will rule out doubt and agnosticism. It is a *positive attitude*:

> it follows immediately that if acceptance of such a presupposition is to be capable of underwriting rational belief in the things to which execution of the project leads, it has to be an attitude which excludes doubt. If there is entitlement of cognitive project, it has to be an entitlement not merely to act on the assumption that suitable presuppositions hold good, but to place trust in their doing so. (Wright 2004*b*, p. 193)

Thus, *putting trust in P* is stronger than merely *acting on the assumption that P*. This is important, because doubting the presuppositions would commit us to doubting the relevance of relevant projects, and agnosticism presumably leads to agnosticism about their relevance.

Why is the relevant doxastic attitude *trust*, and not *belief*? Wright's reason for this is that the concept of belief might be too tightly connected to the possession of evidence. For the reasons sketched above, the possession of evidence is impossible in the relevant cases (Wright 2004*b*, II). Thus, trust is a doxastic attitude stronger than "acting on the assumption that", but a doxastic attitude weaker than belief. I will come back to issues surrounding belief and evidence in the next chapter.

Wright eventually arrives at the following conditions for entitlement of cognitive project. A proposition P is an entitlement of some cognitive project if P is a Wright-presupposition of the project and:

> (i) We have no sufficient reason to believe that P is untrue
>
> And (ii) the attempt to justify P would involve further [Wright-]presuppositions in turn of no more secure a prior standing ... and so on without limit; so that someone pursuing the relevant enquiry who accepted that there is nevertheless an onus to justify P would implicitly undertake a commitment to an infinite regress of justificatory projects, each concerned to vindicate the [Wright-]presuppositions of its predecessor (Wright 2004*b*, pp. 191f).

Presuppositions of basic BFMs are Wright-presuppositions in which condition (ii) is met. In order to know whether condition (i) is met as well, we need to know what sufficient reasons are. Most importantly, "reasons" could be read as external or internal reasons. For example, Pedersen (2009*b*, section 5) distinguishes between *metaphysical reasons* and *epistemic reasons*.

Whether there are metaphysical reasons to believe something is determined by the external world. Even the falsity of P *might* be a metaphysical reason not to put trust in P. On an epistemic conception of reasons, a subject needs to possess at least some (internal) evidence against P in order to have a sufficient reason to believe that P is untrue. Wright must have the epistemic conception of reasons in mind, because entitlements are supposed to be internalistic warrants in the sense defined in 3.4:

> entitlements, it appears, in contrast with any broadly externalist conception of warrant, are essentially recognisable by means of traditionally internalist resources—a priori reflection and self-knowledge—and are generally independent of the character of our actual cognitive situation in the wider world. (Wright 2004*b*, p. 210)

Wright claims that both the non-obtaining of Cartesian scenarios (Wright 2004*b*, XI), and the validity of basic logical rules (Wright 2004*a*, IV) are entitlements of cognitive project because the conditions are fulfilled. We already saw that the non-obtaining of sceptical scenarios are Wright-presuppositions. Assuming that validity is the proper presupposition of logical inferential steps, it is implicit in the sceptical considerations above (4.3.1) that condition (ii) is met for validity. It is also implicit in the sceptical considerations above that condition (ii) is met for Cartesian scenarios.

Finally, it seems clear that we do not possess any evidence whatsoever for the obtaining of any Cartesian scenario. And although arguments against the validity of basic logical rules of inference such as Modus Ponens have been articulated (see e.g. McGee 1985), Wright does not take them to be sufficient to doubt the validity of such rules. Note, however, that I am not interested in ordinary MP, i.e. MP for the English "if", but only in MP for the material conditional ("$\rightarrow$"). And we clearly do not have any evidence to the effect that MP for "$\rightarrow$" is *not* valid. So, clearly, condition (i) is fulfilled in the relevant cases as well.

In the logical case, so the thought goes, entitlements for validity ensure that conservativism about deduction does not lead to the sceptical problems above. And in the external-world case, entitlements for all presuppositions of basic perceptual BFMs ensure that conservativism about perception does not lead to the sceptical problems above.

However, Wright makes a stronger claim. Wright claims that, since the validity of the usual basic logical laws is an entitlement, the respective inferences can be used — ceteris paribus — to extend *internalistic knowledge* (Wright 2004b, VIII). The same is claimed for the perceptual case: as long as we possess entitlements for all the relevant presuppositions — which we do — we can *obtain internalistic* knowledge of ordinary statements about the external world by basic perceptual BFMs (see Wright 2004*b*, p. 208).

If this is correct, then entitlements can be endorsed, at least in some cases, to accomplish both tasks required for the **(Traditional Epistemic Project)**. That entitlements of cognitive project are strong enough for these purposes does of course require further argument. First and foremost, we need to explain how exactly the possession of entitlement can serve the acquisition of internalistic knowledge — knowledge we can rationally claim on the basis of apriori reasoning and self-reflection.

4.5 Leaching and Justification Generation

4.5.1 The Leaching Worry

Even if one accepts Wright's argument that we possess a non-evidential warrant as soon as the conditions of entitlement of cognitive project are met, one might still wonder of whether the entitlement theorist can also defend the following claim, which is essential to endorse entitlements in any non-sceptical epistemology:

(Justification Generation) BFMs with entitled presuppositions ceteris paribus confer or transmit full internalistic warrant to their target beliefs.

The worry that this claim is false — the so-called *Leaching Worry* — poses a major challenge to the entitlement theorist. For example: can Hero really claim knowledge on the basis of his experience as of his two hands, if he (only) possesses entitlements for the presuppositions of this BFM?

Consider the case of Hero. The Leaching Worry is based on the suspicion that the non-evidential status of the presuppositions of perceptual BFMs will affect the epistemic status of their target beliefs. If the Wright-presuppositions of a project have entitled presuppositions, so the thought goes, then the epistemic status of all beliefs in the region will be (negatively) affected. All we can do, so the thought goes, is to put rational trust in the beliefs the project generates, because this weak status of mere entitlement "leaches" upwards from the presuppositions to all our beliefs. In particular, so the thought goes, entitled true belief does not amount to (evidential) justification, and knowledge. So the generated beliefs cannot be claimed to be justified, or to be known.[96]

Remember Jenkins's worry that Wright's proposal might just change the subject. The question arises for the notion of knowledge in particular. Maybe the ordinary notion of knowledge does not work in a way that allows for **(Justification Generation)**, and the best we can do is to adopt a new notion — schmowledge — which has the desired property built in. But this would surely be changing the subject.

What can be said in favour of **(Justification Generation)**? First, note that *presuppositions* are not *premises* of the respective BFMs. And whereas it is very plausible that the epistemic status of the conclusion of an argument cannot be stronger than that of its weakest *premise*, it is *not* immediate that the epistemic status of the target beliefs of a BFM cannot be any stronger than the epistemic status of its epistemically weakest *presupposition*. This is to say that, although the following principle is very plausible:

(Limit Principle) The epistemic status of the target belief of a BFM cannot be any stronger than that of its epistemically weakest propositional source.

Or, equivalently:

(Limit Principle') The epistemic status of the conclusion of an argument (or inference) cannot be any stronger than that of its epistemically weakest *premise*.

The following, stronger principle is in need of further argument:

(Strong Limit Principle) The epistemic status of the target belief of a BFM cannot be any stronger than that of its epistemically weakest propositional source *or presupposition*.

[96]It should not be excluded that some kind of externalistic knowledge is acquired. Wright seems to think that as well (Wright 2004*b*, p.206, fn. 23). But we were interested in knowledge we can (rationally) claim.

As an example for how the **(Limit Principle)** works, consider Hero, who infers "There is no evil scientist who recently envatted me", on the basis of the entitled premise "I am not a BIV", using the conceptual truth "If there is an evil scientist who recently envatted me, I am a BIV" as an auxiliary premise. This argument is not capable of generating an epistemic status any stronger than entitlement for its conclusion. The conclusion cannot count as (evidentially) justified, or known. The former assertion might seem a bit puzzling. Surely, beliefs for which we have provided an argument are justified in some sense or other. For this reason, it makes sense to introduce the following notion:

> **(Inferential Entitlement)** Any proposition based on an argument using one or more entitled premises is called an *inferential entitlement.*

If Hero possesses an inferential entitlement for p, he can provide some backup for p, in the sense that he can show that he can put trust in the proposition. So "backup" has to be read in a stronger sense than the kind of backup one acquires by a conditional proof based on mere assumptions, but in a weaker sense than justification on the basis of evidentially warranted premises.

Entitlement theorists need to reject the **(Strong Limit Principle)**, without rejecting the weaker **(Limit Principle)**. How can this be done?

I think the best way to reject the **(Strong Limit Principle)** is to straightforwardly defend **(Justification Generation)**, by giving an explanation of how a warrant for the claims that we possess justification and knowledge can be generated in relevant cases (where the notions of knowledge and justification are the ordinary notions). If this strategy fails, we could adopt the schmoledge strategy. I do not think, however, that we need to endorse this strategy, and will not say more about it.

4.5.2 Two models of Justification Generation

We need to explain how exactly entitlements for presuppositions can render the respective BFMs apt to confer or transmit full internalistic warrant. One strategy is to argue directly for the claim that Hero possesses a propositional warrant for the higher-order claim by virtue of having applied the BFM in question. Another strategy is to describe how Hero might *actually* arrive at a claim to knowledge in simple cases, or how Hero might *actually* find out that he possesses knowledge. If we establish that this can happen, we will have shown that Hero is in a position to claim knowledge, and thus that he possesses a full internalistic warrant. I propose two models for both strategies in the perceptual case, building up on a remark by Wright (2004*b*).

Model 1 - Externalism The first model endorses an externalist picture about the ordinary notion of knowledge. Wright writes:

> to be entitled to trust that, for example, my eyes are right now functioning effectively enough in conditions broadly conducive to visual recognition of local situations and objects is to be entitled to claim that my vision is right now a source of reliable information about the local perceptible environment and is hence at the service of the gathering of perceptual knowledge. (Wright 2004*b*, pp. 207f)

Suppose Hero undergoes an experience as of his two hands, possesses entitlements for all relevant presuppositions, and forms the belief that he has two hands. Using the Hero metaphor, we can extract the following argument:

1. The obtaining of all relevant presuppositions conceptually entails that Hero's current visual experiences are a reliable guide to the truth.

2. **(Reliabilism about Knowledge)** That Hero's current visual experiences are a reliable guide to the truth conceptually entails that a true belief currently acquired by visual perception counts as knowledge.

3. **(Local Closure)** Ceteris paribus, if a set of premises P conceptually entails q, and all $p \in P$ are entitlements, then q is an entitlement.[97]

4. By 1, 2, and 3, we can infer that, since all the presuppositions are warranted by means of entitlement, Hero also possesses an entitlement for the claim that his belief that he has two hands counts as knowledge, if it is true.

5. If Hero has formed the belief that he has two hands on the basis of visual perception, and possesses an entitlement for the claim that his belief counts as knowledge, if it

[97]Given Hero possesses all the relevant concepts. Other reasons for the ceteris paribus clause will become apparent in the following chapters.

is true, then he possesses an entitlement for the claim that his belief that he has two hands counts as knowledge.

6. **(Claimability)** Entitled propositions can be rationally claimed on the basis of apriori reasoning and introspection.

7. By 5 and 6, Hero can rationally claim that he knows that he has two hands just on the basis of apriori reasoning and intropection.

8. By definition, knowledge that is claimable just on the basis of apriori reasoning and introspection is internalistic knowledge. So Hero acquires internalistic knowledge of the proposition that he has two hands by virtue of forming the belief that he has two hands on the basis of a visual experience as of two hands, assuming he possesses entitlements for all the presuppositions.

Note that the availability of this argument shows that **(Relevance Internalism)** is fully compatible with the claim that the ordinary notion of knowledge is an externalistic notion. Moreover, the resulting position will be compatible with **(Conservativism about Perception)**, assuming our notion of warrant is wide enough. For we might claim that the warrant Hero needs to possess in order to acquire knowledge by visual perception just *consists in* the reliability of the BFM in question (i.e. the de facto reliability of Hero's current visual experiences).

My reading of Wright's explanation can also be formulated as a line of reasoning that Hero can go through to realize that he possesses knowledge of the fact that he has two hands — where "realize" means "acquire an inferential entitlement". This provides a first step towards vindicating the **(Traditional Epistemic Project)** in the perceptual case, and brings to light further features of the proposal.

I introduce some shortcuts for convenience:

- "*Hands*" stands for the proposition that Hero has two hands.

- "$B_{Exp}(p)$" stands for "I have formed the belief that p on the basis of my current experience as of p."

- "*GC*" stands for the proposition that Hero is in the good case, i.e. the conjunction of all relevant presuppositions of visual perception.

- "*Rel*" stands for "My current experiences of as of ϕ are reliable".

- "$K(\phi)$" stands for "I know that ϕ"

Hero can go though the following line of reasoning, all of whose premises are warranted (at least one by means of entitlement of cognitive project), in order to acquire an inferential entitlement for the proposition that he knows that he has two hands:

1	$B_{Exp}(Hands)$	Self-knowledge
2	GC	Hero rationally takes this for granted
3	$GC \rightarrow Rel$	Facts about the world
4	$B_{Exp}(p) \wedge Rel \rightarrow K(p)$	Epistemic facts
5	$K(p)$	1, 2, 3, 4, logic

Hero's entitled reflection on the process of visual perception brings two light two potential problems for the approach. Firstly, the argument requires some logical and conceptual — indeed, epistemological — competence on the side of Hero. So the whole approach faces an ***exclusive club problem.*** We only established that conceptually and epistemically competent subjects can possess internalistic knowledge. I think, however, that this bullet is not hard to bite. After all, my aim was merely to establish that the readers of this thesis possess internalistic knowledge, and the readers of this thesis clearly have the required competence.

Secondly, one might worry that the argument cannot be carried out just on the basis of apriori reasoning and self-knowledge. For, although 1,3, and 4 might be available on this basis, is not GC an external-world condition? How does this argument fit together with the **(Traditional Epistemic Project)**?

If the central point I made in 3.3.3 is correct — namely, that we can eliminate appeal to external-world propositions by being able to establish that we possess a warrant for these propositions just on the basis of apriori reflection and introspection — then there is no problem. For we are able to establish, just on the basis of apriori reasoning and self-knowledge, that GC is an entitlement of cognitive project. Once again, we see the central importance of the reflective accessibility of entitlements in the context of the **(Traditional Epistemic Project)**.

To sum up: I showed that **(Justification Generation)** holds in a paradigmatic case. I thus take **(Justification Generation)** as explained, although I will come back to particular instances of it. Note that Wright wants to apply the same line of thought in the logical case as well, for he writes:

> (...) to be entitled to trust in the soundness of a basic inferential apparatus (...) is to be entitled to regard its correct deployment as serving the generation of proofs and hence, since what is proved is known, to be entitled to claim knowledge of the products of reasoning in accordance with it. (Wright 2004*b*, p. 208)

A small modification of the argument above establishes **(Justification Generation)** for deductive rules whose presuppositions are entitled. Just replace reliability with validity. Hero can then use the conditional that, if the premise is known, and the inference is valid, the conclusion will be known.

Model 2 - Internalism Suppose that reliabilist conditions are not sufficient for knowledge in the relevant cases. Then one cannot establish that one knows just on the basis of the claim that one is undergoing a reliable experience. Because of **(Conservativism about Perception)**, one then also needs to be able to claim that one possesses a warrant for the obtaining of the presuppositions of perception, and these warrants cannot be understood as being identical to the obtaining of the presuppositions anymore. This holds for internalistic notions of warrant in particular. Let us assume that it is sufficient for knowledge in the relevant cases that one undergoes the experience and that one possesses an entitlement for the presuppositions. We can still explain **(Justification Generation)**. Here is how.

Again, suppose that Hero undergoes an experience as of his two hands, possesses entitlements for all relevant presuppositions, and forms the belief that he has two hands. Using the Hero metaphor, we can extract the following argument:

1. **(Internalistic notion of knowledge)** That the presuppositions of visual perception hold, and that Hero possesses an entitlement for them conceptually entails that a true belief currently acquired by visual perception counts as knowledge.

2. **(Local Closure)** Ceteris paribus, if a set of premises P conceptually entails q, and all $p \in P$ are internalistically warranted, some of them by means of entitlement, then q is an entitlement.[98]

3. By 1, 2, and 3, we can infer that, since all the presuppositions are warranted by means of entitlement, and these entitlements are accessible to Hero on the basis of self-knowledge and apriori reasoning (which entails that Hero possess an internalistic warrant to believe them), Hero also possesses an entitlement for the claim that his belief that he has two hands counts as knowledge, if it is true.

4. ...

The remainder of the argument is identical to the argument above. Just as above, the explanation can also be formulated as a line of reasoning that Hero can go through to realize that he possesses knowledge of the fact that he has two hands. I use the shortcuts introduced above. In addition, let "$E(\phi)$" stand for "I possess an entitlement for ϕ".

1	$B_{Exp}(Hands)$	Self-knowledge
2	GC	Hero rationally takes this for granted
3	$E(GC)$	Accessibility of entitlements
4	$B_{Exp}(p) \wedge GC \wedge E(GC) \rightarrow K(p)$	Epistemic facts
5	$K(p)$	1, 2, 3, 4, logic

I think both models are suitable for the entitlement theorist. Note that both models only explain how we can possess entitlements for knowledge claims, and how we can rationally claim knowledge; they do not establish that we can know that we know. In the next chapter, I argue that we cannot claim to know p if we possess an entitlement for p. Thus, both models entail that we cannot claim to know that we know that we have two hands.

4.6 Pragmatism and epistemic consequentialism

Both models of **(Justification Generation)** depend on the fact that entitlement is an (epistemic) warrant. If entitlement was not to count as such a warrant, the above expla-

[98] Given Hero possesses all the relevant concepts. Other reasons for the ceteris paribus clause will become apparent in the following chapters.

nations of **(Justification Generation)** would break down. Most importantly, possessing an entitlement to claim knowledge would not suffice to rationally claim this knowledge in a sense strong enough to allow for the possession of internalistic knowledge in the sense defined in 3.3.1. In other words: if entitlement was not an epistemic warrant, then the proposed lines of second-order justification cannot count as *epistemic justifications* in the first place.

It is one thing to claim that we can account for **(Justification Generation)** as soon as our epistemological toolbox includes a non-evidential notion of warrant. It is another thing to claim that the conditions for entitlement suffice for warrantability in the required sense. This opens the door to an objection. Why should we obtain an *epistemic warrant* for putting trust in a presupposition p just because p is indispensable and unjustifiable in the sense required by the definition of entitlement of cognitive project? Even assuming that we will be able to argue that we *should* put trust in p because we cannot do otherwise (without falling into the abyss of sceptical doubt), the "should" is most plausibly read as carrying pragmatic force, as opposed to *epistemic* force (Pritchard 2005, p. 241).

A classical example for pragmatic warrant is provided by so-called *Reichenbach cases*. Suppose Hero is marooned on a small island, unable to escape. At some point, Hero needs to eat. There is only one type of food to be found on the island: a fruit which Hero has never seen before, and of which he does not know whether it's edible. It seems clear that the rational thing to do in this situation is to assume that the fruit is edible and to eat it. If the fruit is not edible, Hero might die. But Hero will also die if he does not eat the fruit. Making the assumption cannot be any worse than not making the assumption, and potentially saves lives. So Hero should make the assumption. It is a *dominant strategy* with respect to survival.

However, Hero does not have any epistemic warrant to assume that the fruit is edible. He possesses a pragmatic warrant. And one might think that Wright's entitlements are to be conceived in a similar way. Consider the following passage:

> If a cognitive project is indispensable, or anyway sufficiently valuable to us – in particular, if its failure would at least be no worse than the costs of not executing it, and its success would be better ... then we are entitled to — may help ourselves to, take for granted — the original presuppositions without any

specific evidence in their favour. (Wright 2004, p. 192)

The cases look sufficiently parallel. In both cases, assuming something false is no worse than not making the assumption at all, or so the thought goes. So are entitlements pragmatic warrants?

I think that the entitlement theorist should take up this challenge head on and argue that entitled trust in relevant presuppositions is epistemically valuable. That we have pragmatic reasons to believe that sceptical scenarios do not obtain is obvious. And giving up the claim that we can account for epistemic warrants at the basic level is tantamount to giving up the **(Traditional Epistemic Project)**.

How could entitlements be epistemically valuable? Remember the definition of epistemic warrant from 3.2.1. In order to determine whether entitlements can count as epistemic warrants, we need to know what epistemic value is. This gives the entitlement theorist room for manoeuvre.

For Wright's case differs from the Reichenbach case in that the assumption is made to pursue a major objective of cognition: gathering true beliefs. If we pursue a cognitive project on the basis of entitled trust, we have the chance to gather a lot of true beliefs (namely in the case in which the assumptions turn out to be true). If we do not put trust in Wright-presuppositions, we will not gather any beliefs in the first place (at least as long as we are rational). So putting trust in relevant Wright-presuppositions might be a dominant strategy with respect to the promotion of something of epistemic value.

Now suppose we are able to defend the following conception of epistemic value:

(Simple Epistemic Consequentialism) Possessing a doxastic attitude is epistemically valuable if possessing it promotes the epistemic aim of cognition.

Further, suppose that the following principle is true, which might be motivated by a position that Pedersen (2009*a*) calls *veritic monism*, i.e. the position that truth is the only thing of epistemic value:

(True Beliefs) The epistemic aim of cognition is gathering true beliefs.

If "promotes" in **(Simple Epistemic Consequentialism)** is understood in a way that dominant strategies for a certain goal count as promoting that goal, we can infer that entitled assumptions are epistemically valuable.

(Simple Epistemic Consequentialism) can be made plausible by appealing to similar considerations in ethics. Consequentialism — that the moral value of an act depends on it promoting the moral good (e.g. happiness) — is a position that should be taken seriously. Why not also in epistemology?

Unfortunately, there are (at least) two reasons of why the position is problematic, which can be carved out by invoking comparisons with the most straightforward consequentialist position: utilitarianism. Firstly, for utilitarians it is not just the promotion of happiness that counts, but also the avoidance of pain. Similarly, one might think, it is not just the promotion of true beliefs that counts, but also the avoidance of false beliefs. Thus, the **(True Beliefs)** principle is too simple. However, suppose we replace it with:

> **(True Beliefs Without False Beliefs)** The epistemic aim of cognition is gathering true beliefs, and avoiding false beliefs.

Then the assuming presuppositions is not a dominant strategy with respect to the epistemic aim of cognition, and thus not epistemically valuable. For if the presupposition is false, all the beliefs that rest on it will be false as well. Thus, although there still is the possibility of acquiring a lot of true beliefs, there is also the possibility of gathering a lot of false beliefs, and in this context both possibilities have to be regarded as equally likely (Pedersen 2009*a*, p. 450).

Maybe the problem is that it is an *external* factor whether the assumption leads to a lot of true or a lot of false beliefs. Pedersen (2009*a*, p. 447) suggests an *internalistic* notion of epistemic value to save the entitlement theorist. Say that a doxastic attitude is of *teleological value* if its bearer aims at something of epistemic value by taking the attitude, regardless of whether this aim is realized or not. The idea is that entitled assumptions possess such value since they are made with the aim of gathering true beliefs. Consider an analogy in ethics: in some cases, it makes sense to say that an act was good if it has been pursued with a good aim, regardless of whether it has really promoted it. Pedersen gives the example of donating to a charity with the aim to help people, although it turns out that the charity is a hoax, and its members use all the donation money for themselves. This fact does not seem to render the original action bad (at least in some sense).

However, this internalistic version of epistemic consequentialism is still problematic. There is a more general objection to both the externalistic proposal and the teleological proposal, which can again be brought to light by a comparison with utilitarianism. It is a well-known problem for utilitarianism that an act can be a dominant strategy with respect to happiness, or taken to be so, but still be wrong. Take the notorious case in which one person is killed just to use the organs to save five other persons. A similar problem arises in the epistemological case. There are assumptions which lead to many true beliefs, or which are made with this intention, but which are epistemically irresponsible, and thus appear to lack epistemic value. Consider the following case:

> suppose that some quirky goddess has so arranged things that if I believe P—some proposition which I have no other reason to accept and which is in fact false—then she will arrange for the rest of my life to go so fortunately that all the other cognitive acts I ever perform will be absolutely brimming over with all the features that generate epistemic value (whatever they are). However epistemically valuable this consequence of believing P might be, and even if I knew all about the goddess's intentions, the acceptance of P would still be epistemically irrational. (Jenkins 2007, p. 37)

There are many things to say about this case. First and foremost, I think that the epistemic consequentialist (or teleological theorist) could bite the bullet and claim that believing P is not epistemically irrational after all. Of course, one might think, this means giving up the analogy to the ethical case. Clearly, we cannot bite the bullet in the case where one person is killed to save five. However, what if we increase the number of saved people to a couple of billion (or the human race in general)? In this case, intuitions might fade, and one might want to bite the bullet as well. Maybe some cognitive projects are so valuable that the situations become similar.

The utilitarian is most likely to argue either (i) that the action is not really superior in terms of its consequences, or (ii) that we should not focus on what action promotes most happiness in a single situation, but what promotes most happiness in a range of similar situations. In other words: the question is whether the action can be justified by appeal to general rules about which actions promote happiness in general (this position has been called ***rule utilitarianism***).

I do not think that an analogue of (i) is available in the Goddess case. For it is built into the setup of the case that the consequences of accepting P are "absolutely brimming over with all the features that generate epistemic value". However, an analogue of (ii) might be available. In unpublished work, Daniel Elstein has suggested that the entitlement theorist accept an analogy to rule consequentialism in ethics. What counts are not only the direct consequences of a doxastic act, so the thought goes, but whether the *kind* of act has good epistemic consequences in similar cases (i.e. in a lot of situations in which someone promises to provide epistemic goods in abundance). And surely, accepting a proposition just because someone promises to arrange things is highly problematic as a general strategy. Clearly, this needs to be spelled out in much more detail. These issues merit further research.

Jenkins (2007) is pessimistic about consequentialist (and teleological) accounts in general. She insists that the conditions for entitlement cannot suffice for epistemic value, simply because the conditions do not render it more likely that the entitled proposition is true. Her critique rests on the following principle:

> "C: If we are epistemic consequentialists, we ought to think that the epistemic value of a cognitive act depends upon its promotion of those aims which it has in virtue of its being the kind of cognitive act it is." (Jenkins 2007, p.37)

Consider the following auxiliary premise:

> **(Single Aim Of Acceptance)** The single aim of an act of acceptance that it has in virtue of being an act of acceptance is truth.

If both C and **(Single Aim Of Acceptance)** were true, the Goddess case could be explained as a case in which an acceptance of P does not have epistemic value. However, this would also entail that entitled acceptances do not have epistemic value: for nothing about entitlements makes it more likely that the entitled acceptance is true.

Why should we accept **(Single Aim Of Acceptance)**? Jenkins (2007, p. 43) mentions the obvious complaint that acts of acceptance might *also* have the aim of enabling the serious pursuit of cognitive projects, in which case entitled acceptances would count as being epistemically valuable. However, she rejects this move as being *ad hoc*.

I do not think it is obviously ad hoc. The act of "taking something for granted in order to seriously pursue a cognitive project" has the aim of enabling the serious pursuit of a cognitive project. If this is an act of acceptance, and there are other kinds of acceptance with different aims, we would have explained why the aim of acceptance is disjunctive.

As to the act of taking something for granted in order to pursue a cognitive project, Jenkins (2007, p. 43) writes: "I am not aware that I perform cognitive acts like this". Well — what has been said above shows that we *have to* (or have to be able to) perform such acts. However, this does of course not yet *show* that such acts are epistemically valuable. Maybe this is what Jenkins means when she continues writing that "it hard to see any motivation for thinking that I do other than that generated by a desire to rescue the claim that we have entitlement of cognitive project in respect of these propositions" (Jenkins 2007, p. 43).

So Jenkins's objection can be evaluated by asking the following *meta-epistemological* question: do we have to construe our notions of warrant and epistemic value independently of any considerations about what properties these notions need to have in order to enable us to respond to scepticism? If the former is the case, then Jenkins *might* be right, although I think that optimism is warranted because the considerations above point towards modified consequentialist responses that avoid counter-examples like the Goddess case. If the latter is the case, then we can respond that the considerations about scepticism above already show that epistemologists *should* not accept both C and **(Single Aim Of Acceptance)**. For this would be an expression of an incoherent ideal of the notion of (epistemic) warrant that Wright urges us to avoid.

4.7 Intermediate conclusion

To repeat: I have argued that every conservative has to endorse a notion of non-evidential warrant at the basic level, and every internalist has to endorse an internalistic notion of non-evidential warrant at the basic level. I explained how Wright's entitlements of cognitive project can be endorsed to enable to internalists to account for **(Justification Generation)**, which provides a solid starting point for our anti-sceptical internalist foundationalism. However, as we shall see in the next chapter, the proposal also comes with a certain cost, that needs to be avoided by carefully construing the proposed structure of justification.

Before I proceed with my investigation, one note is in order about notions of non-evidential warrant. In this thesis, I will solely be concerned with Wright's notion of entitlement, and with the notion of entitlement cognitive project in particular. I set aside

Burge's (1993), Dretske's (2000) and Peacocke's (2000) notions of entitlement (for a good overview, see Altschul 2011). I do think that Wright's notion is most suitable for my purposes, but a discussion of different notions of entitlement and their suitability for the current project is beyond the scope of this thesis.

5 Transmission-failure

The key results of the last chapter were that we have to endorse non-evidential internalistic warrants at the basic level in order to avoid scepticism, and that there is such a notion: entitlement of cognitive project.

In this chapter, I argue that, although endorsing entitlements for presuppositions enables us to respond to radical scepticism, the response is still concessive in an important sense. Entitlements have a weak epistemic status. First and foremost, they are not sufficient for claimable knowledge, when combined with true belief. So we cannot claim knowledge of merely entitled presuppositions.

However, it is not obvious that initial entitlements for presuppositions cannot be epistemically upgraded. So far, we have seen no reason why **(Justification Generation)** should not also apply to bootstrapping arguments. Since it is desirable to claim knowledge of presuppositions, it is worth examining this possibility.

In this chapter, I argue that such bootstrapping fails. The reason for this is that there are cases where **(Justification Generation)** is defeated because its ceteris paribus clause is violated. The underlying principled reason is a phenomenon that Wright calls *failure of warrant-transmission*. It occurs precisely in cases where we encounter the sort of implicit circularity immanent in bootstrapping cases. Wright argues for this result when it comes to Moorean arguments. I extend it to rule-circular bootstrapping.

So it is indeed impossible to claim knowledge of the presuppositions of basic BFMs. We can only claim entitlement here. Thus, it becomes very important what exactly the presuppositions of relevant cognitive projects are. If the presuppositions include propositions we ordinarily claim to know, our epistemology will yield revisionary sceptical consequences. This sets the stage for the third part of the thesis. What the presuppositions are depends on what our basic BFMs are. Thus, a change in the postulated structure of justification might give rise to different presuppositions, and thus enable us to avoid revisionary claims.

In the last part of this chapter, I show how the structure of justification that Wright proposes in recent work — which is based on entitlement for validity claims as well as on rule-circular arguments — would have the consequence that the validity of basic logical laws has to remain an entitlement and cannot be claimed to be known. In the next chapter, I discuss a similar problem for Wright's newest epistemology of arithmetic. In the last chapter, I argue that we can avoid these problems by changing the proposed structure of justification.

5.1 The weak status of entitlement

I argue that entitlements are weak warrants. In particular, I argue that possessing an entitlement for p is incompatible with possessing a full internalistic warrant for p, given that we cannot upgrade entitlements to full internalistic warrants. I then argue that it is desirable that our epistemology predicts as few entitlements as possible (and as much internalistic knowledge as possible).

5.1.1 Entitlement, evidence, and knowledge

In what sense are entitlements weak warrants? What underlies their weak status is that they are *non-evidential*, i.e. that one does not have to do any evidential work to possess them. That one does not have to possess any evidence is essential, for condition (ii) for entitlement of cognitive project entails that we cannot possess any (initial) evidence for the proposition in question, at least for any notion of evidence that is not crudely externalistic. The *default clause* (i) of the conditions tells us that this does not matter that we cannot possess any such evidence: it is enough that the subject does *not possess* any sufficient reasons to think that P is untrue.

Wright plays with the thought that this might entail that entitlements cannot be construed as warrants to *believe*, for possessing an entitlement for P implies that one does not possess any evidence for P, and "it can seem impossible to understand how it can be rational to believe a proposition for which one has absolutely no evidence, whether empirical or a priori" (Wright 2004*b*, p. 176). Note that this also excludes that the proposition is self-evident or that the proposition seems to be true. All this cannot count as evidence for entitled propositions, so the thought goes, because taking it as evidence would raise further presuppositions in no more secure a prior epistemic standing. It does indeed seem odd to me to say that such a proposition could be rationally believed.

To avoid this worry from the start, Wright construes entitlement as warrants to *trust* (Wright 2004*b*, II). However, the doxastic attitude underlying presuppositions is not really important for my purposes here. What is important is that entitlement would still be a weak epistemic status, even if entitlement was construed as a type of non-evidential warrant to *believe*.

Most importantly, it seems that entitled propositions cannot in principle be rationally claimed to be known. For this requires that the subject can rationally claim that it possesses some evidence for the proposition in question, which is impossible because of condition (ii) for entitlement of cognitive project (and the impossibility to upgrade initial entitlements). Let me explain.

Clearly the conditions for entitlement of cognitive project are remote from all the conditions for knowledge which have been proposed so far. How can a de facto true belief ever amount to knowledge just because doubting it would commit one to doubting the significance of important projects, it is not justifiable without relying on presuppositions in no more secure a prior epistemic standing, and we do not have sufficient reason to believe it to be untrue? Of course, one might construe a notion of schmowledge that includes entitled true beliefs (or acceptances). This might be a fallback position. But it certainly is a revisionary one.

In any case, nothing said so far shows that entitlement is *incompatible* with the possession of knowledge. This is a much stronger claim than that entitled true belief does not entail the possession of knowledge.

In fact, possessing an entitlement for P might be *compatible* with the possession of externalistic knowledge of P. Note that it is not obviously excluded by the conditions for entitlement that the subject possesses such knowledge of relevant Wright-presuppositions. For example, suppose Hero possesses a de facto reliable faculty of intuition about how the world is like. If Hero has the intuition that he is not a BIV, he might count as knowing that.

However, I can ignore this complication here. Remember that the aim is to engage in the **(Traditional Epistemic Project)**. And the possession of entitlement indeed seems to exclude the possession of *accessible justification*, in any sense of justification that requires the possession of evidence. Let us assume that knowledge requires the possession of evidence, either internalistically or externalistically construed. Clearly, possessing an entitlement for P is not yet to be in a position to rationally claim that one possesses such evidence for P. A fortiori, one might think that this is indeed *excluded* by the regress clause

(ii) for entitlement of cognitive project: for how could we ever be in a position to *rationally claim* that we possess evidence for P if any justification for P must rest on presuppositions in no more secure a prior epistemic standing, i.e. if any attempt to justify P leads to epistemic regress? This is just the point of the two sceptical challenges of the last chapter. Ergo: if we possess an entitlement for P, then it seems *impossible* to internalistically know that P (or to rationally claim that one knows that P).

However, this argument is too quick. For the situation we encountered in the sceptical arguments above has changed. We now possess a notion of non-evidential warrant in our epistemological toolbox. Once we possess an entitlement for P, P *is* warranted. Thus, at least one type of justificatory regress is blocked: we do not have to earn a warrant for P anymore, which would commit us to further presuppositions in no more secure a prior epistemic standing.

A fortiori, according to **(Justification Generation)**, once the presuppositions of a basic BFM are entitled, we can use this BFM to acquire internalistic knowledge. The question naturally arises of whether we upgrade initial entitlements for presuppositions to full internalistic knowledge *afterwards*, using the very BFMs whose presuppositions are entitled. That is: whether there is evidence for a proposition, part of which presupposes or relies on entitlement to the very same proposition.[99]

For example, one might try to upgrade a previously entitled validity claim by a rule-circular argument. Nothing said so far excludes this possibility. However, if such upgrading was possible, then it would not be the case that entitlement is incompatible with the possession of internalistic knowledge. The incompatibility argument thus rests on the assumption that one cannot epistemically *upgrade* initial entitlements to internalistic knowledge by bootstrapping arguments, such as rule-circular reasoning.[100]

I think that bootstrapping is impossible because of a phenomenon Wright calls *failure of warrant transmission*. According to Wright's *transmission-failure diagnosis*, the ceteris paribus clause of the **(Justification Generation)** principle is violated in bootstrapping cases. In this chapter, I examine Wright's diagnosis and evaluate some of its consequences.

[99]This formulation has been suggested to me by Robert Williams.

[100]Pedersen (Pedersen 2007, p. 19) argues that clause (ii) of entitlement of cognitive project can be taken to imply — with "a bit of unpacking" — that it is impossible to *improve* the epistemic standing of an entitled proposition. However, I think that Pedersen's argument also ignores the possibility of using a combination of bootstrapping and **(Justification Generation)**. He does not even consider this possibility.

First, I will provide some additional considerations to the effect that claimable knowledge and evidence is valuable, and that it is thus desirable to avoid the postulation of entitlements wherever possible. Secondly, I present Wright's diagnosis and argue that it is correct. After that, I use these results to show that this has revisionary sceptical consequences for Wright's epistemology, because he postulates a structure of justification in which the soundness of basic logical laws and mathematical basic rules are initial entitlements.

5.1.2 The epistemic value of claimable knowledge (and evidence)

Below, I will complete my argument that entitlement is incompatible with claiming knowledge and evidence. Full internalistic warrant is strictly stronger than entitlement. But why should this bother us?

The reason cannot be that only full internalistic warrant has intrinsic epistemic value. For this would render our response to scepticism even more concessive: I argued in the last chapter that this would preclude us from rationally claiming knowledge (and the possession of evidence) in the first place, and suggested epistemic consequentialism to the rescue. Entitlement has intrinsic epistemic value.

However, although both entitlement and full internalistic warrant have intrinsic epistemic value, I think the latter has *more* epistemic value. This provides the basis for a meta-epistemological argument to the effect that it is desirable to be able to claim full warrant wherever possible.

I go through one epistemically valuable feature of full internalistic warrant, and argue that entitlement does not promote the same value. After that, I sketch further features for which I am confident that the same argument can be made. I leave these for further research.

Intellectual stability Being in a position to claim a full warrant for p entails being in a position to claim that one possesses evidence for p. And being in a position to claim evidence for p is more valuable than being in a position to claim an entitlement for p because, if we are in a position to claim that we possess evidence for p, it is more likely that we can uphold our positive attitude towards p in the light of doubt.

Suppose Hero believes that p and that Anti-Hero provides some reason to doubt p.

If Hero can only claim an entitlement for p, then whether or not Hero has to give up his positive attitude towards p depends on whether Anti-Hero's reasons to doubt p are ***sufficient*** reasons to believe that p is untrue, because this would undermine Hero's (claim to) entitlement.

If Hero can claim to possess evidence E for p, Hero can cite this evidence in response. Now, it becomes an interesting issue whether Hero's evidence for E is stronger than whatever the grounds for Anti-Hero's doubts are. In particular, Hero might rationally uphold the belief that p even if Anti-Hero's reason to doubt p was a sufficient to believe that p is untrue, when taken on its own. For Hero might be able to see that his own evidence for p is stronger than Anti-Hero's reasons to the effect that p is untrue.

In this sense, being in a position to claim the possession of evidence promotes intellectual stability. Of course, more needs to be said here. First and foremost, one would need to explicate the notion of intellectual stability, and argue that intellectual stability is epistemically valuable. I cannot do that here, but, all other things being equal, the claim that intellectual stability is epistemically valuable is intuitively correct.

Other considerations There are at least two more considerations that merit further research. The first is mentioned in (Wright 2008), but not explicated in detail. Wright points out that internalistic (evidential) warrant is valuable when it comes to managing a system of beliefs: only if one is aware of what is evidence for what — how one's beliefs hang together — will one be able to keep holding a sufficiently consistent set of beliefs, which might be constitutive of possessing a system of beliefs in the first place. Note that entitlements might also be valuable in such a project, but, just as above, evidential relations might have a special role to play.

The second kind of consideration is implicit in both (Wright 2008) and (Alston 2005). The point is that one might be able to argue that reflective justification is *indirectly* truth-conducive (Alston 2005, p.43), just as following norms of coherence, rationality etc. are indirectly truth-conducive. That is: if I follow these norms — and I can only do so if I possess internalistic justification — I will, ceteris paribus, gather more true beliefs than someone who does not follow these norms.

I agree with Wright (2008, p. 507) that these considerations can also be used as considerations in favour of **(Relevance Internalism)**, because the specific kinds of epistemic value mentioned here are best promoted by reflectively accessible justification. Completing the argument for the value of internalistic evidential warrant presumably also completes the argument for **(Relevance Internalism)** sketched in 3.3.3. However, this claim also requires further research.

5.1.3 We should prefer internalistic knowledge over entitlement

It is a reasonable aim for any epistemology to allow for as much epistemic value as possible. In particular, one should try to find a canonical structure of justification that can be used to acquire as much internalistic knowledge as possible. The underlying thought is that every philosophical theory is subject to cost-benefit considerations. Of course, we also have to take into account the costs. This yields the following principle:

> **(Greater Meta-Epistemological Evil Principle)** Other things being equal, we should prefer an epistemology that accounts for as much epistemic value as possible, without sacrificing something of comparable philosophical worth.

Since the possession of internalistic knowledge is epistemically valuable, and more valuable than the possession of other internalistic warrant (entitlement), we should prefer an epistemology that account for as much internalistic knowledge as possible, without sacrificing something of comparable philosophical worth. This principle will be applied in chapter 7. I will argue that this principle delivers a reason to postulate a BFM that generates internalistic knowledge of basic logico-mathematical principles, since the ensuing structure of justification allows for more internalistic knowledge than a structure of justification that uses entitlements for logico-mathematical basic principles.

Secondly, our epistemology should not be revisionary. It should not yield the consequence that we cannot claim knowledge of propositions in paradigmatic cases of knowledge. For example: ordinary claims about the external-world; the validity of basic logical laws; mathematical axioms. Of course, one needs to be careful here: it might be unavoidable to give up certain claims to knowledge. However, such a radical divergence casts at least some doubt upon our epistemological framework.

5.2 Transmission-failure

The above considerations show that it is desirable to account for more than mere entitlement wherever possible. According to **(Justification Generation)**, one can acquire internalistic knowledge on the basis of BFMs with entitled presuppositions. Thus, the question arises of whether we can acquire more than mere entitlement for the presuppositions themselves by means of bootstrapping arguments. I now examine bootstrapping arguments in two relevant cases and argue that they fail because they are subject to a phenomenon Wright calls *failure of warrant-transmission*.

5.2.1 Moorean upgrading

I begin with the perceptual case. The most prominent argument for bootstrapping in the perceptual case is an argument by Moore (1959).[101] Here is an argument of the same type. Assume that a subject S has a veridical experience as of his or her two hands. Let "AsOfHands" stand for "I have, right now, and experience as of my two hands", let "Hands" stand for "I have two hands", and let "$\neg BIV$" stand for "I am not a brain-in-a-vat"). According to Moore, S can reason as follows:

1	*AsOfHands*	Introspection
2	*Hands*	From 1
3	$\neg BIV$	2, conceptual entailment

The argument's inferential steps seem acceptable. If this is so, then S should be able to acquire knowledge of the non-obtaining of the brain-in-a-vat scenario just on the basis of a perceptual seeming. This would be an odd result. It seems to be too easy to dismiss Cartesian scepticism in this manner.

Fortunately, **(Conservativism about Perception)** provides the resources to explain at least one aspect of our dissatisfaction with the argument. According to conservativism, it a necessary condition for the step from 1 to 2 to transmit evidential warrant that S already possesses a warrant for 3. Thus, no *first* warrant can be obtained for 3 by means of the argument.

[101] Pryor (2004) also thinks that Moorean arguments transmit warrant, although he also thinks they are dialectically ineffective.

Note: no *first* warrant. But what if S already possesses an entitlement for 3? In this case, the step from 1 to 2 transmits full internalistic warrant, in accordance with **(Justification Generation)**. And then, if 2 conceptually entails 3, why should S not acquire a full internalistic warrant (and internalistic knowledge) for 3 in turn?

This cuts both ways. On the one hand, this would be a good result, since it would show that we can obtain internalistic knowledge of Cartesian presuppositions after all. On the other hand, however, this result seems counterintuitive. In fact, the claim that such upgrading is possible appears just as dissatisfying as the claim that we can acquire a first warrant for 3 by means of such reasoning.

5.2.2 Wright's diagnosis

Fortunately (or unfortunately), there is a general phenomenon Wright calls *failure of warrant-transmission* which precludes such upgrading. Let me explain.

What could we say against the possibility of upgrading by means of the argument above? First, it is important to note that **(Justification Generation)** by no means *entails* that we can upgrade initial entitlements by an argument as the one above. For it includes a ceteris paribus clause which might be defeated in special circumstances. Thus, everyone having the intuition that warrant cannot be transmitted in Moorean cases should look for a principled reason why the ceteris paribus clause is violated in the relevant cases.

One might think that upgrading is impossible, for the same reason that no first warrant can be acquired. One already needs to possess a warrant for 3 in order to obtain a warrant for the final premise 2, from which a warrant for the conclusion 3 should be obtained in turn, and this is a bad form of epistemic circularity, or so the thought goes.

However, this is not obvious. There is no epistemic circularity in the sense that one needs to possess a warrant of type T for 3 in order to obtain a warrant of type T for 3. The thought is only that one can rely on a type of warrant T' weaker than T in order to upgrade one's warrant for 3 to T. In particular: that we can rely on an initial entitlement for P to acquire a full internalistic warrant for P by means of the Moorean argument. There is certainly *some* circularity in here — in a pre-theoretic sense of circularity — but it is not clear that it is *vicious*.[102][103] We clearly need an argument against the possibility of upgrading.

Wright (2007*b*, p.36) argues as follows:

- The acquired warrant for the conclusion of an inference cannot be any stronger than the warrant one possesses for the premise.
- Thus, one's acquired warrant for 3 cannot be any stronger than the warrant for 2.
- By **(Conservativism about Perception)** it is an enabling condition to acquire a warrant for 2 that one possesses an antecedent warrant for 3.
- If it is an enabling condition to acquire a warrant for 2 that one possesses an antecedent warrant for 3, then one's warrant for 2 cannot be any stronger than one's antecedent warrant for 3.
- Therefore, one's acquired warrant for 3 cannot be any stronger than one's antecedent warrant for 3.

If this is right, then we can neither acquire a first warrant for 3, nor upgrade an antecedent warrant for 3 by the Moorean argument. This phenomenon Wright calls *failure of warrant transmission* (Wright 2007*b*, p. 36):

> **(Transmission Failure)** An argument or inferential step *fails to transmit warrant* (of type T) if and only if it is impossible to acquire a first warrant (of type T), or to upgrade an antecedent warrant of type T'<T to a warrant of type T, by virtue of carrying out the argument or the inferential step.

It is important to restrict failure of warrant transmission to certain types of warrant. For upgrading might be possible for some warrants, but not for others. In particular, I think that Wright's argument for transmission-failure is flawed, but there is a good argument for transmission-failure for the specific case of upgrading entitlements to full internalistic warrant.

[102] I am indebted to Tobias Wilsch here for a sustained discussion of this issue, and to Philip Ebert, who independently made me aware of it.

[103] Nicholas Silins (2005) argues that the general argument for the impossibility to upgrade one's warrant by Moorean reasoning rests on a confusion of several notions of possession of warrant. Note, however, that even if his argument against general transmission-failure goes through, there might still be room for transmission-failure of full internalistic warrant. Indeed, I think that Silins' arguments fail for this notion of warrant. But I cannot discuss Silins' argument any further here.

As to the first point, note that Wright's penultimate premise is incompatible with **(Justification Generation)**. If it really is the case that the strength of the warrant for 2 is bound by the strength of 3, then, if 3 is an initial entitlement, the epistemic status of 2 cannot be any stronger than entitlement. This is the **(Strong Limit Principle)** from 4.5.1, which contradicts **(Justification Generation)**. I argued that we should reject this principle, and Wright argues for **(Justification Generation)** elsewhere (Wright 2004*b*).

However, Wright (2003, p. 59) sketches another argument for transmission-failure for the specific case of upgrading an antecedent entitlement for 3 to a full internalistic warrant, which I think is sound. That we can acquire a full internalistic warrant for 3 by going through the argument above requires that the subject in question can rationally regard 2 as evidence for 3. However, given the conservative diagnosis, the subject can rationally regard 2 as warranted and thus as evidence at all only because it is antecedently entitled to trust in 3, and this undermines the subject's ability to rationally regard 2 as evidence for 3. This is because of the general fact that one cannot rationally claim that p is evidence for q if p's status as evidence is conditional on q being antecedently warranted. It is the violation of principles of rationality at the second-level, and not the **(Strong Limit Principle)**, that leads to transmission-failure.

Wright (2007*b*) extracts two templates of when he thinks that warrant fails to transmit in the external-world project. Here is the template that applies to the case at hand:

(Information Dependence Template)

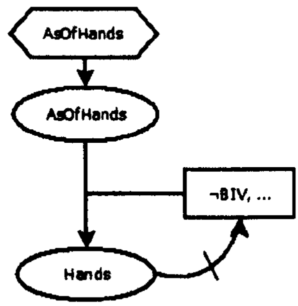
For all propositions e, P, and I, if

- (a) the transmission of internalistic warrant from e to P requires that the subject possesses an antecedent warrant for I, and
- (b) P and other warranted premises logically entail I,

then, if P is obtained from e, a subsequent inference from P to I will fail to transmit internalistic warrant.

The Moorean argument meets these conditions. Thus, according to the **(Information Dependence Template)**, it is not a means to acquire a first warrant for 3, and also no means to strengthen one's warrant for 3. **(Justification Generation)** is not applying to these cases, because its ceteris paribus clause is violated.

This is a correct diagnosis. The template applies only if the argument for transmission-failure of internalistic warrant applies.[104] In the Moorean case, it can be displayed as follows:



The **(Information Dependence Template)** severely constrains our response to external-world scepticism. Presuppositions such as 3 are entitlements, and condition (ii) for entitlement of cognitive project entails that there is no course of justification for them that does not lead to an epistemic regress, or circularity. Moorean upgrading was our only chance to acquire full internalistic knowledge of 3.[105] 3 has to remain an entitlement forever.

Thus, although our epistemological framework can account for internalistic knowledge of ordinary external-world propositions, it predicts that we are unable to acquire internalistic knowledge of the presuppositions of perceptual BFMs. It seems as if the presuppositions must remain entitlements. We can only claim that we can put rational trust into them, but not that we possess evidence for, or knowledge of them. Wright is of course aware of this consequence for his own framework, and call his response to external-world scepticism concessive for this reason (Wright 2004*b*, p. 206). We saw above in what sense this consequence might be relevant and undesired.

In any case, these points are not restricted to the external-world project. There is no reason why these considerations should not generalize to all BFMs with entitled presuppositions. Neither the argument for transmission-failure, nor Wright's **(Information Dependence Template)** depended on the specifics of visual perception or Cartesian

[104]However, the template might not apply to all cases. It is designed for cases in which the conclusion follows by a deductive argument.

[105]Philip Ebert and Tobias Wilsch independently made me aware of the fact that coherentists might deny this. I will ignore coherentist approaches here and assume that there is no other route to full internalistic warrant. In fact, I think that coherentism is incompatible with the **(Traditional Epistemic Project)**, but I cannot argue for that claim here.

scenarios. Both are motivated by quite general considerations about the structure of justification. The following principle holds:

> **(No Upgrade)** The epistemic status of entitled presuppositions of a BFM M cannot be upgraded from mere entitlement to full internalistic knowledge by any line of reasoning using M.

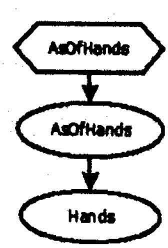
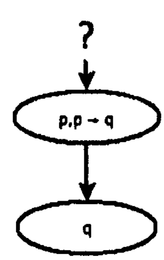
Moreover, such upgrading is our only chance to improve the epistemic standing of entitled presuppositions, for it follows from condition (ii) for entitlement of cognitive project that there is no M-independent route to justifying them that does not lead to an infinite regress. Thus, the following principle holds as well:

> **(No Knowledge Claims)** Entitled presuppositions of our BFMs cannot become items of internalistic knowledge at all.

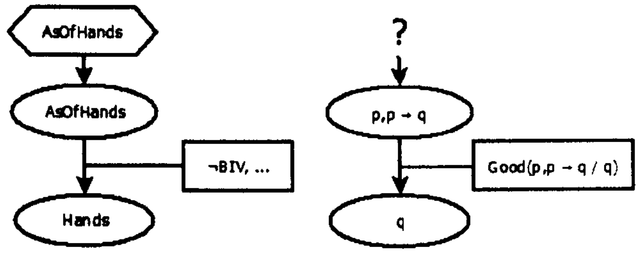
5.2.3 Consequences

In the remainder of this chapter, I apply this result to the logical case. In the next chapter, I discuss the mathematical case. This will eventually lead to an application of the **(Greater Meta-Epistemological Evil Principle)**: I argue that changing the structure of justification avoids some of the sceptical consequences in the logico-mathematical case.

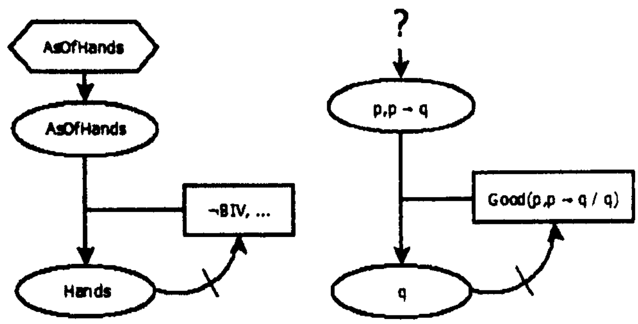
How do our results bear on the logical case? Note we also endorsed **(Conservativism about Deduction)**. From what has been said above, the following table for the transmission of full internalistic warrant can be extracted, comparing the perceptual and the logical cases:

	Perception	Deduction
Liberalism		

Conservativism



Transmission-failure



Transmission-failure will affect every argument for the goodness of a deductive rule of inference which uses that very rule. In short: it affects rule-circular arguments.

This has consequences for Wright's own epistemology. In his latest work on the epistemology of logic, Wright drops his earlier suggestion that we acquire basic logical knowledge on the basis of implicit definition.[106] Wright (2004*a*, IV) argues for the claim that the validity of basic deductive BFMs is an entitlement of cognitive project. However, claims to knowledge of validity are paradigmatic claims to knowledge. Every epistemology that does not account for the possibility to claim such knowledge is revisionary. To avoid the consequence that we can only claim entitlements for validity, Wright suggests using rule-circular arguments to upgrade these initial entitlements (Wright 2004*a*, VIII). However, since rule-circular arguments are attempts to upgrade the epistemic status of the validity claim for a deductive rule M by reasoning using M, they cannot be used for epistemic upgrading in Wright's framework because of **(No Upgrade)**. In fact, because of **(No Knowledge Claims)**, Wright's epistemology has the consequence that validity claims have to remain

[106]He does *not* give up the claim that there is a strong connection between the good standing of a definition of introduction and elimination rules and the good standing of logical concepts. However, he does give up the claim that implicit definition can play a *justifying* role. I discuss this a little bit more in 6.1.3.

entitlements forever and cannot become items of claimable knowledge at all. This is relevant and undesired for the reasons sketched at the beginning of this chapter. First and foremost, this is a revisionary claim. Claims to knowledge of validity are paradigmatic claims to knowledge.

5.3 Rule-circular arguments

At first glance, rule-circular arguments might look like the best option to *justify* validity claims.[107] For how can we justify apriori that basic logical laws are valid? We must either justify their validity inferentially — i.e. through some line of apriori reasoning — or non-inferentially. How could their validity be justified non-inferentially? On the face of it, the basis for a non-inferentially justified belief must be experiential states (Boghossian 2001, p.6). This plausibly amounts to the claim that we can acquire justification through some faculty of rational intuition (Bonjour 1998). However, no plausible account of the exact epistemological workings of rational intuition has been produced (Boghossian 2001, p.6; Wright 2004*a*, p. 156). This leaves us with the option that validity can be justified inferentially. And prima facie, the following looks plausible:

> **(Logicality of Inference)** Every inferential justification of a validity claim makes use of logical reasoning.

Such a justification would be rule-circular in relevant cases. So rule-circular arguments are unavoidable, or so the thought goes.

5.3.1 Transmission-failure

An argument is rule-circular just in case its last line expresses the goodness of a rule of inference used in the argument. An initially plausible candidate for goodness in the logical case is *validity*. In what follows, I assume that validity is necessary truth-preservation.

I can focus on the case of MP here without loss of generality. The following "boxed" universal generalization is a plausible candidate for a validity claim — where the box ("$\Box$") expresses necessity, x and y are variables ranging over propositions (not linguistic entities),

[107]Note: not to warrant validity. Validity claims might be warranted non-evidentially by means of entitlement.

and "$\ulcorner x \rightarrow y \urcorner$" stands for a function mapping the propositions x and y to the proposition $x \rightarrow y$:

(MPV) $\forall xy\,(\Box\,((T\,(x) \wedge T\,(\ulcorner x \rightarrow y \urcorner)) \rightarrow T\,(y)))$

However, in what follows, I will distinguish between *particular* validity statements and *general* validity statements. Whereas a general validity statement entails the goodness of all instances, a particular validity statement expresses the goodness of a particular inferential step. The validity of a particular MP step is plausibly expressed by the following (where p and q stand for the relevant particular propositions):

(MPV-Instance) $\Box\,((T\,(p) \wedge T\,(\ulcorner p \rightarrow q \urcorner)) \rightarrow T\,(q))$

The distinction between particular and general validity claims is in order because conservativism about deduction might be formulated in terms of particular or general validity: the particular validity or the general validity of a particular inference might be claimed to be among its presuppositions. Presumably, Wright holds that it is particular validity claims which feature as presuppositions of particular inferential steps. Firstly, his discussion of Carroll's regress[108] in (Wright 2001, pp. 73f) at least suggests that it is a warrant for the validity of a *particular* inference that enables this inference to transmit warrant. Secondly, Wright's formulation of the presuppositions of deductive inferences in (Wright 2007*b*, p. 30) also seems to focus on the validity of particular inferences, rather than on the validity of the underlying inference type.

Thirdly, someone might understand goodness as (mere) truth-preservation. Again, there is a particular and a general version. We can focus on particular truth-preservation without loss of generality:

(MPT-Instance) $(T\,(p) \wedge T\,(\ulcorner p \rightarrow q \urcorner)) \rightarrow T\,(q)$

I show that rule-circular arguments for the validity of a rule R display transmission-failure in all cases, i.e. no matter of whether we take **(MPV)**, **(MPV-Instance)**, or **(MPT-Instance)** to express goodness.

[108] This refers to the well-known problem of justifying logical inferential steps presented in (Carroll 1895).

Without loss of generality, we can focus on the following argument for **(MPV)**, which does not rest on any premises (only discharged assumptions), and which contains a particular[109] MP step from 4 to 5:

1	$T(p)$	Ass.
2	$T(\ulcorner p \rightarrow q \urcorner)$	Ass.
3	p	1, T-Elim
4	$p \rightarrow q$	2, T-Elim
5	q	3,4, MP
6	$T(q)$	5, T-Intro
7 (=**(MPT)**)	$T(p) \rightarrow (T(\ulcorner p \rightarrow q \urcorner) \rightarrow T(q))$	6, 2x CP
8 (=**(MPV-Instance)**)	$\Box(T(p) \rightarrow (T(\ulcorner p \rightarrow q \urcorner) \rightarrow T(q)))$	7, Nec.[110]
9 (=**(MPV)**)	$\forall x \forall y (\Box(T(x) \rightarrow (T(\ulcorner x \rightarrow y \urcorner) \rightarrow T(y))))$	8, UG[111]

Even ignoring complications about the availability of the required inferential steps — such as the T-rules — we must conclude that the above argument is no means to upgrade the epistemic status of **(MPV)** from mere entitlement to full internalistic warrant, because it exhibits the **(Information Dependence Template)**. Here is why:

If the general validity of MP is a presupposition of MP steps, then the above argument will instantiate the information-dependence template because its last line (9) expresses the general validity of MP. An initially entitled **(MPV)** will remain a mere entitlement by **(No Knowledge Claims)**.

On the other hand, if it is just particular validity statements which feature as presuppositions — as Wright presumably thinks — then the sub-argument from line 1 to line 8 will instantiate the **(Information Dependence Template)** because line 8 expresses the particular validity of the MP step from 4 to 5. Thus warrant cannot be transmitted to line 8, and the previously entitled particular validity statement will remain a mere entitlement by **(No Knowledge Claims)**. A fortiori, everything we derive from merely entitled premises cannot have an epistemic status above mere entitlement, for the epistemic status of the conclusion of a warranted inference cannot be any stronger than the epistemic

[109]The step is meant to be *non-schematic*. p and q stand for arbitrary, but particular propositions.
[110]"Nec." stands for "Necessitation".
[111]"UG" stands for "Universal Generalization"

status of its epistemically weakest premise — this is just the **(Limit Principle)** from 4.5.1.[112] Therefore, subsequent lines of the argument — including **(MPV)** — will remain entitlements as well.[113]

Hence, if either particular or general validity is among the presuppositions of MP inferences, **(MPV)** has to remain an entitlement. Finally, taking only the unboxed "$T(p) \to (T(\ulcorner p \to q \urcorner) \to T(q))$" (=**(MPT)**) to be a presupposition of the MP step does not help, for there would then be transmission-failure in the sub-argument ending with line 7. Other combinations of presuppositions and arguments fail as well, for the same reasons. For example, the schematic argument Wright uses in (Wright 2004*a*).

The weaker the presuppositions are, the worse the situation becomes from an epistemological point of view. For example: we have seen above that, if *particular* validity statements are the proper presuppositions of deductive inferences, both particular validity statements and general validity statements cannot be claimed to be known. In addition to not being able to claim that we know that all instances of MP are necessarily truth-preserving, we could not even claim to know that it is necessarily so that if it is true that it rains, and it is true that if it rains, then there are clouds, then it is true that there are clouds.

If, on the other hand, general validity statements are the proper presuppositions, then it might at least be possible to derive particular validity statements without creating instances of the **(Information Dependence Template)**. Thus, we might be able to acquire full internalistic warrants for particular validity claims, although the general claims have to remain mere entitlements.

5.3.2 The failure of Justification Generation

Interestingly, Wright (2004*a*) nevertheless seriously considers using rule-circular arguments to upgrade the epistemic status of validity claims from entitlement to full internalistic knowledge:

[112]Note that Wright explicitly endorses this claim (Wright 2004*b*, p. 191).

[113]There is another step in the argument above which is problematic with respect to warrant transmission. If we use a *particular instance* of MP in our argument for the *general* validity of MP, the we put epistemic load into the UG step. I agree with Dogramaci (2010) in that the UG step then presupposes an antecedent warrant for the general validity of MP. Thus, there is transmission-failure at the UG step at the latest. However, note that my worry is independent of, and more general than Dogramaci's. Most importantly, it also affects arguments for particular validity claims and arguments for truth-preservation.

> if we are entitled to claim knowledge of a statement which we have recognised to follow from known premises by inference in accordance with entitled rules, then we are surely entitled to claim knowledge of a statement which we have recognised to follow from an empty set of premises by inference in accordance with entitled rules. But – assuming an entitlement to MPP and conditional proof – that is just what a rule-circular derivation of MPP provides for. (Wright 2004*a*, p. 173)

This is the same line of thought that we used to establish the **(Justification Generation)** principle. Wright suggests that it might also apply to rule-circular arguments. However, we have seen that **(Justification Generation)** must be restricted by a ceteris paribus clause, if we accept the transmission-failure diagnosis for full internalistic warrant. Wright argues for transmission-failure at many occasions. It is odd that Wright simply ignores this phenomenon at this point.

However, the question arises of why **(Justification Generation)** is subject to a ceteris paribus clause. After all, our argument for this claim appeared to be entirely general. We need to be able to explain why the argument does not apply in cases of transmission-failure.

In 4.5, I considered two arguments for **(Justification Generation)** — one for an externalistic notion of knowledge, and one for an internalistic notion of knowledge. Both arguments must be unsound in cases of transmission-failure. What makes them unsound?

I said above that transmission-failure occurs because we cannot rationally claim knowledge of p on the basis of an argument that includes the premise that we possess an entitlement for p. However, the argument for **(Justification Generation)** rests on the thought that Hero can acquire an (inferential) entitlement for the claim that he knows the target belief of some BFM (or a chain of BFMs), by realizing that he has appropriate epistemic access to the sources of the BFM, *and* its presuppositions. In rule-circular cases (or Moorean cases), the presuppositions are the same as the target beliefs. And this generates a problem for Hero's second-order argument. It is useful to briefly consider both versions of the second-order argument in the Moorean case to see how exactly they fail. For an externalistic notion of knowledge, the argument is very simple — Hero tries to acquire an inferential entitlement for the relevant knowledge claim by applying a closure principle for externalistic knowledge to the entitled knowledge claim that he has two hands:

1	$K(Hands)$	The second-order argument from 4.5.2
2	$K(Hands) \rightarrow K(\neg BIV)$	Closure of externalistic knowledge
3	$K(\neg BIV)$	Accessibility of epistemic facts

Because of transmission-failure, Hero cannot rationally claim 3 on the basis of apriori reasoning and self-reflection (although he can rationally claim 1 on this basis).

It is problematic to give up closure, because the Instability Argument for **(Relevance Internalism)** assumed that externalists should not give up closure. What this brings to light is that rational claimability cannot be closed under (recognized) logical consequence. A fortiori, since the argument for **(Justification Generation)** rests on the thought that we can claim knowledge of p because we can acquire an inferential entitlement for the claim that we know p, inferential entitlement (and hence, entitlement) is not closed under logical consequence. Presumably, Wright would not want to accept this consequence. I cannot see, however, how we can avoid this consequence if we want to uphold our model for **(Justification Generation)** and allow for transmission-failure without giving up closure. Since both the transmission-failure diagnosis and **(Justification Generation)** must be right, I am willing to bite the bullet here.

A different diagnosis will be available if we hold that the notion of knowledge is an internalistic notion. In this case, we can give up closure without undermining our argument for **(Relevance Internalism)**. Closure will fail in all cases of transmission-failure. Thus, Hero cannot rationally claim (is not entitled to claim) the conditional "$K(Hands) \rightarrow K(\neg BIV)$".

5.4 Two responses

Let us return to Wright's epistemology. Wright either has to give up the transmission-failure diagnosis, or he has to concede that his strategy to vindicate rule-circular arguments fails. I think that the transmission-failure diagnosis is correct. The principles **(No Upgrade)** and **(No Knowledge Claims)** apply. By **(No Upgrade)**, rule-circular arguments cannot be used to upgrade initial mere entitlements to full internalistic knowledge. A fortiori, by **(No Knowledge Claims)**, we obtain the revisionary sceptical consequence that in Wright's epistemology the validity of basic logical rules cannot be claimed to be known at all. Since he thinks that it is particular validity claims that feature as presuppositions, even particular validity statements cannot be claimed to be known.

The only way to avoid these sceptical consequences is to find or construct a disanalogy between the Moorean case and the rule-circular case. That is: one needs to argue that the **(Information Dependence Template)** does not apply to the logical case, but just applies to the external world case.

However, it is hard to see how such an argument would look like. The argument for transmission-failure in the perceptual case is based on entirely general considerations about the structure of justification, and maybe also on entirely general considerations about rational claimability. The underlying considerations, if they are correct in the perceptual case, will also be correct in the logical case and indeed in any conservatively treated area of cognitive inquiry, or so it seems. Nevertheless, there are some peculiarities about the rule-circular case that one might want to use to argue that there is a disanalogy. I close with rebutting two such arguments.

5.4.1 Discharged assumptions

The reader will have noticed that there is a structural disanalogy between the Moorean argument and the rule-circular argument. The rule-circular argument above rests on discharged assumptions, whereas the premises of Moore's proof are not assumptions. Among other things, one might wonder whether one can speak of warrant being transmitted in the former case at all: after all, the assumptions are merely assumptions, and as such not really warranted. We can assume whatever we want to assume.

Firstly, the disanalogy can be removed by changing the rule-circular argument. One might attempt to acquire internalistic knowledge of 5 by means of using an axiomatic calculus which includes MP as its only rule of inference. We can assume that the proof will be based on axioms different than 5, which are internalistically known.[114] There is no structural disanalogy anymore. The situation is just as in the Moorean case: we have a proof of p on the basis of internalistically known premises, where at least one inferential step presupposes p. The **(Information Dependence Template)** applies. Now, it is

[114]If the axioms are not internalistically known, there is no way we can acquire internalistic knowledge of the conclusion in the first place. If 5 is an axiom, we beg the question because 5 would not have to be justified.

very implausible that we can justify the validity of MP by using rules and discharged assumptions, but that we cannot do so if we use an axiomatic calculus. We should conclude that warrant cannot be transmitted in both cases.

Secondly, there is a way to use the transmission-metaphor even in the case of discharged assumptions. We might regard discharged assumptions as temporarily warranted, and ask whether temporary warrant can be transmitted. I just cannot see any reason why temporary warrant can be transmitted in the relevant cases, given that we accept the transmission-failure diagnosis for non-temporary warrant.

5.4.2 Entitlement for rules

The second strategy concerns a possible distinction between entitlement for propositions and entitlement for rules. Suppose we have a notion of an *entitlement to apply* a rule. Perhaps, we can also defend the following principle:

> **(Justification Generation for Rules)** Deductive inferences we are entitled to apply transmit internalistic knowledge from their premises to their conclusion.

The proposal is similar to Wright's in that we need to possess an antecedent entitlement in order for deductive rules to extent internalistic knowledge. However — and this is the crux — we would not need to possess an antecedent entitlement for a *proposition*. Thus, so the thought goes, Wright's arguments for transmission-failure do not apply to the logical case: there is simply no antecedently entitled *proposition* whose epistemic status we desire to upgrade.

However, firstly, one would have to spell out what an entitlement to apply a rule consists in, in such a way that it does not consist in an entitlement for a proposition. On the face of it, it seems natural to say that an entitlement for a rule just consists in an entitlement for the goodness of the rule. Whatever proposition expresses goodness, there would be transmission-failure again: we would not be able to upgrade our entitlement by a rule-circular argument.

Secondly, one might worry that the whole dialectic can simply be replicated for entitled rules.[115] That we are merely entitled to apply a rule would be a sceptical result, just as the result that we possess only entitlements for validity. Of course, one might attempt to

[115] I owe this point to Robert Williams.

upgrade our entitlement such that we become *fully justified* in applying a rule. However, if this upgrading requires making use of the rules in question, it is plausible that we face transmission-failure worries again. If we cannot epistemically upgrade entitled propositions using a line of reasoning including this proposition as a presupposition, then how could it be possible to epistemically upgrade entitled rules using a line of reasoning including this rule?

5.5 Intermediate conclusion: unavoidable sceptical consequences?

In the last chapter, the question has been raised of how presuppositions of basic BFMs can be warranted. A sceptical argument showed us that, if they are warranted at all, they are warranted *non-evidentially*. We endorsed Wright's notion of entitlement to avoid sceptical consequences. We saw that many propositions can be internalistically known because BFMs with entitled presuppositions *ceteris paribus* confer and transmit full internalistic warrant. This chapter revealed that the ceteris paribus clause is crucial. The epistemic status of antecedently entitled presuppositions cannot be epistemically upgraded above entitlement because of the phenomenon of failure of warrant transmission.

This consequence is significant because entitlement is a weak epistemic status and it is desirable to obtain more than entitlement. I provided some considerations to this effect. However, there might be a difference among cases. Some presuppositions might be too remote to generate the problem that our epistemology is revisionary. Other presuppositions might be such that the result that we cannot claim knowledge of them has to count as a revisionary sceptical consequence.

I do not know to what extent the result that we cannot claim knowledge of the non-obtaining of the BIV scenario is a revisionary claim. It might be, but this result is unavoidable. In any case, I think that the result that we cannot claim knowledge of the validity of basic logical laws is a revisionary claim.

It becomes apparent that Wright's epistemology cannot avoid such revision. This raises the question of whether this consequence is devastating for Wright's epistemology — or whether Wright could just bite the bullet: after all, he bites the bullet in the external world case as well. Maybe it is just unavoidable that we do not possess more than mere entitlement for validity, just as it is unavoidable that we cannot claim to know that we are not BIVs. In the next chapter, I will argue that Wright's philosophy of mathematics faces similar problems. Again, the question arises of whether we can bite the bullet that we cannot claim more than entitlement.

In the next chapter I argue that there is the possibility to bite the bullet in the logico-mathematical cases. This is because the resulting framework would still achieve a lot because of what has been called **(Justification Generation)** above: except for cases of transmission-failure, proceeding in accordance with BFMs with entitled presuppositions — including basic logical rules and basic mathematical rules — delivers full internalistic warrant.

So far, so good. However, I also think that we should not bite the bullet in logico-mathematical cases. This is for a methodological, meta-epistemological reason I briefly discussed above: the **(Greater Meta-Epistemological Evil Principle)**.

Ask what the aims of Wright's and our framework were in the first place. Wright accepts conservativism for general reasons concerning the nature of evidence. He accepts the transmission-failure diagnosis in order to explain the intuition that Moorean arguments are unsatisfactory. And he has to postulate entitlements because internalist foundationalism is a non-starter without non-evidential warrants at the basic level. Apart from that, I think that Wright aims at getting as much as possible. In particular, I think that Wright would agree that it is desirable to account for as much internalistic knowledge as possible.

Now it seems that Wright only suggests making use of rule-circular arguments because he lacks alternatives. For example, in his (Wright 2004*a*), Wright considers using rule-circular arguments because he seems to think that the only alternative way to account for evidential warrants for validity is postulating a faculty of rational intuition, and that this position faces insurmountable objections.

However, nothing precludes us from using a defensible alternative course of justification once it is available. Thus, we obtain the following conditional: if we can find a defensible structure of justification SJ which, combined with the conservative diagnosis, the transmission-failure diagnosis, and the entitlement proposal leads to more internalistic knowledge than the use of rule-circular arguments, then we should postulate SJ instead.

In the last chapter, I shall examine such an alternative structure of justification for logic and mathematics. The idea is to (re)consider the proposal that basic logico-mathematical knowledge can be obtained by means of implicit definition. The hope is that suitably embedding Hale's and Wright's earlier **(Implicit Definition Thesis)** in the current framework yields more internalistic knowledge than postulating entitlements for basic principles directly.

Part III

New solutions to old problems

6 Entitled mathematics

In this chapter, I examine the possibility that **(HP)** is an entitlement of cognitive project. After motivating the claim that mathematical basic principles are entitlements, I discuss Wright's explication of a position based on this claim. It transpires that Wright's position yields revisionary sceptical consequences, because of the weak epistemic status of entitlement, and the transmission-failure diagnosis. However, the situation is not quite as bad as it may first seem. In the second part of this chapter, I argue that even the worst case — that all of arithmetic is a mere entitlement — is not devastating to our epistemology overall. A concessive position emerges as a fallback position, which I call ***semi-sceptical foundationalism***. It can be endorsed in case both Wright's position and the position defended in the next chapter prove to founder.

6.1 Entitlement and arithmetic

In Part II, I have argued that internalist foundationalists have to regard certain presuppositions as being warranted non-evidentially, and I have endorsed Wright's entitlements as an explication of the notion of non-evidential internalistic warrant. With this tool at hand, we can now finally reconsider the problems we started off with. First and foremost: how are our beliefs in mathematical axioms — qua basic principles about realms of abstract objects — warranted apriori?

My focus lies on arithmetic, but note that the arguments in this chapter generalize. In 1.1.1, I presented a general puzzle about the justification of axioms. The second-order Peano axioms are mathematical basic principles. In actual mathematical practice, they are simply postulated as being true without any explicit justification.

In chapter 2, we have seen that the neo-Fregean complicates the story by proving these axioms from **(HP)**. According to the neo-Fregean, we can acquire apriori knowledge of **(HP)** because it is an implicit definition of the number operator, and we can warrantedly assume the definitional success of **(HP)** ***by default***. With the notion of entitlement, we now possess a precise notion of warrant by default. However, there are at least two options as to what exactly is warranted by default. We could apply the idea of default entitlement at the level of presuppositions, and we could apply it to **(HP)** directly. In this chapter, I examine three positions according to which it is (the truth of) **(HP)** itself which is warranted by default. Although all of these positions have the consequence that **(HP)** has a weak epistemic status, they differ with respect to their predictions as to the epistemic status of the following two classes of relevant truths:

- (A) The theorems of pure arithmetic (i.e. the theorems of second-order PA).
- (B) The truths of applied arithmetic. That is: true mixed statements such as "The number of planets = 8".

The three positions are:

1. Wright's newest epistemology of arithmetic, outlined in (Wright 2009). It aims at vindicating knowledge claims of both (A) and (B).
2. A position I call ***semi-sceptical foundationalism***, without a phenomenon I call ***extended leaching***. It vindicates knowledge claims of (B), but not of (A).
3. Semi-sceptical foundationalism with extended leaching. It concedes that both (A) and (B) are merely entitled.

I will argue that, under certain conditions, even the most concessive position (3) is not devastating to our epistemology overall.

6.1.1 Axioms as entitlements

At first glance, it appears that **(HP)** meets all the conditions for entitlement of cognitive project. In fact, every proposition we take to be an axiom of a pure mathematical theory seems to meet these conditions. The first condition for a proposition to be an entitlement of cognitive project is that it is a Wright-presupposition of a relevant cognitive project, and the following princple is very plausible:

> **(Axiomatic Presupposition)** Whatever we take to be the axioms of a pure mathematical theory T, they are Wright-presuppositions of the cognitive project of finding out about the subject matter of T, for doubting the truth of the axioms would rationally commit one to doubting the significance of this project.

As an example, consider the project of finding out about the world of natural numbers. **(HP)** is a Wright-presuppositions of this project, because doubting its truth would rationally commit us to doubting the significance of this project (Pedersen 2009*b*, section 4).

Moreover, most axioms of pure mathematical theories also appear to meet condition (i) for entitlement of cognitive project: for in most cases we simply do not seem to have sufficient reason to believe that the relevant axioms or abstraction principles are untrue.[116] Consider the case of **(HP)** again. On the face of it, we can identify two ways of casting doubt upon its truth:

- Option 1 (mathematical doubt): mathematicians have cast doubt upon its truth by casting doubt on the consistency of PA. However, so far all attempts to provide sufficient reason to believe that PA is inconsistent have failed. Consider, for example, the recent discussion of Nelson's claim that PA is inconsistent (Nelson 2011).

- Option 2 (metaphysical doubt): nominalists have cast doubt upon the truth of mathematical basic principles on the basis of the conviction that there are no abstract objects. For example, nominalists might argue that numbers would be queer entities because they would not existent space and time, because there would be infinitely many of them, etc. One possible reaction to such worries is that such considerations fail to provide a sufficient reason to believe that pure mathematical theories are untrue for Moorean reasons. That is: our antecedent warrants for the truth of mathematical theories are stronger than the warrants we might possess for the philosophical premises that lead to the denial of their truth.[117] Of course, a stubborn nominalist will not be satisfied. What exactly counts as a sufficient reason to believe that a proposition is untrue remains to be clarified (see also Pedersen 2009*b*).[118]

I conclude that it is at least initially plausible that condition (i) is met by **(HP)**.

[116]There might be some exceptions: e.g. axioms postulating the existence of certain large cardinals.

[117]Note that the current question is whether we have sufficient reason to believe that mathematical theories are *untrue*. It is another matter entirely to explain how we can warrantedly believe that pure mathematical theories are *true*.

[118]Maybe we can also respond that the worries of the stubborn nominalists are irrelevant, because their world view is so different that their doubts do not matter to those who do not share this worldview. Consider the analogous case of someone who is an idealist, and already convinced that there is no external world. The idealist will not agree that the existence of an external world is an entitlement, but this should not worry the realist too much: the realist is concerned with the question of what can make *his* claims to knowledge rational, *assuming* a realist world view. Maybe there are fundamentally different ways of conceiving of the world, and doubts arising from such deep disagreement are irrelevant. These issues are very difficult to assess, and I cannot even begin to address them in this thesis. Note that this response might be in tension with the Instability Argument for (**Relevance Internalism**), which is based on the desire to arrive at an asymmetric internal epistemic situation.

6.1.2 A regress argument

So the claim that axioms are entitlements rests on condition (ii) — that the axioms cannot be justified without relying on other Wright-presuppositions in no more secure a prior epistemic standing. Again, I focus on the case of **(HP)**.

How can we justify **(HP)**? Either it is justified inferentially or not. In mathematics, inferential justification arguably means proof on the basis of other, more basic propositions. Let us assume that this is so:

> **(Logicality of Inference in Mathematics)** Every inferential justification in mathematics is deductive, i.e. by means of logic.[119]

Given this assumption, justifying **(HP)** inferentially requires deductively inferring it from other propositions which are in a more secure, prior epistemic standing. However, it seems that there are no such propositions. **(HP)** cannot be derived from (second-order) logic alone. One needs an additional non-logical premise (or collection of premises). However, this premise needs to be proof-theoretically at least as strong as **(HP)**. In this particular case, this renders it doubtful that this proposition can be in any better prior epistemic standing. A fortiori, consider any proposition X that entails **(HP)** in the background logic. The question would arise of how X is justified. After all, X entails second-order arithmetic. So we have to appeal to another premise (or collection of premises) that entails X.

We thus enter an epistemic regress (or end up in a circle) — unless, of course, at some point we do not have to justify the relevant basic proposition inferentially. But what are the non-inferential alternatives? And if there is such a way, then why can we not justify **(HP)** non-inferentially in the first place?

Is there a way to justify **(HP)** *non-inferentially*? Firstly, one might refer to some kind of rational intuition. But apart from the question whether such a proposal is tenable at all — and we have seen in 1.3.1 that this option is not very promising — the use of rational intuition (implicitly) relies on the presupposition that it is reliable. And how are we to justify this presupposition?[120] Let us leave the swamps of rational intuition then, and consider alternatives.

[119] This is a mathematical analogue of the **(Logicality of Inference)** assumption from 5.3.

[120] Of course, one might think that the reliability of rational intuition is warranted by means of entitlement: this might be a position worth exploring for someone who thinks that rational intuition is an option, but it is not a position to be explored in this thesis.

Secondly, one might think that one can improve **(HP)**'s epistemic standing indirectly by justifying its consistency or satisfiability. Hilbert, in a letter to Frege, can be interpreted as defending such a view:

> If the arbitrarily posited axioms together with all their consequences do not contradict one another, then they are true and the things defined by these axioms exist. For me, this is the criterion of truth and existence. (Hilbert to Frege 1899, in Frege 1980, pp. 39f)

However, even if the position were correct in general,[121] we would still have to justify **(HP)**'s consistency. Assuming that this justification needs to be inferential, this proof needs to be carried out in some background system. For example, Boolos (1990) has shown that Frege Arithmetic is consistent relative to second-order arithmetic. However,

> in showing FA consistent relative to analysis [=second-order arithmetic] the consistency of FA is held hostage to the consistency of analysis. Whatever its merits, the relative consistency proof does nothing to establish that. Due to Gödel's second incompleteness theorem, the best one can do with respect to the consistency of analysis — indeed, any theory strong enough to express elementary arithmetic — is to establish it relative to some other theory T of consistency strength greater than that of analysis itself. Thus, the consistency of analysis will be held hostage to that of T. The pattern repeats itself, and a regress of relative consistency proofs involving stronger and stronger theories results. (Pedersen 2009*b*, p. 23)

Thus, we cannot justify **(HP)**'s consistency without relying on something in no more secure a prior epistemic standing, and enter an epistemic regress again. It should be clear enough that the same argument applies to satisfiability (Pedersen 2007).

Neither Pedersen (nor Hilbert, of course) consider the option of justifying consistency by non-logical means.[122] For example, one might think one can argue for the consistency of **(HP)** on the basis of the fact that no one has yet observed an inconsistency. However, one might think that such kinds of establishing consistency are inadequate for mathematics. In particular, there is a risk that our justification of consistency turns out to be aposteriori.

[121] It is, of course, another matter entirely whether this train of thought is conducive to a Platonistic interpretation of the axioms.
[122] I owe this point to Robert Williams.

Since these three options — proof on the basis of further, more basic premises, rational intuition, and indirect justification — are exhaustive, so the thought goes, **(HP)** cannot be justified without relying on something in no more secure a prior epistemic standing. In the next chapter, I argue that these options are not exhaustive, and that we should reject **(Logicality of Inference in Mathematics)**. But let us assume, for the sake of the argument, that they are.

Two notes are in order. Firstly, one might think that one can still justify some mathematical theories on the basis of other, more basic theories. Maybe all mathematical theories rest on set theory. Of course, the discussed regress would then arise for the axioms of set theory, and it might well turn out that the axioms of set theory are entitlements. Secondly, as I mentioned above, the same argument can be applied to the conjunction of the second-order Peano axioms. This raises the question of why we should focus on **(HP)**, as opposed to these axioms. Does it really matter which axiomatization we choose?

We need to keep both questions in mind, because they suggest that the positions discussed in this chapter have nothing to do with neo-Fregeanism. In the course of this chapter, it will become apparent that it does make sense to focus on **(HP)** (and on abstraction principles).

6.1.3 Wright on entitlement and arithmetic

In his most recent work on the epistemology of mathematics (Wright 2009), Wright aims at vindicating a form of neo-Fregeanism by arguing that **(HP)** is an entitlement.

Wright's argument for **(HP)** being an entitlement diverges from the simple train of thought above. One way to interpret Wright (2009, §6) is, roughly, this: firstly, the good standing of a project's concepts is a Wright-presupposition of this project. Thus, it is also a Wright-presupposition of this project that the means of fixing these concepts are successful. Now, the concept of number is implicitly defined by **(HP)**: "#" means whatever renders **(HP)** true. Thus, **(HP)** must be true in order for the arithmetical concepts to be in good standing. Therefore, (the truth of) **(HP)** is a Wright-presupposition of the arithmetical project. It is also an entitlement of cognitive project, because the other conditions for entitlement are met as well.

According to this interpretation of Wright, **(HP)** is a Wright-presupposition of the arithmetical project — the project of finding out about the world of (natural) numbers — because doubting **(HP)**'s truth would rationally commit one to doubting that the concept of number is in good standing, and thus rationally commit one to doubting the significance of finding out about numbers.

However, there is a second project of which **(HP)** might be a Wright-presupposition.[123] Wright can also be read as expressing the thought that there are projects of ensuring that our concepts are in good standing, and that such projects rest on entitlements for the success of our meaning-fixing devices, and thus on the truth of meaning-fixing principles. For example: ensuring that the concept of number is in good standing rests on the possibility of rationally regarding **(HP)** as being true, because its truth is required to successfully fix the concept of number. This thought can be summarized as follows:

> **(Metasemantic Presupposition)** Abstraction principles are *Wright*-presuppositions of the project of ensuring that our mathematical concepts are in good standing.[124]

It does not matter here which interpretation we take. Let us assume that **(HP)** is a Wright-presupposition of some significant cognitive project, and let us consider condition (i) for **(HP)** being an entitlement. Do we have sufficient reason to believe that **(HP)** is untrue? Of course, Wright does not think so. But Wright does not think the issue is settled by just considering mathematical doubts and metaphysical doubts as above. According to Wright (2009, §7), more needs to be done here. In particular, he thinks that the classical worries concerning **(HP)**'s status as a definition become relevant. For example: the Bad Company objection (see 2.4.2). Since **(BLV)** has the same form as **(HP)**, so the thought goes, we need to possess reasons for the cases being different, in order to be able to claim an entitlement. Wright discusses rejectionist worries as well, but I cannot discuss this here. What has been said suffices to make the following point: regarding **(HP)** as implicitly defining the concept of number opens up space for relevant sceptical alternatives which need to be considered as potentially sufficient reasons to doubt **(HP)**'s truth.

Of course, we should not set the standard too high if we want to avoid sceptical results.

[123] For a similar observation, see (Pedersen 2009*b*, section 4).

[124] Pedersen (2009*b*) suggests a similar interpretation. However, his observation is importantly different in that he suspects that (HP) might be a presupposition of the project of *fixing* a concept of number in good standing. Pedersen overlooks the fact that cognitive projects are projects of *finding out* about the world. Fixing meaning is a very different kind of project, although there certainly are intimate connections between the two projects.

I agree with Pedersen (2009*b*, p.15) in that only known alternatives should be relevant. For example, Bad Company considerations should only become relevant as soon as bad companions are known. Only in the light of examples such as Russell's Paradox and **(BLV)** do we have to rule out that **(HP)** is not such a case. The entitlement proposal should be interpreted as predicting that Frege possessed an entitlement for regarding **(BLV)** as successfully fixing the concept of extension, and thus an entitlement for **(BLV)** being true, before he received Russell's letter. I take this to be a constraint on what exactly sufficient reasons to doubt are.

In what follows, I assume that condition (i) is met. It has already become apparent that condition (ii) requires further work. In his paper, Wright just makes the general observation that:

> it would be fanciful to suppose that final assurances might be achieved that any particular concept was in definitive good standing. The most that one might hope to do would be to address specific grounds for doubt. And in any case—more important—any investigation of the matter would presuppose—or ancestrally presuppose—an antecedent conceptual apparatus whose good standing would have to be taken for granted. (Wright 2009, p. 9)

Much more needs to be said here. Although it is certainly true that no final assurance can be achieved as to the good standing of any particular concept — just as one cannot achieve final assurance about anything — I do not think one can dismiss the possibility so quickly that one can justify the good standing of at least some concepts without falling into regress. For example, we have seen that we can come to know of the existence of ordinary objects around us on the basis of sense perception. Suppose that we come to know of the existence of a certain dog and decide to call it "Fido". Using my knowledge of the existence of Fido, I will be able to acquire knowledge of the fact that the concept of (being) Fido is in good standing. There is no regress here.

However, we have seen above that there is a prima facie compelling case to be made for the regress clause being met in the mathematical case. If the argument above was correct, then condition (ii) for **(HP)** being an entitlement would be met as well. It would follow that **(HP)** is an entitlement of cognitive project in accord with the **(Metasemantic Presupposition)** principle. Let us assume, for the sake of the argument, that this is so.

Above, I raised the question of whether there is any special role for **(HP)**. Note that Wright's argument (and the **(Metasemantic Presupposition)** principle in particular) not only assigns a special role to **(HP)**, but also connects the proposal to classical neo-Fregeanism. **(HP)** is regarded as fixing the concept of number, and the classical objections to neo-Fregeanism become relevant again.

6.1.4 Wright on avoiding a leaching worry for arithmetic

Very well then. Suppose that **(HP)** is an entitlement. Can this result serve as the basis for a satisfying epistemology of arithmetic? Unfortunately, some reflection seems to reveal that the position yields revisionary sceptical consequences.

The basic problem is this: Wright accepts the **(Limit)** principle[125] which says that the epistemic status of the conclusion of a warrantedly drawn inference cannot be any stronger than the epistemic status of its epistemically weakest premise. Thus, the epistemic status of inferences drawn from **(HP)** cannot be any stronger than the epistemic status of **(HP)**, which is an entitlement. As a consequence, we cannot acquire internalistic knowledge of p by virtue of proving p from **(HP)**: we can only acquire inferential entitlements (in the sense defined in 4.5.1).

Assuming that our warrants for the second-order Peano axioms — and our warrants for all theorems of pure arithmetic — are canonically based on proofs from **(HP)**, this has the consequence that we cannot claim to know any arithmetical truth, but only claim (inferential) entitlements. Our epistemology would massively violate the **(Arithmetical Knowledge)** constraint.

Wright deems this to be an unacceptable sceptical consequence which he has to avoid (Wright 2009, §8). However, he also thinks he can avoid this consequence by tweaking the suggested structure of justification. His argument proceeds in four steps (Wright 2009, p. 18). Firstly, he suggests that the biconditional **(HP)** can also be conceived of as a pair of basic rules corresponding to both directions of the biconditional, because they are proof-theoretically equivalent to **(HP)** in our background logic:

(Hume Rules)

$$(HP_{\rightarrow})\frac{\#(F)=\#(G)}{\exists R(Bij(R,F,G))} \quad (HP_{\leftarrow})\frac{\exists R(Bij(R,F,G))}{\#(F)=\#(G)}$$

[125] See also (Wright 2009, §8).

Secondly, Wright assumes that, if we can argue that the statement **(HP)** is an entitlement, then we will also be able to argue that the *soundness* of the **(Hume Rules)** is an entitlement. Read soundness as *truth-preservation*. Then the assumption is correct. On the face of it, **(HP)** just is a statement of the soundness of the **(Hume Rules)** in this sense.

Thirdly, Frege's proof of the second-order Peano axioms from **(HP)** can be carried out using the **(Hume Rules)** instead of the axiom **(HP)**. The new proof does not have any premises — only discharged assumptions — and only makes use of rules with entitled presuppositions: for both the **(Hume Rules)**, and the used logical rules are entitlements.

Now, the **(Justification Generation)** principle implies that we can acquire internalistic knowledge of the conclusion of a proof with only internalistically known premises, all of whose presuppositions are entitlements. If a proof has no premises (only discharged assumptions), all its premises are trivially internalistically known. Thus, fourthly, the **(Justification Generation)** principle applies and we are able to acquire internalistic knowledge of the proof's conclusion.[126] Wright would have provided a route to internalistic knowledge of the second-order Peano axioms and their deductive consequences.

6.1.5 Transmission-failure

However, Wright's trick leaves us in an odd situation. The following question becomes pressing: can our right to claim knowledge really depend on such a small modification of the underlying deductive system? It seems odd that endorsing the **(Hume Rules)** yields internalistic knowledge of the second-order Peano axioms, whereas endorsing the statement **(HP)** leaves us with merely entitled arithmetic. On the other hand, this odd situation seems to be a straightforward consequence of **(Justification Generation)**. I will come back to this issue.

First, I argue that Wright's proposal is more concessive than one might expect, even if we overlook this odd situation for a moment. This is because transmission-failure considerations imply that the *statement* **(HP)**, and maybe even its instances, cannot be claimed to be known. My argument rests on two assumptions:

1. Conservativism holds for the **(Hume Rules)**, i.e. it is a necessary condition for

[126]Note that this is the same thought that led to an investigation of rule-circular arguments in the last chapter.

being able to use the **(Hume Rules)** to extend internalistic knowledge that one possesses an antecedent warrant for the proposition that they are truth-preserving.

2. **(HP)** expresses the proposition that the **(Hume Rules)** are truth-preserving.

Both assumptions seem very plausible, and Wright seems to be committed to them. As to assumption 1, it is hard to see why the arguments for conservativism should not apply to the mathematical case. As to assumption 2, Wright directly moves from the claim that we possess an entitlement for **(HP)** to the claim that we possess an entitlement for the soundness of the **(Hume Rules)**.[127]

However, assumption 1 is in need of clarification, because we have to distinguish between particular and general soundness claims. Whereas **(HP)** plausibly expresses the fact that *both* **(Hume Rules)** are truth-preserving in general (this is assumption 2), the truth-preservation of particular instances of the **(Hume Rules)** is plausibly expressed by the following two claims (note that F and G are not variables here, but terms standing for particular concepts):

(HP->Sound) $\#(F) = \#(G) \rightarrow \exists R\,(Bij\,(R, F, G))$,

(HP<-Sound) $\exists R\,(Bij\,(R, F, G)) \rightarrow \#(F) = \#(G)$

Corresponding to the logical case, there are two versions of conservativism about the **(Hume Rules)**:

- Firstly, one might claim that it is the truth-preservation of particular instances that feature as presuppositions of particular Hume Rule steps.

- Secondly, one might hold that it is the general soundness claim that is the proper presupposition of all Hume Rule steps, i.e. **(HP)** (by assumption 2).

In both cases, the respective soundness claims have to remain entitlements. This is because the only way to upgrade their epistemic status would be via rule-circular reasoning, and rule-circular arguments fail to transmit warrant because they exhibit the **(Information Dependence Template)**.

[127] In any case, Wright is committed to the claim that **(HP)** directly entails that the **(Hume Rules)** are truth-preserving. Thus, he will not be able to avoid the worry by denying assumption 2, because of the phenomenon of presupposition expansion (see 6.1.6).

Consider the following argument for **(HP)** in the system arising from adding the **(Hume Rules)** to a standard second-order deductive system.

1	$\#F = \#G$	Ass.
2	$\exists R(Bij(R,F,G))$	1, $HP_{\rightarrow}$
3	$\#F = \#G \rightarrow \exists R(Bij(R,F,G))$	1,2, CP
4	$\exists R(Bij(R,F,G))$	Ass.
5	$\#F = \#G$	4, $HP_{\leftarrow}$
6	$\exists R(Bij(R,F,G)) \rightarrow \#F = \#G$	4,5, CP
7	$\#F = \#G \leftrightarrow \exists R(Bij(R,F,G))$	2,5, $\leftrightarrow$-Intro.
8 (=HP)	$\forall F \forall G(\#F = \#G \leftrightarrow \exists R(Bij(R,F,G)))$	7, UG

If **(HP)** is a presupposition of Hume Rule steps, then the argument instantiates the **(Information Dependence Template)** because its last line (8) is **(HP)**, and the argument contains Hume Rule steps. A previously entitled **(HP)** will have to remain a mere entitlement by **(No Knowledge Claims)**.

On the other hand, if it is instances of **(HP->Sound)** and **(HP<-Sound)** which feature as presuppositions, then both sub-arguments (the step from 2 to 3; the step from 6 to 7) will instantiate the **(Information Dependence Template)** because lines 3 and 7 are the relevant instances of **(HP->Sound)** and **(HP<-Sound)**. Warrant cannot be transmitted to lines 3 and 7, and previously entitled instances of **(HP->Sound)** and **(HP<-Sound)** will have to remain entitlements by **(No Knowledge Claims)**. A fortiori, because of the **(Limit)** principle, subsequent lines of the argument — including **(HP)** — must remain entitlements as well.

Hence, regardless of whether it is general truth-preservation, or the truth-preservation of instances that feature as presuppositions of Hume Rule steps, **(HP)** has to remain an entitlement. Thus, Wright's epistemology of arithmetic is concessive just as his epistemology of logic and his epistemology of perception are concessive: the presuppositions of the basic belief-forming methods have to remain entitlements. Unfortunately, one of our aims was to explain how we can claim knowledge of **(HP)**. We would have to give up this aim.

Note that, just as in the logical case, the situation looks worse if it is the truth-preservation of instances that feature as presuppositions. If this is the case, then it cannot even be claimed that instances of **(HP->Sound)** and **(HP<-Sound)** are known. For example, we would not be able to claim to know the following:

> **(Concrete Equinumerosity)** If the number of knives on the table is identical to the number of forks on the table, then the knives and the forks are in one-one correspondence.

For the conditional is of the form **(HP->Val)**, and as such expresses the claim that a particular instance of $(HP_{\rightarrow})$ is truth-preserving. This certainly is a revisionary sceptical consequence. It seems entirely appropriate to make a claim to knowledge of **(Concrete Equinumerosity)** in an ordinary conversation.

6.1.6 Presupposition expansion and extended leaching

If everything goes well, the account will still enable us to claim knowledge of the second-order Peano axioms and its theorems. For example, we can claim to know that every number has a successor, and that there are infinitely many prime numbers. A lot would have been achieved.

Note, however, how odd the consequences of the proposal are. We would be able to claim knowledge of the Peano axioms, but could not claim knowledge of **(HP)**, and maybe could not even claim knowledge of **(Concrete Equinumerosity)**. And does it not look suspicious that such a small change in the underlying deductive system can enable us to claim to ***know*** that every number has a successor, whereas we could only claim an (inferential) ***entitlement*** before we made this change? I think this suggests that something ***must be wrong*** with Wright's argument. In particular, it seems as if Wright's position should either yield knowledge claims to **(HP)** and **(Concrete Equinumerosity)** as well, or imply that we only possess entitlement across the board.

On the other hand, it looks as if **(Limit)**, **(Justification Generation)**, and the **(Information Dependence Template)** just imply this huge epistemological difference between endorsing the **(Hume Rules)** and endorsing **(HP)** as an entitled axiom. And these principles are not negotiable here, because they form a crucial part of our epistemological framework.

I close my discussion of Wright's position by sketching the outlines of an argument to the effect that Wright's argument is flawed after all, and which might avoid the odd situation. The key to a resolution of the situation is that there can be hidden cases of transmission-failure.

According to the transmission-failure diagnosis, we can only acquire claimable knowledge of p by virtue of proving it on the basis of the **(Hume Rules)** and second-order logic if p is not already an entitled presupposition of the **(Hume Rules)**. Thus, if the second-order Peano axioms were entitled presuppositions of the **(Hume Rules)**, they could not be claimed to be known on the basis of such a proof.

I sketch the outlines of an argument to the effect that the second-order Peano axioms will indeed be entitled presuppositions of the **(Hume Rules)**, if **(HP)** is. First, note that it is plausible that certain consequences of entitled presuppositions also count as entitled presuppositions. Consider the case of visual perception. If it is an entitled presupposition of perceptual BFMs that there is no Cartesian demon, then it is certainly also an entitled presupposition of perceptual BFMs that there is no Cartesian demon *who likes playing chess*. Call this phenomenon *presupposition expansion*.

The thought is that there is presupposition expansion in the mathematical case as well. Suppose there is a consequence relation — call it R-consequence — such that:

1. The R-consequences of an entitled presupposition are also entitled presuppositions.

2. The second-order Peano axioms are R-consequences of **(HP)**.

Suppose further that **(HP)** is an entitled presupposition of the Fregean proof. Then 1 and 2 imply that the second-order Peano axioms are entitled presuppositions of the Fregean proof as well. If this was correct, then because of the transmission-failure diagnosis, we could not upgrade the epistemic status of these axioms by virtue of deriving them on the basis of the Fregean proof. This suggests that it is not possible to upgrade the epistemic status of these axioms on the basis of any proof that rests on the **(Hume Rules)** and logical reasoning, for it is plausible that any proof on this basis will display transmission-failure, if the Fregean proof does.

This would resolve the incredulous stare that comes with the result that reconceiving of a biconditional as a pair of rules makes such a huge epistemological difference. However, this also means that Wright's proposal yields much more severe sceptical consequences than he envisages. In fact, it would yield the revisionary sceptical consequence that not only **(HP)**, but also the second-order Peano axioms, and all of second-order arithmetic[128] cannot have an epistemic status above (inferential) entitlement.

I think it is plausible that **(HP)** is a presupposition of the modified Fregean proof. Even if it is not a presupposition of instances of the **(Hume Rules)**, both directions of **(HP)** will be presuppositions of relevant universal generalization steps — namely those steps that generalize particular applications of the **(Hume Rules)**. The crucial question is whether there is a consequence relation with properties 1 and 2. Clearly, there is a phenomenon of presupposition expansion in the perceptual case. However, the inferential gap between **(HP)** and the second-order Peano axioms is relatively wide. So the notion of R-consequence needs to be relatively wide as well.

The thought that the notion of R-consequence is that wide is reminiscent of the rejectionist objection to neo-Fregeanism discussed in 2.4.3. Wright might respond in the same way he earlier responded to rejectionism. For example, he might respond that conceiving of the axioms of arithmetic as presupposition of the Fregean proof just gets the epistemic order wrong. Both the objection and the response to it merit further investigation.

In any case, if the notion of R-consequence was wide enough, Wright's suggestion to base arithmetic on entitled **(Hume Rules)**[129] would imply that not only **(HP)** and its instances, but all of arithmetic cannot be claimed to be known. And this would be a pretty strong revisionary sceptical consequence. Claims to arithmetical knowledge are paradigmatic claims to knowledge.

6.2 Semi-sceptical foundationalism

Be that as it may. In the remainder of this chapter, I argue that it would not be devastating to our epistemology overall if no statement of pure arithmetic could be claimed to be known. This motivates a different position, which does not rest on any tricks in the first place, but straightforwardly concedes that all of arithmetic is a mere entitlement. I call it *semi-sceptical foundationalism* (SSF). It rests on the thoughts that (i) the primary epistemic

[128] That all of second-order arithmetic is affected follows from the **(Limit)** principle, and the assumption that pure arithmetical truths are canonically justified on the basis of the second-order Peano axioms.
[129] Entitled in the sense that all its presuppositions are entitled.

role of pure arithmetic is that it generates inference rules that can be applied to extend knowledge in non-mathematical projects, such as physics or everyday reasoning, and (ii) that this role would not be qualified if all of pure arithmetic was merely entitled.

6.2.1 The idea

According to SSF, **(HP)** is an entitled presuppositions for a different reason than those expressed by the principles **(Axiomatic Presupposition)** and **(Metasemantic Presupposition)**:

> **(Presupposition of Application)** The basic principles of our mathematical theories are presuppositions of all non-mathematical projects in which these theories are applied to extend knowledge, and in which these knowledge-extending applications are essential. For doubting the relevant mathematical basic principles would rationally commit us to doubting the significance of these non-mathematical projects.

This is the fundamental difference to Wright's proposal. There are non-mathematical projects in which arithmetic is applied as a means to extend knowledge, and in which these applications are essential. **(HP)** is a presupposition of such projects because it is the basic principle underlying arithmetic. Because the other conditions of entitlement are met as well, so the thought goes, **(HP)** is an entitled presupposition of such projects. For example:

- Arithmetic and everyday reasoning: arithmetical reasoning is required in cognitive projects of everyday life. For example: the project of calculating whether I have enough money to buy a certain number of sweets, or the project of finding out how to equitably divide a certain number of sweets. We successfully pursue such projects by applying arithmetical lines of reasoning. Doubting the truth of **(HP)** would rationally commit us to doubting the significance of these projects, for it undermines the goodness of the used mathematical inferences. Thus, **(HP)** is a presupposition of such projects, and, because the other conditions for entitlement are met, an entitled presupposition.
- Arithmetic and biology: in biology, there are genuinely mathematical explanations making use of arithmetical reasoning. For example, biologists explain the prime

lifecycle of certain cicada types by noting that prime lifecycles minimize intersection with the lifecycles of other species, and that this is evolutionary advantageous (Baker 2005). Doubting the truth of **(HP)** would rationally commit one to doubting the significance of the project of explaining prime lifecycles. Thus, **(HP)** is a presupposition of such explanatory projects, and, because the other conditions for entitlement are met as well, an entitled presupposition.

All this requires a lot more work, but I hope it suffices to initially motivate the **(Presupposition of Application)** principle. The principle is significant because it makes it look far less devastating if all of pure arithmetic is an entitlement. The thought is that (a) the central role of mathematical theories is knowledge-extending application, (b) that our epistemology of mathematics can be regarded as satisfactory as long as it accounts for this role, and (c) that merely entitled pure theories do not undermine this role because of the **(Justification Generation)** principle.

Note that Wright has all the resources to account for (a), (b), and (c). However, he seems to regard pure mathematics as a cognitive project of its own, i.e. a project of extending mathematical knowledge. This conception of mathematics makes it look devastating if pure mathematical theories cannot be claimed to be known. If there is another role for pure mathematical theories that is not undermined by our epistemological concessions, the situation looks less problematic.

So far, so good. But how exactly can entitled mathematical theories be used to extend knowledge in other projects? Suppose we base pure arithmetic on an entitled **(HP)** or on entitled **(Hume Rules)**, and let us concede that leaching occurs in the sense that no pure arithmetical statements can be claimed to be known.[130] The idea is that we can extract *deductive rules* from our pure arithmetical theory which can be used to expand knowledge in other areas. For example: in the sciences or in everyday reasoning.

Here is how. First, we can use our pure theory to infer conditionals. These conditionals are inferential entitlements. Secondly, we can conceive of these conditionals as rules with entitled presuppositions. For their presuppositions just are these conditionals. Thirdly, because the presuppositions of these rules are entitled, we can use them to extend internalistic knowledge in other cognitive projects, because of **(Justification Generation)**.

[130] We might also remain neutral about such leaching — the point is that it does not matter whether there is leaching or not.

In short:

(Extraction Principle) Conditionals we obtain on the basis of entitled axioms or rules in a pure mathematical theory can be used to extend internalistic knowledge in other projects.

Since the **(Extraction Principle)** is sufficient for accounting for the central role of mathematical reasoning, so the thought goes, it does not matter whether pure theories are merely entitled. According to SSF, pure mathematics should be conceived of as the enterprise of generating new inferences we can rely on to extend claimable knowledge in other areas.

6.2.2 A toy example

Let me go through a toy example of how the **(Extraction Principle)** is supposed to work. Consider the following theorem of pure arithmetic:

(Sample Theorem) For any two prime numbers n and m, the least common multiple of n and m is the product of n and m.

In Frege Arithmetic, we can formalize this as follows:

(Sample Theorem') $$\begin{gathered}(\mathit{Prime}\,(\#F) \wedge \mathit{Prime}\,(\#G)) \rightarrow \\ \forall x\,(\mathit{LCM}\,(x, \#F, \#G) \leftrightarrow x = \#F\#G)\end{gathered}$$

(Sample Theorem') is a universally quantified conditional. Suppose that **(Sample Theorem')** cannot be claimed to be known, but only be claimed to be an (inferential) entitlement. We can conceive of this (entitled) conditional as an (entitled) rule we can endorse to extend internalistic knowledge in other projects as follows (note that F and G are *variables* here):

(Sample Extracted Rule) $\frac{\mathit{Prime}(\#F) \wedge \mathit{Prime}(\#G)}{\forall x(\mathit{LCM}(x, \#F, \#G) \leftrightarrow x = \#F\#G)}$

The reason is that **(Sample Theorem')** is the entitled presupposition of the **(Sample Extracted Rule)**.

Now, how can this rule be put to work to extend internalistic knowledge? We need a case in which a subject S already possesses internalistic knowledge of the premise of an instance of the rule. And this generates a worry. The premise will include arithmetical terms. If all of arithmetic is a mere entitlement, then how can S come to possess internalistic knowledge of such a premise?

First, note that, although the premise will include arithmetical terms (namely "#" and "*Prime*"), it does not have to be a *pure* arithmetical statement, such as "$Prime\,(13) \wedge Prime\,(17)$". Indeed: such a statement cannot be internalistically known (by assumption). However, nothing said so far entails that *mixed* arithmetical statements such as the "The number of knives on the table is prime and the number of forks on the table is prime" cannot be internalistically known. In fact, it seems we *can* acquire internalistic knowledge of such statements by cleverly using the **(Extraction Principle)**. Let me explain.

Our pure arithmetical theory (Frege Arithmetic) proves the following conditionals (note that F is a variable):

(Thirteenfold Existence) $\exists_{13}x\,(F\,(x)) \rightarrow \#F = 13$

And:

(Primeness of 13) $\#F = 13 \rightarrow Prime\,(\#F)$

By logic, we obtain:

(Primeness of Thirteenfold Existence) $\exists_{13}x\,(F\,(x)) \rightarrow Prime\,(\#F)$

Now suppose that an epistemic agent Hero faces a table with 13 knives on it. Our epistemological framework allows for Hero acquiring internalistic knowledge of $\exists_{13}x\,(Knife\,(F))$.[131] By the **(Extraction Principle)** and **(Justification Generation)**, Hero can then acquire internalistic knowledge of the fact that $Prime\,(\#_x Knife\,(x))$ — i.e. that the number of knives on the table is prime — by virtue of inferring it from $\exists_{13}x\,(Knife\,(F))$, using the rule extracted from **(Primeness of Thirteenfold Existence)**.

Now suppose that a biologist already possesses internalistic knowledge of the fact that the length of the lifecycles of two cicada types in years are 13 and 17, and thus prime. According to the **(Extraction Principle)**, he or she can use the **(Sample Extracted Rule)** to acquire internalistic knowledge of the fact that the least common multiple of the length of their lifecycles — and hence the length of the period after which they will "meet

[131] By a successive application of entitled perceptual belief-forming methods and entitled logical rules.

again" — is the first number multiplied by the second number. The biologist can now use further extracted rules[132] to come to possess internalistic knowledge of the fact that the length of the sought period is 221. On this basis, the biologist can make further inferences, use this newly obtained information in explanations, etc.

This completes my toy example. In the remainder of this chapter, I discuss some features of the proposal, consider some objections, and examine which of the constraints laid out in 1.2 the proposal meets, and which it fails to meet.

6.2.3 SSF beyond arithmetic

The idea generalizes beyond arithmetic. As an example, consider the following instance of the **(Presupposition of Application)** principle:

- Analysis and Newtonian physics: the standard way of doing Newtonian physics makes heavy use of analysis. Not only are its concepts intertwined with mathematical concepts: analysis is also used to make predictions and explanations. Both predictions and explanations rest on the possibility of using analysis in inferences extending empirical knowledge — inferences that would otherwise be impossible or a lot more difficult to make.[133] Moreover, doubting the axioms of analysis would rationally commit one to doubting the significance of the standard (analysis-endorsing) way of doing Newtonian mechanics, for it would rationally commit one to doubting the goodness of the used mathematical inferences, or the good standing of some physical concepts.

Thus, the axioms of analysis are Wright-presuppositions of (the standard way of doing) Newtonian physics.[134]

There are examples beyond physics. However, the mature state of physics as a scientific discipline, and the heavy use of mathematical concepts in this discipline make it easier to find suitable examples. In any case, the claims made here require further argument and a

[132] I omit the details, because, by now, the reader will be able to see how this can happen.

[133] This is of course compatible with the claim that there are nominalistic ways of doing Newtonian physics. I come back to the issue of dispensability below.

[134] The above passage brings to light that there are two ways in which mathematical basic principles can be Wright-presuppositions in accordance with the **(Presupposition of Application)** principle. First, they can be Wright-presuppositions because they are required in certain lines of reasoning. Second, they can be Wright-presuppositions because the concepts of scientific theories rest on the good standing of mathematical concepts, and doubting the relevant mathematical basic principles would rationally commit us to doubting that these concepts are in good standing.

detailed investigation of examples. This is beyond the scope of this thesis. But it is worth mentioning that the applicability of mathematics has been a recent focus of attention in the philosophy of mathematics, and some of these investigations fit well together with the picture I am sketching here. For example, Bueno and Colyvan argue that the central role of mathematics is to facilitate inferences in the sciences, in the context of prediction, explanation, and unification (Bueno & Colyvan 2011).

SSF fits well together with an important component of neo-Fregeanism. This becomes apparent by asking the question of how we can ensure that, in general, our reconstructed mathematical theories are conducive to their application. One answer is: **(Frege's Constraint)**. If we follow Frege's advice and make the application of our mathematical theories immediate in their foundations, we ensure that we can apply these theories as intended. This provides a reason to focus on **(HP)** as opposed the second-order Peano axioms. For **(HP)** makes the application of cardinal numbers immediate. At least SSF for arithmetic is naturally combined with the use of abstraction principles.

6.2.4 Inapplicable theories

One might worry that the proposal is not relevant to pure theories that are not designed to be applied, have never been applied, or will never be applied. For example: large cardinal arithmetic. However, to generate an objection to SSF on this basis, one needs to extract undesired consequences from this fact.

One might think that SSF entails that these theories are not even warranted by means of entitlement, because there is no suitable presuppositional role for their basic principles. However, someone endorsing SSF is not committed to this claim: pure theories can still be entitled because they are presuppositions in accordance with the **(Axiomatic Presupposition)** principle, or the **(Metasemantic Presupposition)** principle, or both.

This might raise the worry that the friend of SSF cannot argue that the concession that the pure theories are mere entitlements is not devastating in these cases. For the argument was based on the thought that there is another role for mathematical enquiry than to extend mathematical knowledge: namely extending knowledge in other areas. In the case of inapplicable theories, so the thought goes, the proposal is still unacceptably sceptical because there is no such further role.

I respond that nothing said so far excludes that merely entitled unapplied theories *indirectly* bear on scientific reasoning and everyday reasoning. They might do so because pure mathematical enquiry helps with developing better mathematical theories that can eventually be applied, or because it helps with unifying mathematical theories, which in turn yields new applicable mathematical theories.

6.2.5 Mixed statements

Another objection to the current proposal is that mixed propositions such as "The length of the lifecycle of these cicadas in years = 13" — propositions including both mathematical and other concepts — cannot be internalistically known, if no pure arithmetical statement can be internalistically known. One way to argue for this claim is this: a mixed proposition certainly entails the existence of a number, which is a pure arithmetical statement and thus — by assumption — merely entitled. Thus, the above sketched route to internalistic knowledge of such statements must be blocked somewhere.

There are at least three things to say in response to this worry. The first two responses are based on the epistemological framework outlined in Part II of this thesis; the third response is that there is a fallback position that is still interesting.

Firstly, the framework allows for cases in which p can be internalistically known although an immediate consequence of p cannot be internalistically known. For example: Moorean reasoning. We can internalistically know that we have two hands, although an immediate consequence of this statement — that there is an external world — cannot be more than an entitlement.

Secondly, we already tacitly allowed for internalistically known mixed statements in the logical case. Wright's framework would be radically sceptical if it would not allow for the possibility of acquiring internalistic knowledge of mixed logical statements — i.e. statements including logical vocabulary (e.g. "There is a table in front of me $\wedge$ I have two hands"). Now, why should mixed mathematical statements be any more problematic than mixed logical statements? Everyone who accepts Wright's response to scepticism in the logical case needs to provide additional principled reasons for why this should be so.

A third response is to retreat to an even more concessive fallback position. Let us call the possibility that mixed mathematical statements cannot be claimed to be known *extended leaching*. Even if extended leaching occurs, nothing said so far precludes the possibility to use mathematical reasoning to come to possess internalistic knowledge of propositions which do not contain any mathematical vocabulary. For example, if some mathematical reasoning about the growth of populations allows us to infer that some population of Fs now consist of at least 2000 individuals, then even if "$\#F \geq 2000$" has to remain a "mere entitlement" and cannot be internalistically known on this basis, we might come to internalistically know "$\exists_{2000}x\,(Fx)$" on this basis. Thus, conceding extended leaching still does not have to amount to full-blown scepticism about projects of acquiring knowledge in other areas.

However, whether this response is available depends on the possibility of nominalizing relevant discourses using mathematical vocabulary — i.e. the possibility of rewriting such discourses in a way that does not make use of mathematical concepts. First and foremost, we would have to assess whether we can nominalize relevant scientific theories. Only if we can dispense with mathematical concepts in the sciences will it be possible to acquire scientific knowledge despite extended leaching. There is work in this direction — most notably are Field's efforts (Field 1980) — but it remains an open question whether all relevant theories can be nominalized.

Suppose that relevant theories can be reformulated. Then a different issue might be raised: how does SSF with extended leaching compare to instrumentalist proposals such as Field's? In this context, we should not assume that we have to provide an epistemology that is compatible with Platonism. Are there any other reasons to prefer SSF to instrumentalism?

The answer is yes. SSF potentially does better in that Field's proposal requires the possibility of providing nominalistic reformulations of scientific theories, *and* the conservativeness of relevant mathematical theories over scientific theories. SSF — with or without extended leaching — just requires the former. Thus, SSF can avoid certain objections that arise in conjunction with Field's notion of conservativeness.

For example, Shapiro (1983) argues that Field should not use a semantic notion of conservativeness because it rests on mathematical concepts which cannot be presupposed in a nominalist setting, whereas endorsing a syntactical notion of conservativeness is not available to Field for technical reasons. Field (1985) accepts Shapiro's technical point and responds by construing a modal notion of semantic consequence. I cannot assess here whether Field's response is a good one. But the need to give such a response can be avoided in the first place, by accepting SSF instead of Field's position.

6.2.6 Indispensability

SSF with extended leaching requires that mathematical theories are dispensable in the sense that relevant scientific theories can be reformulated in a nominalistically acceptable way. One might think that this produces an irresolvable tension. For, so the thought goes, the possibility of success of SSF — whether with extended leaching or not — requires the *indispensability* of mathematical concepts and reasoning. What if mathematical reasoning was dispensable? Would we not have grossly overestimated the importance of mathematical theories, and does this not undermine the special role that SSF assigned to them? In particular, so the thought goes, dispensability would cast doubt on the **(Presupposition of Application)** principle: if there is a way of doing science without mathematics, the *truth of* relevant mathematical theories will not be important enough for the scientific enterprise, and thus loses its presuppositional role. If mathematics is dispensable, we might obtain other entitlements, e.g. an entitlement for the *consistency* of mathematical theories, but SSF was meant to be a Platonist theory (or at least compatible with Platonism). Would the ispensability not heavily favour instrumentalism in the end?

Note, however, that although there might be way of doing science without mathematics, there would still be a way of doing science with mathematics (Platonistically construed). And mathematical basic principles will still be presuppositions of these specific projects. Of course, one would then still have to argue for the relevance of these specific projects.

A related response that might also provide the resources to argue for the relevance of doing science with mathematics can be extracted from the literature on the indispensability argument. A mathematical theory T can be indispensable to science in at least two different ways:

- T can be indispensable in the sense that there are no nominalistic reformulations of all scientific theories endorsing T which have the same empirical consequences.
- T can be indispensable in that we need to make use of T in order to provide simple and unified (mathematical) explanations of empirical phenomena.

Baker (2005) argues that indispensability in the first sense does not entail indispensability in the second sense: there are cases of genuinely mathematical explanation where it is hard to see how we can provide a similarly good and unifying explanation without making use of mathematical concepts and reasoning (and without reference to mathematical structures). If this is right, then — modulo weaseling strategies[135] and other complications — indispensability in the first sense does not undermine the **(Presupposition of Application)** principle. For example, we might possess an entitlement for the *truth* of **(HP)** on the basis of its being a presupposition of projects providing simple and unified explanations in biology.

Note that this response is compatible with Bangu's worry (Bangu 2008) that the explananda of genuinely mathematical explanations presuppose the existence of mathematical entities and thus beg the question against the nominalist. But the worse this is for the friend of the indispensability argument, the better it is for the friend of SSF: for the existence of such explanations shows just how much mathematics is *actually* presupposed in vital projects of providing simple and unified explanations in areas different from mathematics. Also note, that if the indispensability argument succeeded, then SSF would not be needed: mathematics would not have to be an entitlement, because the regress clause would be violated: there would be a way of establishing the truth of mathematical theories without relying on something in no more secure a prior epistemic standing.

6.2.7 Meeting the desiderata

How do the sketched positions do with respect to the desiderata from 1.2? Clearly, SSF is compatible with **(Arithmetical Platonism)**. In fact, SSF is meant to be a Platonist position. The entitlements are entitlements for the *truth* of mathematical theories, qua theories about abstract objects. There is no reason, and no need for modesty here.[136]

The motivation for **(Field's Constraint)** is undermined. If mathematical theories are entitlements, they will be warranted non-evidentially. **(Field's Constraint)** is best understood as a constraint on evidential warrant. If we concede that we do not possess proper evidence for the truth of mathematical theories, there will be no need to explain

[135]See (Melia 2000).

[136]It is an interesting question whether accounting for a Platonist position requires invoking Wright's *entitlements of substance*, for Wright (2004*b*) thinks that entitlements of cognitive project cannot serve ontological purposes. This complication cannot be discussed in this thesis.

how our mathematical beliefs are reliably formed.

The **(Arithmetical Knowlege)** desideratum, however, is violated (at least for any notion of knowledge requiring the possession of evidence). This is the downside of any proposal endorsing entitlements for mathematical basic principles. However, I have argued that we can bite this bullet, because a violation of this principle does not preclude us from using mathematics to acquire (claimable) knowledge elsewhere. We might be able to bite the bullet, even if extended leaching occurs and mixed mathematical statements have to remain entitlements.

As a consequence, the **(Apriority Constraint)** and **(Arithmetical Foundationalism)** are violated as well. However, they are only violated because they are claims about *knowledge*. SSF still accounts for mathematical theories and axioms being *warranted* apriori.

The **(Applicability Constraint)** is met head-on. SSF identifies a central role of mathematical theories in that they are applicable in the sciences and in everyday reasoning. This role is supported by regarding abstraction principles as the proper foundations of relevant theories. Firstly, because of their biconditional form, they are suitable for extracting rules. Secondly, if Frege and the abstractionists are right, abstraction principles such as **(HP)** will meet **(Frege's Constraint)**, and thus make the applications of the theories based on them immediate.

Finally, since **(Frege's Constraint)** is met in relevant cases, there is a good chance that the proposal vindicates at least **(Weak Hermeneutic Reconstruction)**, i.e. that the terms of the theories based on the relevant abstraction principles refer to the mathematical objects we all know and love. This is because the neo-Fregean argument for this claim can be repeated for SSF.[137]

6.3 Intermediate conclusion

In the course of this chapter, I carved out several positions arising from a direct application of our epistemological framework to the mathematical case. We can order these positions by their concessiveness, i.e. by the ratio of entitled mathematical truths to internalistically knowable mathematical truths. All these positions can be considered as fallback positions,

[137]For a discussion of the relevance of **(Frege's Constraint)** to **(Weak Hermeneutic Reconstruction)**, see 2.3.5.

in case the less concessive positions fail. In the arithmetical case, these positions are:

- Wright's proposal (assuming Wright's response to leaching works after all): the rules capturing both sides of **(HP)** are entitlements, but the second-order Peano axioms can be claimed to be known on the basis of a modified version of Frege's proof. In fact, with the exception of cases of transmission-failure, all of Frege Arithmetic can be claimed to be known. Transmission-failure affects the epistemic status of **(HP)** itself, and possibly its instances. If it affects instances of **(HP)**, the proposal yields some revisionary sceptical consequences.

- SSF: in case Wright's trick to avoid leaching fails, all of arithmetic is a mere entitlement. This is a revisionary consequence. However, it is not devastating because entitled rules extracted from second-order arithmetic can be used to extend internalistic knowledge in other areas without qualifications.

- SSF with "extended leaching": maybe there is extended leaching to the effect that the weak epistemic status of pure arithmetic negatively affects the epistemic status of mixed arithmetical statements. In this case, all statements containing arithmetical terms are at best entitlements. But the strategy of SSF still delivers knowledge claims for purely empirical consequences of the conjunction of our non-mathematical theories and arithmetic (assuming that we can vindicate knowledge claims for the non-mathematical parts of these theories).

Unfortunately, all these options fall short of some of the aims set out in 1.2. Do we have to accept these consequences? No: for everything that has been said rests on the assumption that mathematical basic principles can be (or have to be) regarded as entitlements. In the next chapter, I will argue that they aren't entitlements, because they can be warranted evidentially without epistemic regress, and condition (ii) of entitlement of cognitive project is not met. In particular, I will reject the **(Logicality of Inference in Mathematics)** principle.

Note that the reason why I reject the proposals presented in this chapter is not that they are concessive, or that they yield revisionary sceptical consequences. The reason is that they get the structure of justification wrong. If the reason was that they yield revisionary sceptical consequences, they could not be used as fallback positions.

7 Knowledge by meta-linguistic stipulation

In the last chapter, I examined positions that rest on the claim that abstraction principles are entitlements of cognitive project. Such positions assume that the conditions for entitlement of cognitive project are met for relevant abstraction principles. In particular: that we cannot justify certain abstraction principles without falling into epistemic regress.

I now reject this assumption. In 7.1, I argue that meta-linguistic stipulations successfully fix meaning given that certain preconditions are in place, and that one can acquire internalistic apriori knowledge of the content of these stipulations by a line of reasoning that makes use of a non-logical inferential step, in case one possesses warrants for the obtaining of these preconditions. This provides a route to justify **(HP)** and the Peano axioms on something in a more secure a prior epistemic standing: knowledge of our own meaning-fixing commands.

In 7.2, I examine the relevant preconditions in detail, and argue that they are warranted, some of them by means of entitlement of cognitive project. In the course of these investigations, I show that my proposal sheds new light on the three big objections to neo-Fregeanism discussed in 2.4. Together with the first part of this chapter, these investigations also close the epistemological gaps of the neo-Fregean proposal I carved out in 2.3.4.

I then argue that my proposal generalizes. In 7.3, I argue that my proposal can be extended to the logical case, and sketch how it can be extended to implicit definition in general. In 7.4, I outline how my proposal vindicates the good standing of something close to the notion of epistemic analyticity.

7.1 Abstractionism reconsidered

7.1.1 The basic idea

Suppose a non-defective[138] epistemic agent mastering second-order logic — to keep up the tradition, call him *Hero* — sits in his armchair and decides to introduce a new term-forming type-lowering operator "$\#F$" ("the number of the Fs"). To do this, he consciously and explicitly stipulates that "#" is to be assigned a meaning such that the Hume Rules are *sound* (i.e. necessarily truth-preserving)[139]:

(Hume Rules)

$$(HP_{\rightarrow})\frac{\#(F)=\#(G)}{\exists R(Bij(R,F,G))} \quad (HP_{\leftarrow})\frac{\exists R(Bij(R,F,G))}{\#(F)=\#(G)}$$

This leads to:

(Cognitive Success) By virtue of making the stipulation, Hero comes to understand the new operator "#".

And, more importantly:

(Semantic Success) By virtue of making the stipulation, a meaning[140] is assigned to "#" such that the **(Hume Rules)** are sound.

Moreover, Hero knows what he has just done. So he is able to reflect on his stipulation as follows: I stipulated that "#" is to be assigned a meaning such that the **(Hume Rules)** are sound. So — Psst! Unless something went wrong, but I can assume that nothing went wrong — $(HP_{\rightarrow})$ and $(HP_{\leftarrow})$ are sound.

I contend that, by virtue of this simple inference, Hero can acquire internalistic apriori knowledge of the soundness of the **(Hume Rules)**:

[138] A *non-defective epistemic agent* is an agent with properly working cognitive faculties, who is competent in English and possesses the relevant concepts.

[139] There is a meta-language reading and an object-language reading of soundness. In what follows, I always mean the former. The stipulation is a *meta-linguistic* stipulation: it is about terms and inferential patterns. However, I do not think that a lot hangs on this distinction when it comes to the *conclusion* of the inference I defend in this chapter. I think that it could also be construed as an inference to a object-language version of the claim. If a subject understands the rules in question, the object-language version can always be inferred from the meta-language version (and vice versa). Ebert (2005*b*) and Jenkins (2008*a*) argue that inferring object language soundness claims from meta-language soundness claims are problematic in the current context. I disagree, but I cannot discuss this issue in this section. However, I will briefly come back to it below (see 7.4.6).

[140] I come back to how meaning should be understood here in section 7.2.

(Epistemic Success) The described inference transmits internalistic apriori knowledge. Hero can acquire such knowledge of its conclusion — the soundness of the **(Hume Rules)** — by virtue of inferring it from instances of apriori self-knowledge — that he stipulated that "#" is to be assigned a meaning such that the **(Hume Rules)** are sound — given he possesses warrants for the claim that nothing went wrong with his stipulation.

The restrictive clause at the end captures the Psst-part of the intuitive idea. Hero's acquisition of internalistic knowledge of the soundness of the **(Hume Rules)** will effect an unconditional (and rational) acceptance of the **(Hume Rules)** on the side of Hero.

In the course of this chapter, it will become apparent that the structure of justification is analogous to the structure of justification in the following case which we encountered in 4.2.1: Hero looks around and observes his two hands. On the basis of this experience, he acquires internalistic knowledge of the fact that he has, right now, an experience as of his two hands. He then reasons as follows: I have, right now, an experience as of my two hands. So — Psst! Unless something went wrong, e.g. if I was a brain-in-a-vat, but I can assume that nothing went wrong — I have two hands. By virtue of this inference, Hero can acquire internalistic knowledge of the fact that he has two hands.

This requires much more argument. In this section, I provide an argument for the existence of the inference underlying **(Epistemic Success)**, and some details about its workings. In 7.2, I argue that Hero can warrantedly assume that nothing went wrong.

7.1.2 An argument for the basic idea

My proposal rests on a metasemantic, and an epistemic principle. The metasemantic principle, which implies both **(Cognitive Success)** and **(Semantic Success)**, is this:

(Metasemantic Inferentialism) By virtue of stipulating that certain logico-mathematical terms are to be assigned a meaning[141] such that their introduction and elimination rules are sound, we can (i) come to understand these terms, and (ii) bring it about that they get assigned a meaning such that their introduction and elimination rules are sound.

The epistemic principle is this:

[141] Again, I wish to remain neutral as to how meaning should be understood here. I come back to this question in section 7.2.

(Implicit Definition Inference) There is a primitive type of inference allowing non-defective epistemic agents to acquire full internalistic apriori warrants for believing that our explicit meaning-fixing stipulations are in good standing, on the basis of internalistic knowledge of what these meaning-fixing stipulations are, in case we also possess antecedent warrants for certain conditions ensuring that nothing went wrong with the stipulation in question. First and foremost: that no scenarios obtain that undermine the claim that the inference is ***truth-preserving***[142].

Since the rules corresponding to both directions of abstraction principles can be conceived of as introduction and elimination rules of the respective abstraction operator, these two principles yield the following principle:

(Abstractionist Inference) There is a primitive type of inference allowing non-defective epistemic agents to acquire full internalistic apriori warrants for believing that rules corresponding to both directions of good abstraction principles are necessarily truth-preserving, on the basis of internalistic knowledge of our explicit meaning-fixing stipulations of these rules, in case we also possess antecedent warrants for certain conditions ensuring that nothing went wrong with the stipulation in question. First and foremost: that no scenarios obtain that would undermine the inference's truth-preservation.[143]

It implies **(Epistemic Success)**. Since the stipulation of the **(Hume Rules)** is semantically successful, the **(Hume Rules)** are sound. Thus, the full internalistic warrant for their soundness acquired by the **(Abstractionist Inference)** amounts to internalistic knowledge.

Presumably, the semantic part of my proposal — **(Metasemantic Inferentialism)** — is less contentious than its epistemic part. For the time being, I assume its truth without a defense, and come back to it in the next section. Note, however, that the principle is relatively weak: it is the claim that meaning of new operators *can* be fixed by explicit meaning-fixing stipulations of non-defective agents, and how an understanding of new terms *can* be generated. It is not a claim about how the meaning of *English* terms is

[142]That is: the material conditional encoding the inference is true.

[143]I focus on rules here. However, the proposal can be extended to sentential stipulations. Hero could also stipulate the (necessary) truth of (**HP**). I come back to this in 7.4.1.

actually fixed, or how *English* terms *actually* become understood.[144]

In any case, the epistemic part of my proposal is rather unorthodox. The claim is that there is a primitive type of inference that does the described job, and this is meant to imply that there is only *one* inferential step. Two questions arise immediately: firstly, why should there be a one-step inference from facts about actual stipulations of inferential patterns to the soundness of these patterns? Secondly, if there is such an inference, how exactly are we to conceive of its epistemic workings?

As to the first question, there are at least three considerations motivating the claim that there is a belief-forming method of the kind described in the **(Implicit Definition Inference)** principle:

- We have seen above that an important part of our cognitive lives consists in the successful pursuit of *cognitive projects*, i.e. projects of finding out about the world — ideally by acquiring knowledge of it. The good standing of certain concepts is a precondition of the successful execution of any cognitive project. And the good standing of our concepts rests on the good standing of our meaning-fixing devices. For example: the soundness of meaning-fixing rules. The good standing of these devices will be an entitlement, if there is no route to justifying them without engaging in epistemic regress. However, we might regard establishing the good standing of our meaning-fixing devices as a cognitive project of its own: a project in which we aim at acquiring knowledge of their good standing. After all, it is the aim of any reasonable epistemology to account for as much knowledge as possible. Because such a project would be very basic, we can expect that it has its own *primitive* belief-forming methods. The **(Implicit Definition Inference)** is a good candidate. Below, I argue that our epistemological framework predicts that every attempt to reduce this belief-forming method to other, more basic inferences — such as conceiving of it as an enthymeme — fails. So, if there is a project of acquiring internalistic knowledge of the good standing of our meaning-fixing devices, it must include a basic belief-forming method along the lines of the **(Implicit Definition Inference)**.

[144] This limitation makes the principle much harder to attack. On the other hand, it generates the worry that the proposal only applies to artificially introduced operators, which would limit its epistemic potential considerably. In the end, so the thought goes, it might turn out that we can only establish the soundness of the **(Hume Rules)** for an artificial operator — "$\#F$" — and not for the *English* operator "the number of the Fs", or so the thought goes. I briefly return to this worry in 7.1.8.

- We can find out which basic belief-forming methods we can take as basic by examining (i) how certain beliefs are canonically formed, and (ii) which of the belief-forming methods underlying canonical belief-formation cannot be reduced to a combination of others. As to (i), when we make an explicit stipulation to fix meaning, other things being equal, we come to believe the soundness, validity, or truth of the stipulation[145] without making any use of additional reasoning. For example, mathematicians write down a definition on the board, and after that, in most cases, the truth of the definition is believed. So there must be a belief-forming method which is based on something grounded in our stipulations, and which leads to beliefs in the good-standing of these stipulations. As I said, I will show that this belief-forming method cannot be reduced to other, more basic belief-forming methods. It must work along the lines of the **(Implicit Definition Inference)**.

One might think that what has been said could also be said in favour of postulating primitive rational insights: a faculty of rational intuition. However, there is a further general point to be made that allows us to respond to this worry:

- At some point, every epistemologist has to make choices as to what he or she regards as basic belief-forming methods. These choices depend in part (!) on cost-benefit considerations. Call a set of basic belief-forming methods an *epistemic framework*. Other things being equal, we should prefer an epistemic framework accounting for as much internalistic knowledge as possible, without sacrificing something of comparable philosophical worth. In particular: if postulating the **(Implicit Definition Inference)** is a means to explain the possibility of claiming knowledge of certain important propositions, and the framework does not have any unacceptable philosophical consequences, then we should postulate the existence of such an inference.

The argument rests on a meta-epistemological principle we already encountered in 5.1.3, namely the:

(Greater Meta-Epistemological Evil Principle) Other things being equal, we

[145]Note that I also call the stipulated pattern a stipulation. So there is some ambiguity here between stipulation *as an act* and stipulation *as a statement or inferential pattern* (corresponding to a similar ambiguity of the term "definition"). However, it is always easy to determine which of the two senses of stipulation I use, so I do not introduce artificial terms to distinguish the two.

should prefer an epistemological framework accounting for as much internalistic knowledge as possible, without sacrificing something of comparable philosophical worth.

The principle applies to the postulated structure of justification in particular. It enables us to rule out the postulation of a faculty of rational intuition: the considerations against rational insight in 1.3.1 show that postulating a faculty of rational intuition means sacrificing something of comparable philosophical worth. However, it allows us to postulate the **(Implicit Definition Inference)**. Among other things, it is not as unclear as the rational intuition proposal.[146]

Assuming that the proposed basic belief-forming method exists, the **(Implicit Definition Inference)** principle can be obtained if two principles hold. Firstly:

> **(Default Entitlement)** Presuppositions of basic belief-forming methods of fundamental cognitive projects, which cannot be warranted evidentially without falling into epistemic regress, are warranted by default.

Presuppositions are to be understood here in the way I defined the term in 4.2. They are propositions that need to be in the informational context of a subject — i.e. the subject needs to possess antecedent propositional warrants for them — in order for the respective belief-forming method to deliver evidential warrants for its conclusion. The second principle is:

> **(Anti-Leaching)** Basic belief-forming methods of such projects, whose presuppositions are warranted by default, generate internalistic knowledge, if no relevant sceptical scenario obtains.

We obtain the **(Implicit Definition Inference)** principle as follows: there is a basic belief-forming method allowing us to infer the good standing of our stipulations. We can assume here that the presuppositions of this inference are propositions to the effect that

[146]Note that a full defense of my proposal requires a comparison with other approaches on the market place. In particular, the reader might have the worry that it will not suffice to set aside the naïve rational intuition proposal without considering more sophisticated rationalist theories such as Bealer's (2000) and Peacocke's (2000). This is correct. Although I agree with Jenkins as to some of the objections against these theories (Jenkins 2008*b*, 2.5), I cannot discuss these theories in this thesis and I will simply set them aside. I am indebted to Carrie Jenkins and Andrew McGonigal for making me aware of this gap in my argument.

scenarios undermining the truth-preservation of the inference do not obtain.[147] The non-obtaining of these scenarios cannot be warranted evidentially without falling into epistemic regress. Thus, they are warranted by default **(Default Entitlement)**. And because of **(Anti-Leaching)**, the belief-forming method in question generates internalistic knowledge, if no relevant sceptical scenario obtains.

My epistemological framework meets both principles. Entitlement of cognitive project can play the role of warrants by default demanded by **(Default Entitlement)**, and **(Justification Generation)** is an **(Anti-Leaching)**-principle for entitlement. What remains to be established is that the proposed inference is the best way to model the canonical belief-forming process in question. This is the only way to make sure that it is subject to the **(Greater Meta-Epistemological Evil Principle)**. In particular, we need to argue that the current account does better than other accounts of knowledge by stipulation, such as Boghossian's. I argue for this claim by simultaneously clarifying the **(Abstractionist Inference)** by means of examples, and arguing for the made choices regarding the structure of justification.

7.1.3 The Abstractionist Inference

Consider again our non-defective epistemic agent — Hero — who already masters second-order logic, and who sincerely makes the following stipulation:

> **(Meaning-Fixing Stipulation)** "#" is to be assigned a meaning such that the **(Hume Rules)** are necessarily truth-preserving.[148]

Making a stipulation can be conceived of as a command of Hero to himself, which immediately brings it about that a pattern of use is fixed for "#", and Hero becomes disposed to reason in accordance with this pattern of use. By virtue of this process, Hero comes to understand "#".

However, making such a command does not have to effect an actual uncondition acceptance of the rules. Consider the case in which Hero stipulates that "ϵ" is to be assigned

[147] I discuss some potential exceptions in 7.2.

[148] Hero might also stipulate that "#" is to be assigned a meaning such that the Hume Rules are truth-preserving. Considerations of eligibility suggest that the assigned referent will be the same. In both cases the stipulation assigns the function from concepts to objects that maps concepts to their cardinalities. However, building it into the stipulation that the rules are to be necessarily truth-preserving increases the immediate epistemic payoff.

a meaning that renders the rules belonging to both directions of BLV sound:

(BLV Rules)

$$(\epsilon\text{-I})\ \frac{\forall xy(Fx\leftrightarrow Gx)}{\epsilon(F)=\epsilon(G)} \quad (\epsilon\text{-}E)\ \frac{\epsilon(F)=\epsilon(G)}{\forall xy(Fx\leftrightarrow Gx)}$$

Hero might be willing to seriously make this stipulation in order to understand "ϵ", or to obtain a concept of extension, but not be willing to use these rules, because he knows that BLV is inconsistent. This knowledge can override the generated disposition to reason in accordance with these rules.[149] Full acceptance is not ensured by having made a stipulation.

This case shows that I need a notion of stipulation such that one can seriously make a stipulation to the effect that a certain meaning should be assigned without it being (fully) successful, and even if one knows that it cannot be (fully) successful. Since I conceive of the stipulation as a command, this claim is not particularly problematic.[150] Ceteris paribus, that one has sincerely made a command does not depend on (the possibility of) its success, and one can sincerely make a command even if one knows that it cannot be succesful. A general can command his troops to advance, knowing that they won't.

As an example which is even more like the meaning-fixing case, consider a world in which there are several demigods, which have the power to stipulate or command that the world is to be in a certain way, and that these commands are successful unless God intervenes because he does not like the world to be as the command says. Clearly, a demigod can sincerely command that the moon is to consist of green cheese, although God intervenes and the command is not successful. And clearly, a demigod can make the command although he knows that God does not like the moon to be made of green cheese. The demigod might just be interested in provoking an argument with God.

The premises of the **(Abstractionist Inference)** are propositions to the effect that a meaning-fixing command has been sincerely made. For example:

(Premise) I sincerely made the **(Meaning-Fixing Stipulation)**.

[149] I say a little bit more about this in 7.2.1.

[150] It would be problematic to hold that sincerely making a meaning-fixing stipulation comes with the *intention* to fix meaning in all cases. If one knows meaning cannot be fixed, one might not be able to have this intention. On my view, an intention to fix meaning (reference) is not required. Of course, it is required that there is an intention to make the command itself. And some weaker intentions might be required in addition, such as the intention to come to understand the term on the basis of the stipulation.

I call such premises *stipulation facts*. My proposal is that, from the stipulation facts, the good standing of meaning-fixing devices — e.g. the soundness of the **(Hume Rules)** — can be inferred in a single step.

It better turns out that **(Premise)** can be known internalistically, and apriori. For otherwise, Hero cannot acquire internalistic apriori knowledge of the conclusion of the inference. I think that Hero can acquire internalistic apriori knowledge of **(Premise)**: everyone who sincerely made a meaning-fixing command knows apriori that he made this command sincerely, and has access to the fact that he knows that. Why? It is an instance of what has been called *maker's knowledge* — it is an instance of self-knowledge of one's own intentional actions — the kind of knowledge for which Anscombe (1957) argues. Ceteris paribus, if Hero knows that he intended to make the meaning-fixing command, he knows that he made the command. And this knowledge is apriori. It is not based on perceptual experience.[151][152][153]

The idea underlying the **(Abstractionist Inference)** is that it is truth-preserving in relevant good cases, because in these cases making the meaning-fixing command brings about **(Semantic Success)**. We need to have a closer look at the metasemantic part of the proposal. It involves two claims: (i) that Hero comes to understand the undefined operator and (ii) that meaning is fixed appropriately. Applied to our example, the claim is that, by virtue of making the command, "#" becomes understood and gets assigned a meaning such that the **(Hume Rules)** are sound.

To establish it, we just have to choose a suitable metasemantics. As to (i), I contend that sincerely making the meaning-fixing command brings it about that a pattern of use is fixed for "#" and accepted by Hero, i.e. that Hero becomes disposed to reason in

[151]Although this knowledge is apriori, it is knowledge of a contingent proposition. I do not think that this generates any problem for the proposal.

[152]Alternatively, one might describe the process of coming to possess internalistic knowledge of **(Premise)** as follows. Hero's command comes with a certain phenomenology of understanding the command and obeying it, and this phenomenology serves as a source for forming the belief that **(Premise)** holds. On this basis, **(Premise)** can be known internalistially and apriori by introspection, or so the thought goes. Of course, our notion of apriority needs to be wide enough to allow for knowledge by introspection to count as apriori knowledge.

[153]I thus endorse a relatively wide notion of apriority. The reader might worry that this wider notion of apriority allows for apriori knowledge of substantial external facts on the basis of knowledge of one's own mental states, if one also assumes semantic externalism. Consider, for example, Putnam's famous argument that we cannot be BIVs because we can think of water and we can only think of water if we are not BIVs (Putnam 1981, chapter 1). I reject such arguments. It is my view, however, that one can possess apriori *warrants* for (note: not apriori *justification* for or apriori *knowledge* of) substantial claims about the external world. For example, one can possess such a warrant for the claim that the BIV hypothesis does not obtain, because it is an entitlement.

accordance with the rules. Assuming a metasemantics according to which understanding consists in accepting a pattern of use, "#" can become understood by virtue of sincerely making the command. Note that the command fixes the logical form of the stipulation. Consider the **(Hume Rules)**. That "#" is a term-forming operator is settled, for this is part of the *content* of the meaning-fixing command.

As to (ii), we need a metasemantics allowing for meaning to be fixed by the stipulation of introduction and elimination rules. For now, we can assume that there is a plausible metasemantics delivering the desired result in the case of the **(Hume Rules)**. However, there are numerous examples where meaning-fixing commands of introduction and elimination rules fail. Certain preconditions have to be in place in order for the **(Abstractionist Inference)** to be truth-preserving.

We can focus on the bad companions of **(HP)** discussed in 2.4.2. The problematic abstraction principles, taken as pairs of rules, will be examples for bad companions of the **(Hume Rules)**. As an example, consider again the **(BLV Rules)**. Assuming unrestricted comprehension, these rules cannot be sound, because they lead to Russell's Paradox.[154] This suggest consistency as a minimal precondition.

There is much more to say. However, let us assume for now that we found the appropriate preconditions (and the relevant abstraction principles turn out to meet them). We then still face an epistemological problem: if not all stipulations can fix meaning in accordance with the stipulation, then how can Hero responsibly infer that a certain stipulation is in good standing, just on the basis of the claim that he made it? After all, the stipulation could be a bad case.

My suggestion, which is implicit in the **(Abstractionist Inference)**, is that we apply the epistemological framework carved out in Part II, and draw parallels to other cases of belief-forming methods that can go wrong in some way or other.

I regard the situation as being analoguous to the case of Hero forming the belief that he has two hands on the basis of having the experience as of his two hands. The existence of sceptical scenarios shows that this belief-forming method can go wrong. However, I argued in Part II that all this shows is that whether the experience can count as evidence depends on Hero's prior information to the effect that sceptical scenarios do not obtain. If Hero

[154] Just replace the MP steps in accordance with instances of BLV by the respective BLV-rule steps.

possesses such information, Hero can acquire a full internalistic warrant for there being a table by virtue of executing this belief-forming method, because he can rationally regard the experience as evidence for the external-world belief.

Here is how this bears on the mathematical case: bad companions show that the fact that a certain stipulation has been made cannot already be rationally regarded as evidence for the claim that the respective meaning-fixing devices are in good standing. Whether it can be regarded as evidence depends on the subject's prior information to the effect that the stipulation is not a bad case, which includes that the stipulation in question is not a bad companion. However, in the context of such prior information, knowledge that the respective stipulation has been made is evidence for the good standing of its meaning-fixing devices — e.g. the soundness of rules corresponding to both directions of abstraction principles.

In chapter 4, I examined how we can handle information-dependent evidence in general. The upshot is this: we endorse a conservative interpretation of the belief-forming method in question — i.e. the claim that we need to possess antecedent warrants to accept certain presuppositions ensuring the non-obtaining of undermining (sceptical) scenarios — and argue that the presuppositions of these belief-forming methods are warranted. If the belief-forming method is sufficiently basic, we will need to endorse non-evidential warrants for the presuppositions: entitlements of cognitive project. However, because of the **(Justification Generation)** principle, this does not preclude us from using these belief-forming methods to acquire internalistic knowledge, for it tells us that belief-forming methods all of whose presuppositions are entitlements ceteris paribus generate or transmit internalistic knowledge.

I conceive of the **(Abstractionist Inference)** as a basic belief-forming method which is treated conservatively. It has characteristic presuppositions, ensuring its good standing in particular cases. Suppose that all presuppositions are met. Then meaning is fixed in the appropriate way, the inference is truth-preserving, and the conclusion is true. If the presuppositions are also warranted — maybe by means of entitlement of cognitive project — the subject can acquire internalistic knowledge of its conclusion — the soundness of the relevant rules — because of the **(Justification Generation)** principle. This is precisely the suggestion expressed by the **(Abstractionist Inference)** principle.

What are the relevant presuppositions in the case of the **(Hume Rules)**? One might simply say that the presupposition is one of the following:

- Nothing went wrong with this stipulation (that would undermine the inference's truth-preservation).
- This inference is truth-preserving.
- If I make this stipulation, then the stipulated rules will be sound.

All these entail that the stipulation in question is not a bad companion (e.g. that it is consistent). However, although it is very plausible that some of these are presuppositions, claiming that these are the only presuppositions, and leaving it at that, would not be very illuminating.

Firstly, we need to know what exactly constitutes the conduciveness of the environment, and we need examples for bad cases. Only then can we examine how exactly the presuppositions are warranted in relevant cases. Consider the perceptual case again: after noting that it is a presupposition of perceptual belief-forming methods that the cognitive environment is conducive, Wright provides concrete examples of what might go wrong.

Secondly, as we shall see below, knowing what the presuppositions are is important to evaluate the epistemological consequences of the proposal, because of transmission-failure considerations.

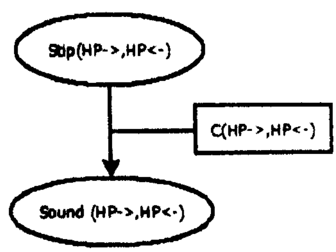
Thirdly, simple presupposition such as those above won't do for more sophisticated subjects. If Hero has a firm grasp of logic and knows about the inconsistency of the **(BLV Rules)**, he cannot simply assume that nothing went wrong in order to regard his stipulation as evidence for the soundness of the **(Hume Rules)**. Hero needs to assume that the **(Hume Rules)** do not fall short of the same difficulty as the **(BLV Rules)**. That is: he needs to assume that the **(Hume Rules)** are consistent. Consistency is a presupposition of the inference in this case.[155]

[155] An alternative is to say that it is indeed only one of the simple propositions above that features as the presupposition, even for more sophisticated subjects, and to say that specific conditions such as consistency are relevant not because they are proper presuppositions, but because having a reason to doubt them is a sufficient reason to believe that the simple presupposition is untrue. Thus, having a reason to doubt that specific conditions such as consistency are met would defeat our entitlement for the simple presupposition. Something similar could be said about perceptual belief-forming methods, but I cannot discuss this alternative picture in this thesis.

So we need to say more. We need to extract appropriate presuppositions, and argue that they can all be warranted in relevant cases. Given the regress argument of the last chapter, we should expect that some presuppositions — such as consistency — will be warranted by means of entitlement. In the remainder of this subsection, I will simply assume that the consistency of the **(Hume Rules)** is among the presuppositions of a stipulation of the **(Hume Rules)**, that it is an entitlement, and that all other presuppositions of a stipulation of the **(Hume Rules)** — whatever they are — are likewise warranted. I thus obtain a picture I can work with in the remainder of this section. However, I will discuss the presuppositions in much more detail in 7.2.

7.1.4 Defending the proposed structure of justification (1)

For a stipulation of a pair of abstractionist rules R_{Intro} and R_{Elim}, I denote the stipulation facts by "$\mathit{Stip}\,(R_{Intro}, R_{Elim})$", and the conditions that feature as presuppositions by "$C[R_{Intro}, R_{Elim}]$". The soundness of a rule R will be denoted by "$\mathit{Sound}\,(R)$". Now consider the case of the **(Hume Rules)**. We can display the relevant instance of the **(Abstractionist Inference)** as follows:



(Abstractionist Inference - Diagram)

The suggestion is that, given that the relevant presuppositions are warranted, Hero can acquire internalistic knowledge of $\mathit{Sound}\,(HP_{\rightarrow})$ and $\mathit{Sound}\,(HP_{\leftarrow})$ by virtue of going through the following simple line of reasoning:

(1) $\mathit{Stip}\,(HP_{\rightarrow}, HP_{\leftarrow})$(= **(Premise)**)

(2) $\mathit{Sound}\,(HP_{\rightarrow})$ and $\mathit{Sound}\,(HP_{\leftarrow})$ (1), **(Abstractionist Inference)**

And this leads to the following objection to my proposal: why should Hero be able to infer (2) ***directly*** from (1)? Is it not a much better explanation that he goes through a metasemantic line of reasoning directly mirroring the claim that the stipulation is semantically successful, citing **(Metasemantic Inferentialism)**? Is my envisaged **(Abstractionist Inference)** not just an *enthymeme* for a sustained reflection on stipulations along the following lines?

(1')	$Stip(HP_{\rightarrow}, HP_{\leftarrow}) \wedge C(HP_{\rightarrow}, HP_{\leftarrow})$	
(2')	$Stip(HP_{\rightarrow}, HP_{\leftarrow}) \wedge C(HP_{\rightarrow}, HP_{\leftarrow}) \rightarrow$ $Sound(HP_{\rightarrow}) \wedge Sound(HP_{\leftarrow})$	By **(Metasemantic Inferentialism)**
(3')	$Sound(HP_{\rightarrow}) \wedge Sound(HP_{\leftarrow})$	1', 2', MP

We can summarize this idea as follows:

(Metasemantic Reasoning Model) The canonical belief-forming method underlying the process of acquiring knowledge of the good standing of meaning-fixing devices on the basis of knowledge of meta-linguistic stipulations is best described by a deductive, metasemantic line of reasoning using **(Metasemantic Inferentialism)** as a premise.

The problem with this proposal is that it is subject to leaching, and thus fails the **(Greater Meta-Epistemological Evil Principle)**. Suppose it turns out that some presuppositions will only be warranted by means of entitlement. We already saw that this is very plausible, and I will argue for the claim in full detail below. In this case, the second conjunct of the argument's premise (1') will at best be an (inferential) entitlement.[156] But if an argument's premise is only an (inferential) entitlement, then — by the **(Limit Principle)** from section 4.5.1 — its conclusion cannot have an epistemic status above (inferential) entitlement. Thus, the soundness of the **(Hume Rules)** cannot be internalistically known by virtue of going through the argument.

The proposal would be just as sceptical as Wright's position (see 6.1.3). However, the aim of the current proposal is to vindicate the possibility of possessing internalistic knowledge of propositions such as (3'). If the canonical structure of justification is displayed by an explicit argument using the conditions for semantic success as a premise, we cannot achieve this aim.

[156]This result is obtained by the **(Limit Principle)** from section 4.5.1. Since some of the presuppositions are entitlements — by the **(Limit Principle)** — $C(HP_{\rightarrow}, HP_{\leftarrow})$ it will (at best) be an inferential entitlement because it is a conjunction one of whose conjuncts is a mere entitlement. For the same reasons, (1') will (at best) be an inferential entitlement since it has to be inferred from both conjuncts — $Stip(HP_{\rightarrow}, HP_{\leftarrow})$ and $C(HP_{\rightarrow}, HP_{\leftarrow})$.

This is just another instance of the leaching worry, and we already know a strategy to avoid it. In particular, although I accepted the **(Limit Principle)**, I rejected what I called the **(Strong Limit Principle)**: that entitled *presuppositions* preclude an inference from transmitting internalistic knowledge. Thus, if we can regard the preconditions for semantic success — and $C(HP_{\rightarrow}, HP_{\leftarrow})$ in particular — as presuppositions of the abstractionist's line of reasoning, as opposed to premises, we will avoid the leaching worry.

This is precisely my suggestion. The **(Abstractionist Inference)** principle construes the conditions for semantic success as presuppositions. It does not matter that some of them are entitlements. By **(Justification Generation)**, we can acquire internalistic knowledge of the conclusion of the inference from (1) to (2) by virtue of drawing the inference on the basis of internalistically known premises. As I mentioned above, I think the stipulation facts can be regarded as items of *self-knowledge* in the relevant cases, and self-knowledge is internalistic knowledge. Thus, we can acquire internalistic knowledge of (2).

At this point, the reader might still worry (i) that we cannot construe the structure of justification *as we please*, and (ii) that my proposal cannot be correct, because the proposed change in the structure of justification cannot account for internalistic knowledge, if the **(Metasemantic Reasoning Model)** cannot account for internalistic knowledge; in short: that my proposal *pulls the internalistic knowledge rabbit out of the entitlement hat.*

As to (ii), the worry amounts to the claim that **(Justification Generation)** cannot be correct. However, I already argued at length for the entitlement proposal and the **(Justification Generation)** principle. I explained how internalistic knowledge can be generated on the basis of belief-forming methods with entitled presuppositions, and I will explain how exactly it works in the case at hand in the next subsection. There are no magic tricks here at all.

As to (i), note first that I am engaging in a reconstructive project. And it is not clear whether the complaint really has force in this context. I am not claiming to track the *actual* structure of justification.

Be that as it may. If the reader feels there is the need for an argument, here is an argument. Because of the **(Greater Meta-Epistemological Evil Principle)**, we should postulate the structure of justification delivering as much internalistic knowledge as possible, without sacrificing something of comparable philosophical worth. Construing the **(Abstractionist Inference)** as I do looks as if it can deliver more internalistic knowledge than all the other proposals we have seen so far, such as the **(Metasemantic Reasoning Model)**. Since I am able to clearly explain how the belief-forming process works, and I am able to defend the proposal against objections, the **(Greater Meta-Epistemological Evil Principle)** applies, and delivers justification for the postulation of the described primitive inference.

7.1.5 The generation of internalistic justification

I complete the model by explicating how **(Justification Generation)** is to work in the case of the **(Abstractionist Inference)**. Consider the case of Hero. The basic thought is this: because Hero possesses warrants for all presuppositions — let us assume that all these warrants are entitlements — Hero possesses an entitlement for the claim that the stipulation facts constitute evidence for the soundness of the **(Hume Rules)**, and thus that he can extend knowledge by virtue of the **(Abstractionist Inference)**. Since Hero can also claim that he possesses knowledge of the stipulation facts — they are internalistically known — he can acquire an inferential entitlement for the claim that he possesses knowledge of the conclusion.

In 4.5.2, I presented two models for **(Justification Generation)**. Both models can be applied here, but I just consider the first for the sake of simplicity. Let "*Sound*" stand for "$Sound(HP_{\rightarrow}) \wedge Sound(HP_{\leftarrow})$", let "$K(Sound)$" stand for "I know that *Sound*", and let "$B_{IDI}(p)$" stand for "I formed the belief that p on the basis of drawing the appropriate instance of the implicit definition inference". According to our first model, Hero can acquire an (inferential) entitlement for $K(Sound)$ by going through the following argument:

1	$K(Stip(HP_{\rightarrow}, HP_{\leftarrow}))$	Accessibility of self-knowledge
2	$B_{IDI}(Sound)$	Self-knowledge
3	$C(HP_{\rightarrow}, HP_{\leftarrow})$	Hero rationally takes this for granted
4	$B_{IDI}(Sound) \wedge C(HP_{\rightarrow}, HP_{\leftarrow}) \rightarrow (K(Stip(HP_{\rightarrow}, HP_{\leftarrow})) \rightarrow K(Sound))$	Metasemantics and epistemology.
5	$K(Stip(HP_{\rightarrow}, HP_{\leftarrow})) \rightarrow K(Sound)$	2,3,4, MP
6	$K(Sound)$	4,1, MP

Two notes are in order. Firstly, Hero's entitlement for the claim that the stipulation facts constitute evidence for the conclusion (line 4) will rest on the availability of an argument similar to the **(Metasemantic Reasoning Model)**. The crucial difference is that it is applied at the *second-level*. Second-level considerations are merely entitled because they rest on entitled premises, but the first-level argument — the argument for *Sound* — does not.

Secondly, the argument above involves citing conditions like consistency, which is to be understood as a formal concept. Does the current proposal not exclude subjects that do not possess the relevant concepts? Does it not, so to speak, face an *exclusive club problem*?

I do not think this is so. For a less sophisticated Hero might begin his second-order reflection with lines 1 and 4. It is not always required that one realizes just *how* the entitlement for the conditional in line 4 is canonically grounded. A less sophisticated Hero — or Hero who has no reason to consider conditions like consistency because he never realized that there are bad companions — will be directly entitled to take it for granted that nothing went wrong, or that the **(Abstractionist Inference)** is truth-preserving. Thus, he will also be able to acquire an inferential entitlement for the claim that the **(Abstractionist Inference)** extends knowledge.

What has been said enables us to reconsider the abstractionist position. For the **(Abstractionist Inference)** precisifies the epistemic workings of the neo-Fregean **(Implicit Definition Thesis)**. Let me retell the neo-Fregean Hero story using the **(Abstractionist Inference)**.

7.1.6 A refined Hero story

The story again begins with Hero, a non-defective epistemic agent who grasps second-order logic, and can use it to extend internalistic knowledge. Hero stipulates the **(Hume Rules)** to fix the meaning of "#". He does so by sincerely making the **(Meaning-Fixing Stipulation)** — that "#" is to be assigned a meaning such that the **(Hume Rules)** are sound.

By virtue of this, Hero comes to understand "#", and the meaning that is assigned to it renders the **(Hume Rules)** sound. I argue below that the assigned meaning is the number operator that maps concepts to their cardinalities.

Moreover, Hero can acquire internalistic knowledge of the fact that he has just made the appropriate meta-linguistic stipulation. On the basis of this knowledge, Hero infers that the **(Hume Rules)** are sound, by drawing the appropriate instance of the **(Abstractionist Inference)**. Note that Hero acquired knowledge by an inferential, but non-deductive route. The **(Logicality of Inference in Mathematics)** assumption which underlied the mere entitlement proposals from chapter 6, needs to be rejected.

Because Hero possesses entitlements for the presupposition of this inference — in particular, he does not possess a sufficient reason to believe that the **(Hume Rules)** are not consistent, etc. — internalistic justification (and knowledge) is transmitted in accordance with the **(Justification Generation)** principle. Hero thus acquires internalistic knowledge of the soundness of the **(Hume Rules)**.

Furthermore, either the soundness of the **(Hume Rules)** is the only presupposition of reasoning in accordance with them, or Hero can acquire an internalistic warrant for the presupposition on the basis of his internalistic knowledge of soundness.[157] Thus, we can assume that Hero possesses a warrant for whatever the presupposition of reasoning in accordance with instances of the **(Hume Rules)** is. Also, Hero now unconditionally (and rationally) accepts the **(Hume Rules)**, and is willing to make inferences in accordance with these rules. Thus, Hero can use the **(Hume Rules)** to extend internalistic knowledge, because of the **(Justification Generation)** principle.

Hero can now go through Frege's proof: after having made the necessary explicit definitions, he can use the modified, rule-based version of Frege's Proof to infer versions of the second-order Peano axioms, just using second-order logic and the **(Hume Rules)**. We assumed that Hero can use second-order logic to extend internalistic knowledge. So all the used inferential steps transmit internalistic knowledge. Hero thus comes to possess internalistic knowledge of the second-order Dedekind-Peano axioms. From this, Hero can make further inferences and acquire knowledge of further, specific properties of the natural

[157]Note that there is a question here of whether the presupposition of the rules is a meta-language, or an object-language version of the claim that they are sound (or truth-preserving), and what kind of claim the **(Abstractionist Inference)** delivers. Presumably, the presuppositions are not couched in meta-linguistic terms. Thus, the current argument either requires that the **(Abstractionist Inference)** directly delivers an object language version of the soundness claim, or Hero needs to disquote, i.e. infer an object language version of the soundness claim — a version of the claim in which "#" is used, and not only mentioned — from the meta-language claim. I do not think that such a disquotational step is problematic, since Hero already fully understands "#". However, note that such steps have been problematized by Ebert (2005*b*) and Jenkins (2008*b*).

numbers. Moreover, as sketched in the last section, Hero will be able to apply arithmetical rules in other contexts. This time, however, Hero not only possesses entitlements for the soundness of the arithmetical rules, but internalistic knowledge.

7.1.7 Transmission-failure and the justification of consistency

Hero can acquire a lot of internalistic knowledge on this basis. However, there are limits. They come with transmission-failure.

The general lesson of transmission-failure considerations can be put like this: entitled presuppositions remain entitlements, and can never be claimed to be known. Thus, whatever the entitled presuppositions of the **(Abstractionist Inference)** are, these propositions will remain entitlements, and can never be claimed to be known.

For example: since the consistency of **(HP)** is an entitled presupposition of the **(Abstractionist Inference)** for **(HP)**, Hero will not be able to acquire internalistic knowledge of its consistency. The following argument displays transmission-failure, because the step from (1) to (2) presupposes its conclusion:

(1)	$Stip\,(HP_{\rightarrow}, HP_{\leftarrow})$(= **(Premise)** above)	
(2)	$K\,(Sound)$	(1), *Abstractionist Inference*
(3)	If $K\,(Sound)$, then HP is consistent.	Sound rules do not generate contraditions
(4)	HP is consistent.	(2), (3), MP

One might think that this is a revisionary sceptical consequence, because knowledge of the consistency of arithmetic is a paradigmatic claim to knowledge. I am not sure whether it is. But it is certainly worth examining whether we can somehow avoid this consequence.

Can we avoid it by recourse to stronger theories? Suppose Hero not only stipulates the **(Hume Rules)**, but that he also stipulates abstraction principles which yield theories with a stronger consistency strength than Frege Arithmetic (FA). Suppose there is a pair of rules corresponding to a consistent abstraction principle for sets, which delivers a theory as strong as ZF. Call the corresponding abstraction principle *Magic Law V*, call the corresponding rules the *Magic Law V Rules*, and call the system *Frege Set Theory* (*FST*).[158] Prima facie, we should be able to tell a Hero story about this system. If this is correct, then Hero will be able to prove a statement expressing the consistency of FA and hence the consistency of the **(Hume Rules)** — in FST. Thus, Hero will be able to acquire internalistic knowledge of its consistency after all, or so the thought goes.[159]

But this does not really help with avoiding sceptical consequences, given that consistency claims for mathematical theories are paradigmatic claims to knowledge. For now the consistency of FST will be an entitlement. Of course, one might hope to find another, even stronger theory that proves the consistency of this theory in turn. But this does not avoid the problem, for then the stronger theory's consistency will be an entitlement.

Moreover, on closer consideration it is not at all clear that the availability of a proof in FST establishing a claim that can be taken to express the consistency of FA shows that the consistency of FA cannot be an entitlement. First of all, what exactly can we show in FST? We could use Gödel's technique to construct a statement $Con(FA)$ that is provable in FST and of which we can show, *in some meta-theory*, that $Con(FA)$ holds if and only if FA is consistent. But we can only show this as long as we assume, in the meta-theory, that FST is consistent. We will have to use the consistency of FST as a premise in our proof of the consistency of FA and HP. And if this premise is, as we assumed, only an entitlement, we cannot obtain more than an (inferential) entitlement for the conclusion of the meta-theoretic argument, i.e. the claim that FA and HP are consistent.[160]

In other words: a proof of the consistency of FA in FST relies on presuppositions of the same kind as the consistency of FA, and in no more secure a prior epistemic standing than the consistency of FA. The consistency of FA remains an entitlement.

Consistency claims for abstractionist theories are entitlements, and have to remain entitlements, if they are presuppositions of the **(Abstractionist Inference)**. This is a bullet I have to bite. It is a consequence of the epistemic framework presented in Part II.

[158] I am grateful to Filippo Ferrari and Robert Williams for making me aware of this line of reasoning, and the potential complications this brings with it.

[159] Robert Williams has pointed out to me that this might lead to a potentially devastating objection to the proposal. Suppose we can find stronger and stronger theories, each of which proves the satisfiability of the weaker ones. One might think that in this case *no* satisfiability claim is an entitlement, because there is always an argument available which does not rest on the same presuppositions as the weaker theory. But how does the abstractionist project get off the ground then? It seems we are caught in a regress, because we always have to *prove* consistency by using a stronger theory. The following argument shows how to rebut this objection.

[160] This is because of the **(Limit Principle)** from 4.5.1.

7.1.8 Hermeneutic abstractionism

How does the generated knowledge relate to ordinary arithmetical knowledge? Can the described route enable Hero to possess internalistic knowledge of the properties of the numbers we all know and love? In other words: can my proposal provide for hermeneutic reconstruction?

Fortunately, at least in the arithmetical case, we can build on the arguments provided by the neo-Fregeans. In particular, we can build on the idea that **(HP)** meets **(Frege's Constraint)** — that the primary applications of cardinal numbers are rendered immediate by **(HP)** — and that this implies that Hero will come to refer to and acquire knowledge of the (properties of the) numbers we all know and love, i.e. the numbers every sufficiently sophisticated user of ordinary number talk refers to: 1, 2, 3, and so on.

According to one line of reasoning (and this is the line of reasoning I want to endorse), this is because the fact that **(HP)** meets **(Frege's Constraint)**, and the definitions of the other terms of FA, entails that the terms of FA displays the same pattern of use as ordinary arithmetical terms (see 2.3.5). Thus, by the neo-Fregean **(Reference Supervenes on Use)** principle, the terms of our reconstruction and our ordinary arithmetical terms have the same meanings (referents). The reconstructed knowledge really is arithmetical knowledge: the subject matter is the same. Moreover, we have established that the mathematicians, the scientists, and sufficiently competent users of mathematics in everyday discourse have true arithmetical beliefs. This is because the truth values of relevant ordinary arithmetical propositions are the same as the truth values of corresponding theorems of our reconstructed theory. In other words, the proposal promises to vindicate at least **(Weak Hermeneutic Reconstruction)**. All this of course depends on whether some objections, discussed in 2.3.5, can be met.

Ideally, Hero (and hence the reader of this thesis) can come to know the hermeneutic claim apriori. If this is the case, then Hero will not only be able to acquire claimable knowledge of arithmetical facts, but also be able to claim, on the basis of his reconstruction, that he possesses knowledge of the ordinary arithmetical statements he made before he stipulated **(HP)** (and that these statements are true).

Can Hero come to know apriori that the terms of his reconstructed theory have the same reference as the terms of ordinary number talk? I think it is plausible that he can, for the neo-Fregean argument for **(Weak Hermeneutic Reconstruction)** is available to Hero apriori, given that it apriori what the logical form of ordinary number talk is, and that Hero possesses relevant mathematical and philosophical concepts and skills. In 1.2, I assumed that a relevant transparency principle holds, and in 2.3.5, I argued that it might hold even if it is an epistemic possibility that further linguistic research reveals that ordinary number terms are not really singular terms. Of course, the assumption merits further research.

Be that as it may. As I said in 2.3.5, that Hero's knowledge meets the demand of **(Weak Hermeneutic Reconstruction)**, and hence that my account meets the **(Arithmetical Knowledge)** constraint, would still be true even if the argument for it was aposteriori. So this result would not be devastating.

Moreover, if the neo-Fregean strategy for **(Strong Hermeneutic Reconstruction)** can be vindicated, then it is probably also available to my proposal. However, the question of how to establish strong hermeneutic claims, and whether such claims are available apriori, is wide open, and I want to remain neutral on this issue.

7.1.9 Meeting the constraints

I complete my exposition of the new abstractionist position by examining how it fares with respect to the constraints I imposed in 1.2.

The first constraint is **(Arithmetical Platonism)**. It should be clear through my discussion of the hermeneutic aspect of my proposal that my aim is to account for knowledge of mathematical theories, Platonistically construed. The idea is that we can acquire knowledge of **(HP)** with its apparent logical form taken at face value. In particular: once we come to possess internalistic knowledge of the existence of certain bijections — even trivial ones as e.g. $\exists R\,(Bij\,(R, [x = x], [x = x]))$ — we can use $(HP_{\leftarrow})$ to acquire internalistic knowledge of a number identity — e.g. $\#_x[x = x] = \#_x[x = x]$ — and acquire internalistic knowledge of the existence of a certain number in turn — e.g. $\exists y\,(y = \#_x[x = x])$. This claim is meant to be interpreted as a claim about what there is. It is meant to carry a genuine ontological commitment. This commitment is one to abstract objects. I have not explicitly argued for this claim yet, but I think that it is not hard to provide such an argument. **(HP)** can only be necessarily true if the objects it is about necessarily exist.[161]

This might raise rejectionist complaints. I think that my proposal enables us to reconsider this objection in a new light: namely as a challenge to identify suitable presuppositions for stipulations with heavyweight ontological impact. I discuss the objection and my response below.

My proposal provides the resources to meet **(Field's Constraint)**. For we can extract an account of how our practise of stipulation and corresponding belief-formation can be reliable.[162]

The upshot is that in most cases of meta-linguistic stipulation, nothing goes wrong, and that bad cases are such that responsible epistemic agents realize, sooner or later, that they are bad, which effects giving up their false beliefs and also raises the standards for further stipulations. Consider **(BLV)**, for example. Once the inconsistency is brought to Hero's attention, he looses his warrant for the presupposition that the stipulation is consistent and for the claim that nothing went wrong. Thus, he will give up the belief that **(BLV)** is true. Moreover, once such cases are known, the epistemic standards for further stipulations are raised. From this moment onwards, it needs to be excluded that new stipulations fall short of the same difficulty, for otherwise the agent has sufficient reason to doubt that the relevant presuppositions are met, and they will not be entitlements.

The postulated structure of justification ensures that stipulations are made responsibly, in a way which is conducive to reliability. Only if a subject does not have sufficient reasons to doubt that nothing goes wrong — i.e. that none of the relevant bad scenarios obtains — does a subject acquire an internalistic warrant for the conclusion of relevant instances of the **(Implicit Definition Inference)**. This ensures that full internalistic warrant is generated, first and foremost, in good cases.

Moreover, in order to be fully responsible, a subject needs to check that he or she possesses entitlements for the claim that nothing goes wrong, i.e. that he or she possesses entitlements for all relevant presuppositions. This will uncover risky stipulations and effect an examination to the effect whether the stipulation can really be made. Specific pitfalls of identified bad cases will be avoided by fully responsible epistemic agents. I have dis-

[161] See also the considerations in 7.2.3. I am also sympathetic to Rosen's arguments in (Rosen 1993).

[162] I am indebted to Carrie Jenkins and Andrew McGonigal for a very interesting and helpful discussion of this issue.

cussed above what the essential presuppositions of the **(Abstractionist Inference)** are. Someone who makes sure that he or she possesses entitlements for these presuppositions will avoid all known pitfalls. And, for all we know, avoiding these pitfalls means that the stipulation in question is successful.

I have already discussed how I intend to account for **(Weak Hermeneutic Reconstruction)**. Although the proposal endorsing **(Frege's Constraint)** merits further research, we have reason to believe that the account meets the **(Arithmetical Knowledge)** desideratum.

It also meets the **(Arithmetical Foundationalism)** constraint, and the **(Apriority Constraint)**. Knowledge of the stipulation facts will be apriori, because it is makers knowledge, and both the **(Abstractionist Inference)** and logical rules preserve apriority.

The **(Applicability Constraint)** is met in the same way as in the case of the proposals discussed in the last chapter. Because of **(Frege's Constraint)**, the applicability of arithmetic is built into the foundations of our reconstructed theory.

The epistemic payoff of the proposal is limited by cases of transmission-failure. I have to bite this bullet. Such cases are rare, and in any case the proposal does much better than the proposals discussed in the last section. However, it might still yield some moderate sceptical consequences. For example: if consistency claims about mathematical theories count as ordinary mathematical claims (and they probably do), then some ordinary mathematical truths cannot be claimed to be known.

The general idea underlying my proposal motivates reconsidering the **(Same Source)** desideratum. If basic mathematical rules can be justified by means of the **(Abstractionist Inference)**, why should we not say something similar about logical rules? Can we extend the proposal in such a way that we obtain an instance of the **(Implicit Definition Inference)** for logic? A positive answer would be most welcome for at least two reasons:

- We would avoid the revisionary sceptical consequence that the validity of basic logical laws is warranted only by means of entitlement. In fact, if the proposal can be extended to the logical case, validity claims could be internalistically known apriori, and inferentially so, on the basis of self-knowledge of our own meta-linguistic stipulations.

- Moreover, we would obtain the result that the **(Same Source)** constraint is met and thus vindicate genuine logicism. Mathematics and logic would be canonically based on the same belief-forming method: the **(Implicit Definition Inference)**.

I argue for a positive answer below. First, however, I need to examine in more detail the presuppositions of the **(Abstractionist Inference)**.

7.2 The presuppositions of the Abstractionist Inference

To complete the argument that has emerged in the last sections, it needs to be established (i) that all presuppositions of relevant instances of the **(Abstractionist Inference)** can be (internalistically) warranted, and (ii) that at least some of the warrants for these presuppositions are entitlements. Establishing (i) will enable us to appeal to the **(Justification Generation)** principle to establish that one can acquire internalistic knowledge by virtue of the **(Abstractionist Inference)**. Establishing (ii) is required as a premise in the motivation for endorsing the **(Abstractionist Inference)**: if all conditions could be justified without regress, our argument against the **(Metasemantic Reasoning Model)** would be undermined, because there would be no leaching problem.

This requires a detailed investigation of the question of what the presuppositions of the **(Abstractionist Inference)** are. Above, I worked with the example of consistency. But consistency cannot possibly be the only presupposition. We saw in 2.4.2 that there are consistent, but mutually inconsistent abstraction principles. This shows that not all consistent abstraction principles can be true. Thus, the **(Abstractionist Inference)** is not truth-preserving for some consistent abstraction principles. And everything that is relevant to ensure the truth-preservation of this inference is a candidate for a presupposition: if the inference is not truth-preserving, something must have gone wrong, and the presuppositions — by definition — are the conditions ensuring that nothing went wrong.

I am not interested in any presuppositions whatoever. The aim is to identify a collection of presuppositions of the **(Abstractionist Inference)** which delivers a complete and informative account of what can go wrong with stipulations of abstraction principles, such that someone who knows of all these presuppositions, and who has made sure that he possesses warrants for them, can be called a *maximally informed and responsible stipulator.*[163]

In particular, the identified collection of presuppositions should be maximal in the sense that its members entail all other propositions that could be identified as presuppositions. In particular, they should entail the ***trivial presuppositions*** such as the propositions that

[163] I am sympathetic to the idea that logico-mathematical sophistication *ups the ante*, i.e. that only unsophisticated stipulators can claim knowledge solely on the basis of being able to claim entitlements to simple presuppositions such as "Nothing went wrong". More sophisticated subjects need to be able to claim entitlements for more specific presuppositions (such as consistency). By making sure that we can claim entitlements for all the essential presuppositions, we establish that even the most sophisticated subjects can rationally claim knowledge on the basis of the **(Abstractionist Inference)**.

nothing goes wrong, or that the **(Abstractionist Inference)** is truth-preserving.

Call the collection of presuppositions meeting these conditions the *essential presuppositions* of the **(Abstractionist Inference)**.[164]

7.2.1 Three presuppositions of success

A satisfying account of what the essential presuppositions are can only be given in the light of an appropriate semantic and metasemantic background theory which tells us (a) what meaning is, and (b) how it can be brought about that meaning is fixed.[165]

Let us assume that we have available a suitable metasemantic theory. It will reveal what the *relevant* conditions are that need to be in place such that a stipulation of a certain abstraction principle brings it about that a suitable meaning is assigned to the relevant abstraction operator. Call these conditions the *presuppositions of metasemantic success*. Whatever these are, some of them will be essential presuppositions of the **(Abstractionist Inference)**.

Are these all the presuppositions? I'm afraid not! It turns out that there are also *presuppositions of cognitive success* and *epistemic presuppositions*. This is because there is more that can go wrong with drawing the **(Abstractionist Inference)** than that no meaning can be assigned that renders the stipulated rules sound: the stipulation might not be able to generate understanding, and the stipulation might be epistemically ineffective.[166] Let me explain.

Metasemantic success vs. cognitive success Remember that the stipulation not only has to bring it about that meaning is assigned, but also has to effect an *understanding* of the abstraction operator — and the relevant rules — on the part of the stipulating sub-

[164] I assume that there is only one collection of essential presuppositions. I do not have an argument for this claim.

[165] One way to draw the distinction between *semantics* and *metasemantics* is this: whereas a semantic theory for a language explains how the meaning of complex linguistic expressions of this language depends on the meaning of their constituents and the way they are put together, a metasemantic theory explains by virtue of what a semantic theory is selected for a given language, how linguistic expressions acquire meaning in the first place, what constitutes meaning, etc. Compare the case of ethics: ethics is the discipline of determining which actions are the good ones, using an ethical theory, and meta-ethics is the discipline in which one tries to explain how we should select ethical theories, how an action can be good or bad in the first place, what constitutes goodness, etc. Instead of metasemantics, one might also use the expressions "foundational semantics", "theory of meaning" or "the philosophy of linguistic representation" (Williams 2008, p. 603).

[166] These three kinds of presuppositions of course correspond to Ebert's three dimensions of achievement: effectiveness, success, and epistemic productiveness (see Ebert 2011).

ject. Otherwise, the abstraction operator could not be *used* by the subject in any project of extending knowledge, but at best be mentioned.[167] For example, without understanding "#", Hero could not come to understand Frege Arithmetic, and this is a necessary condition to acquire knowledge of its axioms and theorems. Note that the **(Metasemantic Inferentialism)** principle also involves both dimensions of success.

We should not conceive of the process of coming to understand the abstraction operator as a process of grasping *the meaning that is assigned*, if "meaning" is read in a certain way. Draw the Fregean distinction between sense and reference and consider the **(BLV Rules)**. It is very plausible that the extension operator can be understood by virtue of a stipulation of these rules,[168] but there is no referent to be assigned that renders the **(BLV Rules)** sound, because they are inconsistent. Thus, understanding the extension operator cannot consist in grasping the assigned referent. If we want to talk of meaning being assigned and grasped here, we must read "meaning" as "Fregean sense" (or "concept"), the thought being that a stipulation of the **(BLV Rules)** still assigns a sense (concept) to the extension operator, which is grasped.

This raises an interesting question. If we make the Fregean distinction, should we read the meaning-fixing command as being about sense *and* reference, or only about one of them? It might be interesting to explore different options, but I think that someone endorsing the Fregean distinction could just make something like the following stipulation:

> **(Meaning-Fixing Stipulation')** "#" is to be assigned a referent such that the **(Hume Rules)** are sound, and the sense determined by accepting the **(Hume Rules)** as the device to fix the meaning of "#".

In this thesis, I will not complicate the picture by talking about senses being assigned as well. In particular, I cannot discuss issues concerning concept acquisition and concept possession. I focus on the understanding question: the question how the number operator is *understood* by virtue of making the meaning-fixing command. I think there are at least two plausible answers, corresponding to two views of what understanding consists in: what Heck (2006) calls the *Use Theory* and the *Cognitive Conception*. These can be characterized, very roughly, as follows:

[167] For a related critique of arguments based on meta-linguistic stipulation, see (Ebert 2005*b*).

[168] For an extensive discussion and an argument that BLV provides at least some understanding, see (Ebert 2005*a*, pp. 125ff).

- The *Use Theory* (see Horwich 1998; Heck 2006): understanding consists in using a term appropriately, and this can be brought about by accepting a fundamental pattern of use for the term, i.e. by becoming disposed to reason in accordance with it. Applied to our case, the idea is that the **(Hume Rules)** express a *fundamental* pattern of use for "#". By virtue of making the **(Meaning-Fixing Stipulation)**, so the thought goes, one becomes disposed to reason in accordance with these rules, and this suffices to effect an understanding of "#" on the part of the subject.

- The *Cognitive Conception* (see e.g. Heck 2006): understanding a term t consists in tacitly having knowledge of, or at least beliefs about relevant semantic properties of t. One way to make this explicit is in terms of beliefs about truth conditions. Williams (2012) points out that cognitive conceptions come in different strengths. According to the weak version, one only needs to hold appropriate beliefs about the truth conditions of sentences containing the term. According to the strong version, one needs to hold beliefs about the reference of the term. According to the ultra-strong version, one needs to have a complete semantic theory about the term. Regardless of the particular version of the cognitive conception one might defend, the thought must be that sincerely making the **(Meaning-Fixing Stipulation)** brings it about that one holds the relevant beliefs.

I think my proposal is compatible with both theories of understanding. I endorse the Use Theory here, because it is simpler. Thus, my claim regarding cognitive success is that, given certain preconditions are in place, sincerely making the meaning-fixing command brings it about that one becomes disposed to reason in accordance with the rules which are stipulated to be valid, and that this suffices for understanding.

The reader might worry that I need to endorse a version of the Use Theory which renders it possible to understand meaning-fixing rules without being willing to actually use the rules. For I think that one can stipulate the **(BLV Rules)** to be valid, and come to understand them on this basis, although one knows that they are inconsistent. And someone knowing that these rules are inconsistent will not be willing to use them (at least not unconditionally). How can it be possible to understand "ϵ" although one does not actually endorse the rules, if understanding consists in being disposed to reason in accordance with these rules?

I think that one can understand "ϵ" by virtue of being disposed to reason in accordance with the **(BLV Rules)**, without these dispositions ever being active (or only active in some rare cases). Dispositions can be overridden in various ways. Such cases are common ground in the literature. In particular, dispositions can be *masked*, i.e. an existing disposition cannot be manifested because this is prevented by another disposition (for an overview, see Choi & Fara 2012). I contend that knowledge of the inconsistency of the **(BLV Rules)** brings with it other dispositions that mask the disposition to reason in accordance with these rules. There is more to say here, but I think this is a plausible response, and I will assume that it is a good response in what follows.[169]

One might want to make it explicit that the term is to be understood by grasping a pattern of use and prefer the following formulation of the **(Meaning-Fixing Stipulation)**:

> **(Meaning-Fixing Stipulation")** Let "#" get assigned a meaning such that the **(Hume Rules)** are sound, and "#" should be understood by virtue of accepting the **(Hume Rules)** as the fundamental pattern of use for "#".

At least in the mathematical case, I have nothing against such more sophisticated commands. But it is desirable to keep the stipulation as simple as possible. The official view thus is that the way understanding is generated is *implicit* in the stipulation, and that the stipulation invokes a grasp of a pattern of use without this being explicitly stipulated, because sincerely making the meaning-fixing command automatically effects an acceptance of the stipulated rules as determining a pattern of use.

Now, there are ways in which fixing a pattern of use can go wrong. And the non-obtaining of such scenarios may be called the *presuppositions of cognitive success*. Some of these will be essential presuppositions of the **(Abstractionist Inference)**.

The reader might have a worry at this point. The worry is that the presuppositions of cognitive success should not be regarded as presuppositions of the **(Abstractionist Inference)**. For one might think that all that matters is the truth-preservation of the **(Abstractionist Inference)**. If a suitable meaning is assigned, so the thought goes, the

[169] One issue here is that the mask would be intrinsic, i.e. other dispositions *of the agent* would mask the agent's dispositions. There is a debate of whether intrinsic masking is possible. For an argument for the possibility of intrinsic masks, see (Ashwell 2010). For an argument against, see (Handfield & Bird 2008).
Note that Williamson (2007, chapter 4) also argues extensively against the outlined idea. Unfortunately, I cannot pursue the issue any further here.

relevant instance of the **(Abstractionist Inference)** will be truth-preserving, no matter of whether the abstraction operator is understood or not.

However, the presuppositions of a belief-forming method are, by definition, the propositions which need to be in a subject's informational context in order for the inference to deliver an internalistic warrant for its target beliefs. And a lack of warrant for whatever the presuppositions of cognitive success are certainly undermines the subject's warrant to regard the stipulation facts as evidence for the soundness of the rules in question. This is because a lack of warrant for a precondition of cognitive success entails a lack of warrant for the claim that the stipulation assigns any meaning. After all, the term is meant to be introduced into the language of the stipulator, and from the point of view of the stipulating subject, if the term is not even understood, it does not have any meaning. Ergo: the presuppositions of cognitive success are presuppositions of the **(Abstractionist Inference)**.

Epistemic presuppositions The abstractionist might have to recognize a further type of presuppositions: *epistemic presuppositions*. Even if all presuppositions concerning the generation of understanding and the assignment of meanings (referents) are met (and entitled), drawing the **(Abstractionist Inference)** can still be bad from an epistemological point of view. For it might be such that a move from the stipulation facts to the conclusion of the **(Abstractionist Inference)** is epistemically irresponsible in such a way that one cannot warrantedly regard the stipulation facts as evidence for the soundness of the relevant rules by default. I have in mind some stipulations that Hale and Wright (2000) want to handle with their non-arrogance constraint. I come back to this below.

I now go through all three kinds of presuppositions in turn, and extract essential presuppositions. I argue that all essential presuppositions are warranted in the case of the **(Hume Rules)**, at least some of them by means of entitlement of cognitive project.

7.2.2 Presuppositions of cognitive success

We can find out what the presuppositions of cognitive success are by considering cases in which no pattern of use can be grasped. However, at first glance, all the stipulations we considered so far were good cases in this sense.

We can easily generate at least one bad case, namely a case in which some terms other than the abstraction operator are not already understood. For example:

(Nonsense) $\forall FG\,(\alpha\,(F) = \alpha\,(G) \leftrightarrow xzyjd\,(F,G))$

Such cases are not particularly interesting. One might think that in such cases the meaning-fixing command cannot be made in the first place, and the question about presuppositions does not arise.[170]

However, one might also want to cover such cases by invoking the presupposition that all terms except for the abstraction operator are already understood. Would this be an *essential* presupposition? No. For there is a more fundamental presupposition that entails it.

The more fundamental presupposition is that a version of Evan's *Generality Constraint* is met for the abstraction operator. The Generality Constraint can be put as follows (Hale & Wright 2000, p. 22; Evans 1982, pp. 100-105):

(Generality Constraint) In order to understand t, one needs to understand all relevant contexts $\phi[t]$, where the matrix $\phi[\_]$ is already understood.

The relevant presupposition can be put as follows:

(Generality Presupposition) Accepting the rules in question as a fundamental pattern of use for the abstraction operator O can bring it about that one understands all relevant contexts $\phi[O]$, where the matrix $\phi[\_]$ is already understood.

Two question arise. Firstly: do the **(Hume Rules)** meet this constraint? And, secondly: what kind of warrant can Hero possess for this presupposition?

The first question brings us back to classical objections to neo-Fregeanism. One might construe the Caesar problem as the worry that this presupposition is not met. One version of the argument goes as follows:

1. The term "Julius Caesar", the predicate "x is a planet", and "=" are already understood.

2. Thus, the matrix "_(x is a planet)=Julius Caesar" is already understood.

[170]It is my view that one can simultaneously fix the meaning of several terms, but note that the command is supposed to be one that fixes the meaning only of "α".

3. Thus, if a stipulation of the **(Hume Rules)** meets the **(Generality Presupposition)**, a stipulation of the **(Hume Rules)** will bring it about that one understands (C) "$\#_x$(x is a planet)=Julius Caesar".

4. However, a stipulation of the **(Hume Rules)** cannot bring it about that this statement is understood. For it only fixes a pattern of use for identity statements where both terms begin with the number operator.

5. Therefore, a stipulation of the **(Hume Rules)** does not meet the **(Generality Presupposition)**.

The question is, of course, whether the second last line of this argument should be accepted. Why can a stipulation of the **(Hume Rules)** not bring it about that statements like (C) are understood? Why cannot a use be fixed for mixed identity statements as well?

One might want to avail oneself of one of the easy responses to the Caesar problem sketched in 2.4.1. For example, one might think that Hero can determine that such statements should be rejected on independent grounds. Numbers exist necessarily, and persons do not. Therefore, so the thought goes, numbers and persons cannot possibly be identical. The idea then is, very roughly, that everyone who fully understands what persons are and what numbers are, will never introduce an identity statement involving a person and a number, always reject such a statement etc.

Another response is that statements like (C) involve category mistakes, and are meaningless for this reason. Thus, so the thought goes, it cannot be a requirement that (C) has to be understood, and it cannot be a statement relevant for the **(Generality Presupposition)**.

However, we saw in 2.4.1 that there are harder versions of the Caesar problem. What about identity statements involving different sorts of abstracts? The above responses are not readily available in these cases.

The abstractionist needs to account for what I called the **(Categorization of Abstracts)**, i.e. a principled and metaphysically motivated partition of different abstracts into categories. In 2.4.1, I argued that this will be a difficult task. If the task succeeds, however, then it seems as if the neo-Fregean standard response can be applied. All this requires further work, but here is the basic idea. If two kinds of abstracts belong to the same category, then there will be a shared criterion of identity, and a pattern of use will have been fixed. If two kinds of abstracts do not belong to the same category, there are two options. If cross-categorial identity statements always need to be decided negatively, relevant statements need to be rejected tout court, and a pattern of use will have been fixed. If cross-categorial identity statements are evidence-transcendent, then the current objection is not a specific objection against Hero's ability to grasp a pattern of use on the basis of a stipulation of the **(Hume Rules)**, but it is a worry that affects understanding in general.

I certainly cannot decide the issue here. Let us assume that there is a solution to this version of the Caesar problem, and that the presupposition is true in the relevant cases. Then how will the **(Generality Presupposition)** be warranted in relevant cases?

I focus on the **(Hume Rules)**. There are two options: either the **(Generality Presupposition)** can be justified without regress on the basis of something in a more secure a prior epistemic standing, or not.

As to the former option, one might think that it can be apriori justified on the basis of theoretical views on understanding and its preconditions. If it turns out that every such justification involves presuppositions in no more secure a prior epistemic standing — this would be an interesting result and merits further research — the **(Generality Presupposition)** will be an entitlement in the relevant cases, because the three conditions for entitlement of cognitive project are met:

- It is a Wright-presupposition of the project of finding out that our meaning-fixing devices are in good standing.
- Condition (i): by assumption, it cannot be justified (without relying on something in no more secure a prior epistemic standing).
- Condition (ii): there is no sufficient reason to believe that the **(Generality Presupposition)** is not met (assuming there is a solution to the Caesar problem).

Thus, given that there is a solution to the Caesar problem, the essential presupposition of cognitive success will be internalistically warranted. This completes my discussion of such presuppositions.[171]

[171] A complete discussion of presuppositions of cognitive success would have to involve a discussion of the notion of impredicativity. However, this is beyond the scope of this thesis.

7.2.3 Interpretationism

I now turn to the presuppositions of metasemantic success — the presuppositions ensuring that meaning (=reference)[172] can be fixed such that the stipulated rules turn out to be sound, or, in other words, the presuppositions ensuring the truth-preservation of instances of the **(Abstractionist Inference)**.

What these presuppositions are will depend on the metasemantic background picture. It will tell us: (i) what meanings are; (ii) under what conditions meanings are assigned; (iii) which meanings are assigned. A suitable metasemantics for implicit definitions is what Williams (2007, 2008) calls ***interpretationism***. According to interpretationism, complete semantic units such as sentences have semantic priority, and we can easily extend the picture to the rule case. Let me explain.

Classical interpretationism is motivated by the task of radical interpretation, i.e. the task of interpreting speakers of a remote language and their language from scratch. Lewis (1974) nicely explains the task: given all the physical facts **(P)** about the speaker K of a foreign language L and his environment, solve the following three unknowns (Lewis 1974, p.332):

(Ao) K's attitudes, beliefs and desires, as expressed in our language.

(Ak) K's attitudes, beliefs and desires, as expressed in K's language.

(M) The "meanings" of K's language.

Determining (M) includes determining the meanings of K's sentences, and determining how the meanings of these sentences are determined by the meanings of sub-sentential expressions of K's sentences.[173] Interpretationism is designed to solve the problem of how we can come to know (M) in particular, and rests on two claims, which correspond to two stages of an interpretation process.

The first claim is that there are observable facts about K and his environment **(P)** which can be used to determine the semantic values of sentences of K's language. There are many ways of spelling this out. Depending on one's choice of what kinds of things the semantic values of sentences are, one can make a choice as to how particular semantic

[172] Throughout this subsection, and until 7.2.6, "meaning" should be read as "reference".
[173] Thus, it also includes determining what the meanings of the sub-sentential expressions are.

values are to be assigned on the basis of **(P)**. Collecting these values is the first stage of the interpretation process.

The second claim is that the primary criterion of success for a theory of **(M)** — and the primary criterion of success for a semantic theory of L — is that it matches the semantic values collected for the sentences in the first stage of the interpretation process. This is not to say that the first and the second stage have to be carried out separately. This is just to say that the semantic values of sentences are fundamental: matching the semantic value of the sentences is the primary criterion of success.

This formulation leaves open that it is not the only criterion. In fact, fitting the sentential data cannot possibly be the only criterion, because of inscrutability of reference objections. Lewis (1984), for example, argues that the semantic theory is additionally constrained by criteria of naturalness or eligibility. I agree with Lewis.

Here are three well-known interpretationist theories:

- According to Davidson's programme (Davidson 1973), the semantic values of sentences are truth conditions, which can be extracted from observable utterances using a *Principle of Charity*, i.e. we proceed on the assumption that most of the subject's utterances are true in the context in which they are uttered.

- According to Lewis's programme (see e.g. Lewis 1974), the semantic values of sentences are functions from possible worlds to truth values. The meanings of subsentential expressions are chosen accordingly. For example, the semantic values of names will be construed as functions from possible worlds to objects. What emerges is "an account of language (...) which is truth conditional and intensional, couched in the framework of possible worlds" (Holton 2003, p. 2). The propositions can be assigned to sentences by a procedure which is related to Davidson's, but (much) more sophisticated. I cannot go into detail here.

- Global descriptivism is a very simple version of interpretationism, discussed by Lewis (1984) and Williams (2007). The idea is that the meaning of a range of terms can be fixed *by description*, whereas *fixing meaning by description* just means laying down a *total theory* containing all relevant term-introducing statements. The semantic values of sentences are truth values, and the correct semantic theory for the language of the theory is a semantic theory which renders the total theory true. If the theory can be expressed within the resources of first-order or second-order logic, the obvious choice for a semantic theory is the theory that can be extracted from a model of the theory. What are the term introducing statements? An example of where global descriptivism can be applied is when it comes to determining the meaning of some range of *theoretical terms* operating on top of a more fundamental language (e.g. the terms of folk psychology). We could extract their term introducing statements from an existing language by just collecting the *platitudes* containing the terms, i.e. those statements containing the terms which are accepted in every situation (Williams 2007, p.367).

I think that all three frameworks can be used for my purposes. However, global descriptivism is the simplest theory, and it suffices to render the basic idea clear enough, so I will endorse it here. The applicability of global descriptivism is obvious. The thought is that the correct semantic theory for "#" is specified by an interpretation under which the term-introducing statement for "#" comes out true, i.e. an interpretation under which it comes out true that the **(Hume Rules)** are necessarily truth-preserving.

Consider Hero, who has just made a meaning-fixing stipulation. We can assume here that Hero means the same with "*necessary truth-preservation*" as we do. Now add the claim that the **(Hume Rules)** are necessarily truth-preserving to Hero's current theory about the world. Any function from Hero's concepts to objects under which the **(Hume Rules)** come out as (necessarily) truth-preserving will be a candidate meaning for "#".

We need to endorse (something like Lewis's) notion of eligibility. For there are many models of the **(Hume Rules)**, even if it is fixed how many (and what) objects there are, and what concepts there are. There are many choices even for the categorical part of the theory. On the face of it, "$\#_x[x \neq x]$" might be interpreted as $\emptyset$ in one model, and as 0 in another (assuming $\emptyset \neq 0$).

I contend that there is only one most eligible interpretation of "#" in all relevant cases: the interpretation that lets natural number terms refer to the natural numbers. Why are the natural numbers assigned? One reason might be **(Frege's Constraint)**. The term-introducing princippes should be interpreted as being about the objects that most naturally fit their description, and in the case of the **(Hume Rules)** it is very plausible that these are the (natural) numbers we all know and love (see also 7.1.8). This claim requires much more argument, but I cannot do more here than just to assume that the notion of eligibility works this way.

There are other aspects of this proposal that require further work. It is my view that the meaning assigned to "#" will be a function mapping concepts to objects. However, there are problems with using such a semantic theory, at least when we endorse a set-theoretic notion of "function".

Firstly, there might be too many concepts for the domain of the function to be a set. Secondly, and more importantly, one might think that the semantics for "#" should be formulated without using mathematical terms (e.g. sets). One might think that this is required because the theory should be such that it can be warrantedly used by Hero before he has introduced any abstract objects by abstraction. These are very interesting and tricky issues. My aim here can just be to provide the outlines of one suitable metasemantics of the proposal, in such a way that we philosophers can extract the essential presuppositions of the **(Abstractionist Inference)**.

7.2.4 Mathematical presuppositions

The above picture of how meaning (reference) is to be assigned immediately generates presuppositions of metasemantic success. All relevant conditions on abstraction principles ruling out cases in which there is no suitable model-theoretic interpretation for the relevant abstraction operator — i.e. a suitable function from concepts to objects — will count as such presuppositions.

Consistency becomes a presupposition on this picture. Inconsistent abstraction principles do not have any models. Moreover, we see that there must be presuppositions other than consistency: for Bad Company objections show that not all consistent abstraction principles can be interpreted as being true (together). We need to find criteria that single out a maximal collection of abstraction principles that has a model. I call these criteria:

> **(Mathematical Presuppositions)** Mathematical presuppositions are the conditions ensuring that an abstraction principle is part of the relevant maximal collection of abstraction principles that has a model.

My strategy to find these presuppositions is to use the criteria that neo-Fregeans discovered when carrying out **(Fine's Programme)** — the mathematical project I sketched in in 2.4.2. This project exactly corresponds to the project of finding suitable mathematical presuppositions. For it just is the project of finding a maximal collection of abstraction principles that can be jointly true. I briefly go through the most relevant conditions, and argue that they are warranted by means of entitlement of cognitive project.

Consistency The consistency of the **(Hume Rules)** will be a mathematical presupposition of the relevant instance of the **(Abstractionist Inference)**. It will be warranted by means of entitlement of cognitive project, because all conditions are met (for the conditions, see 4.4):

- It is a Wright-presupposition of the project of finding out that our meaning-fixing devices are in good standing.
- Condition (i): it cannot be justified (without relying on something in no more secure a prior epistemic standing). We saw in 6.1.2 that we cannot justify the consistency of the **(Hume Rules)** without presupposing a stronger theory, just as we cannot justify the consistency of **(HP)** without presupposing a stronger theory.
- Condition (ii): we do not have sufficient reason to believe that the **(Hume Rules)** are inconsistent, just as we do not have sufficient reason to believe that **(HP)** is inconsistent. To the contrary: for all we know, these rules are consistent.

Conservativeness Although consistency is a presupposition, it is not an *essential* presupposition in the sense defined above. For it is entailed by another mathematical presupposition: *conservativeness*. Conservativeness is a mathematical presupposition because there are consistent, but mutually inconsistent abstraction principles. It is a technical notion that can be put in a precise way (see 2.4.2). The basic idea is that:

(Conservativeness) An abstraction principle is conservative just in case it does not yield new consequences about any old domain (i.e. any domain of objects we might recognize before we introduced the abstraction principle).

How is it warranted? Again by means of entitlement of cognitive project:

- It is a Wright-presupposition of the project of finding out that our meaning-fixing devices are in good standing.

- Condition (i): it cannot be justified (without relying on something in no more secure a prior epistemic standing). Conservativeness entails consistency. Since the regress clause for entitlement of cognitive project is met in the case of consistency, it will also be met in the case of conservativeness.

- Condition (ii): we do not have sufficient reason to believe that the **(Hume Rules)** are not conservative, just as we do not have sufficient reason to believe that **(HP)** is not conservative. To the contrary: for all we know, these rules are conservative.

Irenicity However, conservativeness shares the same fate with consistency. Although conservativeness is a presupposition, it is not an *essential* presupposition in the sense defined above. For it is entailed by yet another mathematical presupposition: *irenicity*.

To repeat: the notion of irenicity has been suggested by Weir (2003), after he observed that there are conservative but mutually unsatisfiable abstraction principles. For my current purposes, the following informal characterization suffices: an abstraction principle is *irenic* just in case it is consistent with every conservative abstraction principle. Using set theory, it can be shown that **(HP)** is irenic (Weir 2003, pp. 32f).

Although the discussion about mathematical presuppositions goes on, and there are known complications with irenicity (see 2.4.2), my tentative suggestion is that irenicity is the only essential mathematical presupposition of the **(Abstractionist Inference)**.

How is this presupposition warranted? Again by means of entitlement of cognitive project. By now, the reader will be able to see how the argument goes, so I omit it here.

7.2.5 Ontological presuppositions

Many will be unsatisfied with the presuppositions proposed so far.[174] For example, so the thought goes, I have ignored ontological questions. What if there are only finitely many objects? In this case, there won't be any interpretation for "#" under which the **(Hume Rules)** come out as truth-preserving. This motivates postulating a further type of presuppositions, namely:

[174] This corresponds to a worry I discussed in 2.4.2, namely that carrying out **(Fine's Programme)** does not yet amount to carrying out what I call the **(Neo-Fregean Programme)**.

(Ontological Presuppositions) *Ontological presuppositions* are the conditions that guarantee that nothing goes wrong from an ontological point of view.

The following is an obvious candidate for such a presupposition:

(Presupposition of Existence) There is a domain of objects that can serve as the domain for a model of the rules in question.

I now argue (i) that it yields undesired consequences for the abstractionist to acknowledge a **(Presupposition of Existence)**, and (ii) that there are strategies to avoid such presuppositions. The idea is that the ontological issue is not really an issue, once the mathematical presuppositions are met, and thus does not raise further presuppositions.

Transmission-failure Suppose that the **(Presupposition of Existence)** is a presupposition of the **(Abstractionist Inference)**. Then it is plausible that the following condition is a presupposition as well, for it is entailed by the **(Presupposition of Existence)**, and the entailment seems direct enough to make the case eligible for presupposition expansion (for a discussion of this phenomenon, see 6.1.6):

(Infinity) There are infinitely many objects.

Thus, Hero needs to possess an antecedent warrant for it in order to acquire internalistic knowledge by virtue of carrying out the **(Abstractionist Inference)**. What kind of warrant could Hero possess for this claim? It is clear that this warrant could only be an entitlement of cognitive project.[175] Part of the neo-Fregean idea was that we can prove the existence of infinitely many objects on the basis of the **(Hume Rules)**. And if a justification on the basis of abstraction principles (or corresponding rules) is the only way to justify the existence of infinitely many objects — and it is hard to see what other route there could be for a neo-Fregean without giving up too much of the basic idea — then we cannot justify **(Infinity)** without presupposing it. Condition (ii) for entitlement of cognitive project is met.

[175]Wright (2004*b*, VIII) claims that entitlements of cognitive project are not suitable for ontological purposes. For reasons I cannot go into here, I do not think that this restriction applies to the case at hand, and I shall simply assume here that we can use the notion of entitlement of cognitive project.

Moreover, if **(Infinity)** was a presupposition of the **(Abstractionist Inference)**, it would be a Wright-presupposition of the cognitive project of ensuring that our meaning-fixing devices are in good standing. Thus, in this case, whether or not **(Infinity)** is an entitlement depends on whether there is sufficient reason to believe that there are only finitely many objects. I have already discussed this condition in the last chapter, and argued that — disregarding dubious considerations of stubborn nominalists — we do not have such reasons. So it is plausible that **(Infinity)** would indeed turn out to be an entitlement of cognitive project.

If **(Infinity)** was a presupposition, this would not preclude us from using the **(Abstractionist Inference)** to acquire internalistic knowledge of the soundness of the **(Hume Rules)**. However, if **(Infinity)** was an entitled presupposition, we could not possibly acquire internalistic knowledge of **(Infinity)** because any argument for it would fail to transmit warrant due to the **(Information Dependence Template)** (see 5.2.2). And it was one of the most precious aims of the neo-Fregeans to establish that we can *learn of* the existence of an infinite domain by virtue of Frege's proof. This aim would be impossible to achieve if there was transmission-failure.

There are two additional reasons of why we should avoid an **(Infinity)** presupposition. Firstly, Frege Arithmetic would be a theory that ultimately has a mixed epistemological status. Some of its members could be internalistically known — e.g. the Peano axioms — and some of its members could not be internalistically known — the second-order version of **(Infinity)**. This would be an odd result.[176]

Secondly, that we cannot claim to know that there are infinitely many objects is a revisionary sceptical consequence. Mathematicians and logicians often claim to know that there are enough objects to satisfy axiom systems that only have infinite models.

Existence What can be said in favour of a **(Presupposition of Existence)**? I believe that the considerations in favour of ontological presuppositions exactly match the considerations that lead to *epistemic rejectionism*, the major objection to neo-Fregeanism I discussed in 2.4.3. Rejectionism rests on the worry that the stipulation of an abstraction principle (or corresponding rules) might go wrong for the reason that there might not be

[176]This result is reminiscent of the odd results generated by endorsing merely entitled **(Hume Rules)**, which I used to criticize Wright's position in 6.1.6.

enough objects to render the stipulations good. For example: that there might not be enough objects to render the **(Hume Rules)** sound. So whether or not we can avoid existence as a presupposition depends on whether the rejectionist objection to neo-Fregeanism can be avoided.

To repeat: I think that the neo-Fregean can avail himself of various strategies to undermine rejectionism. Some of these strategies require further research, and some of them might be anti-realist in spirit. But I think there are enough options on the table to draw an optimistic conclusion.

The current context enables us to assess rejectionism from another angle. It gives us a criterion to decide whether we need to invoke ontological presuppositions or not. The criterion is whether there is a specific way the process of fixing meaning can fail from an ontological point of view — given all other presuppositions are met. So we can avoid ontological presuppositions by showing that there is no such way. Here are three such strategies:

- Plain maximalism: we might endorse *meta-ontological maximalism*, i.e. that everything that can exists, does exist (see Eklund 2006). From the point of view of the maximalist, **(Existence)** is trivially met as soon as the mathematical presuppositions are met. Hale and Wright reject this strategy, because (i) they think they do not need it and (ii) they think it has undesired metaphysical impliciations (Hale & Wright 2009, §4). I regard it as an attractive fallback position.

- Priority of meaning-fixing: this strategy corresponds to the "ought implies can" strategy sketched in 2.4.3. Wright (1999, pp. 311f) argues that we cannot even intelligibly ask the question about the existence of a realm of infinitely many numbers before we possess the concept of number, and that possessing the concept of number already requires accepting the truth of (an un-conditionalized version of) **(HP)**. This thought can be applied to my proposal as follows: (i) the relevant ontological presupposition would not simply be **(Existence)**, but a claim to the effect that *numbers* exist, and (ii) this claim cannot be a presupposition, because it must be possible for a thinker to be able to grasp all presuppositions of the relevant instance of the **(Abstraction Inference)** before having made the stipulation of the **(Hume Rules)**.

- Content-recarving: Hale and Wright (Hale 1997; Hale & Wright 2009) reject rejectionism by appeal to a meta-metaphysical picture entailing that the state of affairs of both sides of good abstraction principle are the same. Thus, so the thought goes, we do not need to make sure that the objects referred to by instances of the left-hand side of **(HP)** exist before we can responsibly stipulate that **(HP)** is to be true. All that is required is discharging some specific worries to the effect that both sides cannot express the same state of affairs, such as Bad Company worries. As I said in 2.4.3, making sense of this option, and arguing that it is compatible with Platonism certainly requires a lot of further work.

I do not want to commit myself to any particular response here, but I conclude that it is far from clear that we have to accept **(Existence)** as a presupposition. Moreover, there might be intermediate positions. Conceding that there are ontological presuppositions forces us to give up knowledge claims for these propositions, but it does not yet entail that we cannot claim knowledge of mathematical axioms. Granted, such positions look somewhat odd, but they should still be considered as fallback positions.

7.2.6 Presuppositions of uniqueness

There might not be a most eligible interpretation of the abstraction operator in relevant cases, and this might entail that meaning cannot be fixed at all. This worry is not specific to my proposal. For example, Lewis (1970) argues that the success of definitions depends on there being only one candidate meaning. This generates a presupposition of:

> **(Uniqueness)** There is a most eligible candidate meaning for the abstraction operator in question.

Do the **(Hume Rules)** meet this presupposition? Above, I have claimed that they do. The thought was that the **(Hume Rules)** meet **(Frege's Constraint)**, and that our notion of eligibility works in such a way that it assigns the natural numbers for this reason.

Of course, this assumption does not license us to dispense with the **(Uniqueness)** presupposition. It seems that there are bad stipulations where all presuppositions are met, except for **(Uniqueness)**. And these stipulations need to be ruled out. As an example, consider the following abstraction principle:

(Interestingly Underdetermined) $\forall FG\,(a\,(F) = a\,(G) \leftrightarrow x = x)$

On what basis are we to decide what object "α" assigns to concepts? Nothing specific comes to mind. This suggests that the stipulation should be ruled out as bad because it fails the **(Uniqueness)** test.

Two notes are in order. Firstly, one might think that all such cases are already ruled out by the **(Generality Condition)**, and that this renders the **(Uniqueness)** presupposition inessential. If reference is not uniquely determined, so the thought goes, the subject will not even be able to understand the abstraction operator. This thought, which is related to the Caesar problem, merits further research.

Secondly, one might hold that **(Interestingly Underdetermined)** is not a bad case. For example, one might think that our notion of eligibility works in such a way that it comes with its own specific kind of object — the *identity abstractum* — which we can only come to refer to by virtue of making the relevant stipulation.

In any case, we should remain open-minded about the possibility that **(Uniqueness)** is a further presupposition, although it is not clear whether this presupposition is essential. This completes my discussion of presuppositions of metasemantic success.

7.2.7 Epistemological presuppositions

There seem to be cases in which all presuppositions above are met and warranted, but in which something goes wrong from an epistemological point of view. This suggests that there is a further type of presuppositions, which we may call *epistemological presuppositions*. A good candidate is Hale's and Wright's *non-arrogance constraint* (see Hale & Wright 2000). Let me explain.

Non-arrogance Put very crudely, the non-arrogance constraint forbids that the stipulation *cries out for* additional epistemic work on the part of the subject. In 2.3, we have seen that the non-arrogance constraint admits of different interpretations:

- It might provide a special role for abstraction principles such as **(HP)**. The thought is that direct stipulations of axiom systems and existence claims — e.g. the Peano axioms — are arrogant, whereas stipulating abstraction principles is not arrogant because of their biconditional form.[177]

- It might rule out abstraction principles which allow for easy knowledge. For example: **(HP+FLT)** from 2.3.4, which allows for an easy proof of *Fermat's Last Theorem* (FLT).
- It might help to rule out stipulations like "Jack the Ripper is the unique perpetrator of these killings" or "Let God be the greatest being that can be conceived", because further *aposteriori* epistemic work needs to be carried out to license such stipulations.

We can ignore the first and the third role here, because they are not directly concerned with the stipulation of abstraction principles.[178] The case of **(HP+FLT)** is the one that is of interest here. The problem with **(HP+FLT)** is that it is a good stipulation from a metasemantic point of view, and that it is hard to see how the relevant presuppositions of metasemantic success could not be entitled. After all, we know that it fixes the same meaning as **(HP)**, since we know that FLT is a theorem of PA. However, Hero should not be able to acquire knowledge of FLT that easily. This motivates ruling out **(HP+FLT)** by postulating a further presupposition.

One idea is to simply demand that the stipulation should not allow for *easy knowledge*. However, this requires independently motivated criteria of what easy knowledge is. This is beyond the scope of this thesis.

Alternatively, one might argue that we can do without the non-arrogance constraint to rule out such stipulations. For one might think that arrogant stipulations are such that they provide sufficient reasons to believe that some of the other presuppositions of the **(Abstractionist Inference)** are not met, and that this defeats our entitlement for these presuppositions.

However, at first glance it is hard to see which of the conditions we already introduced could do the job. For example, intuitively, Hero would not have sufficient reasons to believe that **(HP+FLT)** is inconsistent.

Things might look better once we regard the truth-preservation of the **(Abstractionist Inference)** as one of its presuppositions. Arrogant stipulations might be such that we have

[177] We may call this thought *Hale's credo*. I do not remember a single talk on neo-Fregeanism, in which he did not stress that in order to generate existence claims out of **(HP)**, one first needs to establish appropriate instances of its right-hand-side.

[178] Of course, these roles become relevant when we extend the proposal to implicit definition in general.

sufficient reason to doubt that they are truth-preserving. Consider the case of **(HP+FLT)**. Once we observe that the stipulation builds in a substantial claim about the properties of the objects we want to introduce, we should ask what license we have to build in such a claim. This consideration might defeat our entitlement in relevant cases. However, a lot depends on what is required for having a sufficient reason to believe that a proposition is untrue. Seeing that **(HP+FLT)** builds in a substantial claim might provide a *sufficient reason to doubt* that we can responsibly make this stipulation, but it might not provide a sufficient reason to believe that the abstraction principle is *untrue*.

My conclusion is that whether or not we need a non-arrogance constraint, and how it would look like, can only be decided by further research.

7.2.8 Summary: essential presuppositions

This completes my discussion of the presuppositions of the **(Abstractionist Inference)**. To sum up. The following conditions are essential presuppositions of the **(Abstractionist Inference)**:

- The **(Generality Presupposition)**.
- Irenicity

Maybe we have to add one or more of the following:

- **(Uniqueness)**
- Non-arrogance

The **(Generality Presupposition)** is a presupposition of cognitive success — it is a presupposition ensuring that the abstraction principle generates understanding. Irenicity is a mathematical presupposition: it is used to avoid specific possible shortcomings of abstraction principles that arise from (mathematical) Bad Company considerations. I provided considerations to the effect that we can reject the claim that there are ontological presuppositions. This is important because such presuppositions would constrain the epistemological payoff of my proposal. However, there *might* be a metaphysical presupposition ensuring that a unique referent can be assigned: **(Uniqueness)**. The fourth *potential* presupposition is an epistemological presupposition. It might be required to rule out irresponsible stipulations such as **(HP+FLT)**.

In the course of my investigations, it has become apparent that the classical objections to neo-Fregeanism are connected to more general issues regarding presuppositions of the **(Abstractionist Inference)**:

- The Bad Company problem can be regarded as the problem of selecting appropriate mathematical presuppositions.
- One version of the Caesar problem threatens the claim that we possess an entitlement for the **(Generality Presupposition)**.[179]
- In the light of my proposal, epistemic rejectionism arises as the claim that we need to invoke ontological presuppositions.

[179] I have not discussed this above, but the reader will be able to see that some issues regarding **(Uniqueness)** are related to the Caesar problem for reference.

7.3 Extending the proposal: logic

Extending the proposal to the logical case is straightforward, given that a stipulation of the validity of relevant introduction and elimination rules can bring it about that the meaning of logical operators is fixed in such a way that the respective rules are valid. This is Gentzen's idea (Gentzen 1934) — **(Metasemantic Inferentialism)** for logic. I think this idea is very plausible. For example:

- The meaning of the material conditional "$\rightarrow$" can be fixed by virtue of stipulating that it is to be assigned a meaning that renders its elimination rule *Modus Ponens* (MP) and its introduction rule *Conditional Proof* (CP) valid:[180]

$$\textbf{(MP)}\ \frac{\Gamma,\phi\vdash\psi}{\Gamma\vdash\phi\rightarrow\psi}\quad \textbf{(CP)}\ \frac{\Gamma\vdash\phi\ \mathit{and}\ \Delta\vdash\phi\rightarrow\psi}{\Gamma,\Delta\vdash\psi}$$

- The meaning of conjunction "$\wedge$" can be fixed by virtue of stipulating that it is to be assigned a meaning that renders $\wedge$-*Introduction* ($\wedge$-I) and $\wedge$-*Elimination* ($\wedge$-E) valid:

$$(\wedge\text{-I})\ \frac{\phi,\psi}{\phi\wedge\psi}\quad (\wedge\text{-}\boldsymbol{E})\ \frac{\phi\wedge\psi}{\phi}$$

Corresponding to such meaning-fixing stipulations, there is another instance of the **(Implicit Definition Inference)**:

(Logicist Inference) There is a primitive type of inference allowing non-defective epistemic agents to acquire full internalistic apriori warrants for believing that introduction and elimination rules are valid, on the basis of internalistic knowledge of our explicit meaning-fixing stipulations of these rules, in case we also possess antecedent warrants for certain conditions ensuring that nothing went wrong with the stipulation in question.

One note is in order. Although I want to remain neutral as to what validity amounts to, it can be understood as necessary truth preservation throughout this chapter. The general proposal is neutral with respect to the notion one uses, although different notions might come with different presuppositions of the **(Logicist Inference)**.

[180] Throughout this section, I assume that "$\vdash$" is already understood.

7.3.1 Another Hero story

Consider again our non-defective epistemic agent Hero. Hero sits in his armchair and makes the following meta-linguistic stipulation, to fix the meaning of a new symbol "$\rightarrow$":

(Meaning-Fixing Stipulation for Arrow) "$\rightarrow$" is to be assigned a meaning that renders MP and CP valid.

The stipulation can again be conceived of as a command, which brings it about that a pattern of use is fixed for "$\rightarrow$" and Hero becomes disposed to reason in accordance with this pattern of use (i.e. Hero becomes disposed to reason in accordance with MP and CP). By virtue of this process, Hero comes to understand "$\rightarrow$".

After that, Hero reflects on his stipulation, by drawing the **(Logicist Inference)** on the basis of the following premise:

(Premise for Arrow) I sincerely made the **(Meaning-Fixing Stipulation for Arrow)**.

Let "$Stip(MP, CP)$" stand for **(Premise for Arrow)**, and let "$Valid(R)$" stands for "R is valid". Hero's argument is this:

(1) $Stip(MP, CP)$

(2) $Valid(MP)$ and $Valid(CP)$ (1), **(Logicist Inference)**

If Hero possesses internalistic knowledge of (1), he can come to possess internalistic knowledge of (2), given that he possesses warrants for the presuppositions of the **(Logicist Inference)**. We can assume that Hero possesses internalistic apriori knowledge of **(Premise for Arrow)**, because it would be an instance of maker's knowledge (see 7.1.3).[181] But what are the presuppositions of the **(Logicist Inference)**, and does Hero possess warrants for them?

[181] Note that I assume that Hero is already competent in English, and possesses the relevant concepts. This is important, because making relevant meaning-fixing stipulations and forming justified beliefs about them presupposes some logical, syntactical and conceptual competence.

7.3.2 Presuppositions in the logical case

The presuppositions of the **(Logicist Inference)** are the conditions ensuring that meaning can be fixed in the right way. We can determine these conditions by considering cases in which meaning-fixing commands of introduction and elimination rules fail. Consider the *tonk rules*, which are stipulated to be valid in order to fix the meaning of the operator "τ" (Prior 1960):

$$(\tau\text{-I})\ \frac{\phi}{\phi\tau\psi} \quad (\tau\text{-}E)\ \frac{\phi\tau\psi}{\psi}$$

These rules cannot be valid, because they lead to triviality. Choose any theorem of the background system for ϕ. A successive application of (τ-I) and (τ-E) can be used to establish any sentence whatsoever. It is thus impossible that the right meaning is fixed by virtue of stipulating them to be valid.

This motivates imposing additional conditions $C\,(R_{Intro}, R_{Elim})$ ensuring that a stipulation of introduction and elimination rules R_{Intro} and R_{Elim} is not defective in the way a stipulation of (τ-I) and (τ-E) is defective. There are at least two candidates for $C\,(R_{Intro}, R_{Elim})$, depending on how one analyses the failure of the tonk-rules.

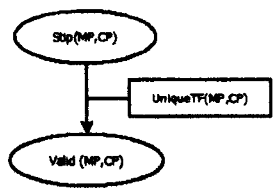
Firstly, one might argue that there is no truth function corresponding to the tonk-rules (Wagner 1981, Hjortland 2009, section 2.2.2). Secondly, Dummett (1991) proposed that the problem with the tonk-rules is that they are not in *harmony*. Very roughly, for a pair of introduction and elimination rules to be in harmony the results of applications of the elimination rule must not be stronger than the conditions for applying corresponding introduction rules and vice versa (Hjortland 2009 section 2.3.1).[182] It is immediate that (τ-I) and (τ-E) are not in harmony in this sense. We only need p to introduce $p\tau q$, but we can obtain $p \wedge q$ in turn, which is stronger than p.

This leads to two proposals of what $C\,(R_{Intro}, R_{Elim})$ include: one the one hand, one might demand that the stipulated introduction and elimination rules must determine a unique truth function. Let us denote this condition by "$UniqueTF\,(R_{Intro}, R_{Elim})$". On the other hand, one might demand that the stipulated introduction and elimination rules must be in harmony. Let us denote this condition by "$Harmony\,(R_{Intro}, R_{Elim})$". If one of these conditions is met, so the thought goes, by **(Metasemantic Inferentialism)**,

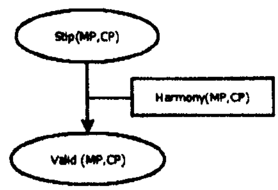
[182] This is the same harmony constraint that Hale and Wright proposed in their (2000, p.136).

meaning can be fixed in such a way that the stipulated inferential pattern turns out to be valid. Thus, one of these conditions is the only essential presupposition of the **(Logicist Inference)**.

We obtain two candidate models for how the structure of justification looks like. The first model may be called the *realist model*, since what is at stake is the existence of certain truth functions. The model invoking the Dummettian condition may be called the *anti-realist model*, because all that matters are syntactic considerations. Here is a graphical example for both models:



The realist model



The anti-realist model

I contend that the presuppositions of both models are entitlements of cognitive project in the relevant cases. I cannot argue for this claim in full generality. I will just complete the argument for the case of MP and CP.

7.3.3 Entitlements for presuppositions

The realist model According to the realist model, the presupposition of the **(Logicist Inference)** is that the stipulation determines a unique truth function, i.e. that there is a unique way of mapping all propositions to truth values such that the stipulated rules turn out to be valid.

Simple meta-theoretical reasoning — e.g. model-theoretic reasoning — shows that MP and CP determine a unique truth function. However, such arguments require a lot of resources. In particular, they make use of logical reasoning. Since logical reasoning presupposes the validity of the used logical rules, and, according to the current proposal, acquiring warrants for the validity of logical rules presupposes the existence of appropriate truth functions, we cannot justify that a unique truth-function is determined without implicitly relying on something in no more secure a prior epistemic standing. The regress clause for entitlement of cognitive project is met, at least in cases that are sufficiently basic. We can assume here, without loss of generality, that the case of MP and CP is such a case.

It is easy to see that the other conditions for entitlement of cognitive project are met as well. First, the conditions for the success of stipulations of basic logical rules are Wright-presuppositions of an important general cognitive project of finding out about the good standing of our meaning-fixing devices.[183] Secondly, we do not have sufficient reason to believe that MP and CP do not determine a unique truth function.

That MP and CP determine a unique truth function is an entitlement of cognitive project. According to the realist model, this is the only presupposition. All presupposition of the **(Logicist Inference)** for MP and CP are warranted. Thus, by **(Justification Generation)**, it transmits internalistic knowledge from its premise (the stipulation facts) to its conclusion (the validity claim).

The anti-realist model A similar argument applies to the anti-realist model. According to this model, the presupposition of the **(Logicist Inference)** for MP and CP is that MP and CP are in harmony.

It is very simple to argue for the claim that MP and CP are in harmony when harmony is understood along the lines sketched above (Gabbay 2007, p. 3). However, there is a variety of technical explications of the harmony constraint (see e.g. Hjortland 2009). How the argument looks like in full detail will depend on one's exact notion of harmony.

In any case, it is to be expected that the argument for MP and CP being in harmony requires logical reasoning. Since logical reasoning presupposes the validity of the used logical apparatus, and, according to the current proposal, our justification for the validity of the used logical apparatus presupposes that the used rules meet a harmony constraint, we cannot establish harmony in all cases without implicitly relying on something in no more secure a prior epistemic standing. The regress clause for entitlement of cognitive project is met in at least some cases, namely in those cases that are sufficiently basic. Again, we can safely assume that the case of MP and CP is such a case.

It is easy to see that the other conditions for entitlement of cognitive project are met as well. The reasons are analogous to those that apply to the realist model. According to the

[183] This is the same project of which the presuppositions of the **(Abstractionist Inference)** are entitlements (see 7.1.2).

anti-realist model, harmony is the only presupposition. All presupposition of the **(Logicist Inference)** for MP and CP are warranted. Thus, by **(Justification Generation)**, it transmits internalistic knowledge.

7.3.4 Defending the proposed structure of justification (2)

Analyzing the structure of justification in the logical case provides yet another argument for the claim that we need to conceive of the **(Implicit Definition Inference)** as primitive, with the preconditions for semantic success construed as presuppositions, as opposed to premises. The upshot is that alternative models cannot account for the generation of (internalistic) knowledge in our epistemological framework. The considerations are similar to those presented in 7.1.4, but due to the nature of the logical case there is an additional argument available.

One might think that the most natural way to acquire knowledge of validity on the basis of knowledge of one's own meta-linguistic stipulations is by reflecting on the metasemantic process. Why should Hero *directly* infer the validity of MP and CP from that he stipulated MP and CP to be valid? Why should Hero not rather go through an argument such as the following:

(1)	$Stip(MP, CP) \wedge C(MP, CP)$	
(2)	If (1), then "→" gets assigned a meaning such that MP and CP are valid.	Our metasemantics
(3)	MP and CP are valid.	(1), (2), MP

It is worth also considering a similar argument that Boghossian defends in his (1996). Boghossian's idea is that Hero can reason as follows (see also Ebert 2005*b*, Jenkins 2008*a*):

(1B)	"→" means what it does.	
(2B)	If "→" means what it does, then MP and CP are valid.	
(3B)	MP and CP are valid.	(1B), (2B), MP

Both arguments are subject to two devastating objections. The first problem is that both (1) and (1B) cannot have an epistemic status above (inferential) entitlement.

We have already seen that $C(MP, CP)$ is an entitlement. Thus, the conjunction (1) cannot have an epistemic status above (inferential) entitlement, because of the **(Limit)** principle. Something similar holds for (1B). Our warrant for it cannot be any stronger than our warrant for $C(MP, CP)$. For $C(MP, CP)$ are the conditions ensuring that "$\rightarrow$" gets assigned a meaning such that MP and CP are valid. If the notion of "means what it does" is to render (2B) true, it is plausible that we could only acquire a warrant for (1B) by using $C(MP, CP)$ as a premise.[184] Thus, by the **(Limit)** principle, (1B) cannot have an epistemic status above (inferential) entitlement.

So both arguments rest on an entitled premise. By the **(Limit)** principle, the conclusions cannot have an epistemic status stronger than (inferential) entitlement. This means that we cannot acquire internalistic knowledge by virtue of going through these arguments. The arguments are subject to leaching. Note that this problem is analogous to a problem that arises for the **(Metasemantic Reasoning Model)** in the mathematical case (see 7.1.4).

However, there is an additional problem in the logical case. The problem is that both arguments endorse an MP step. They are rule-circular. Thus, they fail to transmit warrant (for the argument, see 5.3). No first warrant can be obtained by virtue of going through the argument, and we cannot upgrade an antecedent warrant for validity by virtue of going through the argument.[185]

Ergo: we need to construe our argument for validity in such a way that it does not rest on entitled premises, and that it does not endorse logical inferential steps. The solution is to endorse the **(Logicist Inference)**.

7.3.5 Justification generation, and hyper-circularity

In the logical case, an interesting further circularity worry arises, which becomes apparent by applying the standard model of how **(Justification Generation)** is supposed to work. Let me explain.

[184] In other words: (1B) and (2B) need to be construed as claims about *reference* (see Ebert 2005*b*, Jenkins 2008*a*).

[185] A different argument for the claim that Boghossian's template fails to transmit warrant is provided by Jenkins (2008*a*). Jenkins argues that one cannot possess a warrant for (1B) without already possessing a warrant for a disquoted version of (3B) — note that the conclusion of Boghossian's template is a meta-linguistic statement about inferential patterns, in which "$\rightarrow$" is only mentioned and not used — and that the disquoted version of (3B) is the real target of the argument. Thus, so the thought goes, the argument already presupposes the possession of a warrant for what it is meant to establish.

Suppose Hero carried out the **(Logicist Inference)** for MP and CP correctly, from internalistically known premises $Stip\,(MP, CP)$, and that Hero possesses an entitlement for $C\,(MP, CP)$. **(Justification Generation)** tells us that Hero thus acquires internalistic knowledge of $Valid\,(MP)$. This requires that the possession of this knowledge is available to Hero on reflection. One explanation of why it is available to Hero — which I endorsed in both the perceptual and the mathematical case — is that Hero is able to go through some line of reasoning to the effect that he knows the conclusion of the **(Logicist Inference)**. Consider the simplest proposal of what Hero's argument might look like:

(1) $C\,(MP, CP)$

(2) $K\,(Stip\,(MP, CP))$

(3) $C\,(MP, CP) \wedge K\,(Stip\,(MP, CP)) \rightarrow K\,(Valid\,(MP))$

(4) $K\,(Valid\,(MP))$ (3), (4), MP

The availability of this line of reflection is supposed to show that Hero can come to possess an (inferential) entitlement for (4). However, the above reflection contains an MP step. So the one-step argument endorsing the **(Logicist Inference)** is circular in the following sense:

(Hyper-Circularity) An argument A for the goodness of a belief-forming method M is ***hyper-circular*** if and only if any (second-level) argument for the claim that one can acquire internalistic knowledge of, or internalistic justification for the conclusion of A by virtue of going through A, makes use of M.

The question arises of whether **(Hyper-Circularity)** is bad in that it undermines **(Justification Generation)**. If the answer to this question was positive, then the validity of our basic logical apparatus would have to remain a mere entitlement. This would not undermine our attempt to vindicate the proposal for mathematics, but the account would be incomplete. This issue merits further research.

7.3.6 Intermediate conclusion: logicism?

Let me sum up. In principle, the neo-Fregean Hero story can be extended to the logical case. On the basis of meta-linguistic stipulations and subsequent reflection in accordance with the **(Implicit Definition Inference)**, so the thought goes, Hero can acquire internalistic knowledge of the validity of basic logical laws.

As in the mathematical case, the proposal is limited by cases of transmission-failure. What exactly these limits are will be revealed by examining what the presuppositions of the **(Implicit Definition Inference)** are in the relevant cases. According to the realist model, claims about the unique determination of truth functions will have to remain entitlements. According to the anti-realist model, claims about the harmony of introduction and elimination rules will have to remain entitlements. Someone who thinks that claims to knowledge of harmony, or claims to knowledge of the unique determination of truth values are paradigmatic claims to knowledge will think that these are revisionary consequence. I regard these consequences as a philosophical insight into the limits of our (claimable) knowledge.

If the proposal for logic works, we will have shown that there is a canonical way of forming logico-mathematical beliefs that bases mathematics and logic on the same kind of basic belief-forming method. The **(Same Source)** constraint would be met. We will obtain a genuine version of logicism. This motivates further works on extending the proposal to the logical case.

7.4 Extending the proposal: implicit definition in general

The proposal generalizes further. The second generalization to be made is not directly concerned with the subject matter of stipulations, but with their form. So far, I focused on the stipulations of (the soundness or validity of) (introduction and elimination) *rules*. I now argue that the proposal can also be applied to *sentential* stipulations, and that the proposal thus provides an account of the epistemic workings of *implicit definitions* in general, including meaning-fixing stipulations of the truth of sentential matrices containing one or more undefined terms. However, I can only touch the surface here. The main purpose of this section is to carve out aspects of my proposal that merit further research.

Remember the distinction, drawn in 2.3.1, between the **(Traditional Conception of Definition)** and the **(Liberal Conception of Definition)**. According to the first conception, definitions have to meet the criteria of **(Eliminability)** and **(Conservativeness)**. Definitions meeting these criteria can be put in normal form, i.e. they can always be stated in terms of introduction and elimination rules (Gupta 2009).

Sentential stipulations of the truth of a certain explicit definition can be handled by my account, because they can be regarded as meta-linguistic stipulations to the effect that certain introduction and elimination rules are truth-preserving. The criteria of **(Eliminability)** and **(Conservativeness)** can be regarded as *essential presuppositions* of relevant instances of the **(Implicit Definition Inference)** (for the definition of "essential presupposition", see 7.2).

Hale and Wright (2000) defend the **(Liberal Conception of Definition)**, in order to argue for the apriori knowability of abstraction principles (Hale & Wright 2000). There are more applications for this conception of definition:

- It has been suggested that mathematical terms can be defined by virtue of stipulating the truth of (consistent) axiom systems. Carnap's idea in (1950) must have been along these lines, although there are some complications regarding the endorsed notion of truth here. And Hilbert can be read as expressing this idea in a letter to Frege (Hilbert to Frege 29.12.1899; in: Frege 1980).

- It has been suggested that theoretical terms of scientific theories can be defined by virtue of stipulating the truth of so-called *Carnap conditionals*. The idea can be traced back to Carnap (1966). The idea is taken up, among others, by Lewis (1970), Horwich (1998), and Hale and Wright (2000).

It does not appear too difficult to extend the account to these cases as well. All that needs to be done is to determine the essential presuppositions of the inferences associated with relevant cases of sentential definitions.[186]

7.4.1 The Sentential Stipulation Inference

As soon as we have found the right presuppositions, we can incorporate implicit sentential stipulations as follows. First, we need to make a metasemantic claim, corresponding to **(Metasemantic Inferentialism)**:

> **(Metasemantic Interpretationism)** By virtue of stipulating that a sentential matrix (or a collection of sentential matrices) containing one or more undefined terms is to be true (or necessarily true), we can (i) come to understand these terms, and (ii) bring it about that these terms get assigned a meaning such that the sentential matrix (or the collection) is true (or necessarily true), given that certain preconditions are met (which will also be the presuppositions of the corresponding inference).

To this we add an epistemic claim, corresponding to the **(Implicit Definition Inference)** principle:

> **(Sentential Stipulation Inference)** There is a primitive type of inference that allows us to acquire full internalistic apriori warrants for the (necessary) truth of our sentential stipulations on the basis of internalistic knowledge of what these stipulations are, in case we also possess antecedent warrants for the preconditions ensuring that nothing went wrong with the stipulations in question.

The **(Sentential Stipulation Inference)** can be defended in the same way as the **(Implicit Definition Inference)**, assuming that **(Metasemantic Interpretationism)** holds. And once we accept **(Metasemantic Inferentialism)**, it seems natural to accept **(Metasemantic Interpretationism)** as well.

[186]We can expect that new presuppositions arise for liberal stipulations of sentential matrices, or collections of sentential matrices. For it is much easier to make bad stipulations if one is not restricted to making explicit stipulations, or to stipulating introduction and elimination rules.

7.4.2 Stipulating abstraction principles

This opens up an alternative approach to abstractionism. We could tell our Hero story in such a way that Hero — who already grasps second-order logic — makes the following *sentential* stipulation to fix the meaning of "#":

> **(Meaning-Fixing Stipulation*)** "#" is be assigned a meaning such that the statement that **(HP)** holds necessarily is true (i.e. "$\Box HP$" is true).

The meaning that is assigned to "#" by this stipulation is exactly the same as the meaning assigned by the **(Meaning-Fixing Stipulation)** above. Moreover, it is plausible that both stipulations have the same presuppositions. I have argued above that Hero possesses entitlements for these presuppositions, so all the presuppositions of the relevant instance of the **(Sentential Stipulation Inference)** are warranted. Thus, Hero can come to possess internalistic knowledge of the truth of a boxed version of **(HP)** by virtue of the **(Sentential Stipulation Inference)**.

Hero can then apply a disquotational step to acquire internalistic knowledge of $\Box HP$, and of **(HP)** in turn. The disquotational step is unproblematic, because Hero already understands **(HP)**, by virtue of being disposed to use "#" in the right way, which has been brought about by sincerely making the meaning-fixing command.[187] Hero can then carry out Frege's proof as it is usually presented, using **(HP)** as a premise, as opposed to the **(Hume Rules)**.

7.4.3 A special role for abstraction principles?

If we allow sentential stipulations, two further questions arise for the abstractionist. Firstly, the question arises of what distinguishes abstraction principles from axiom sets with regard to their special role as foundations of (reconstructed) mathematical knowledge. Why should we not stipulate the (necessary) truth of the second-order Peano axioms directly, bring it about that the relevant terms have meaning, and come to know them by using an appropriate instance of the **(Sentential Stipulation Inference)**? If this possibility cannot be excluded, why should we take the long route via Frege's proof and (a stipulation of) **(HP)**, or the **(Hume Rules)**?

[187]The disquotational step has been problematized by both Ebert (2005*b*) and Jenkins (2008*a*). I briefly discuss the problem below (7.4.6).

To make this worry more precise, consider the following variation of our Hero story. In the alternative scenario, Hero* — who already grasps second-order logic — makes the following sentential stipulation to fix the meaning of "0", "$S(x)$" (successor), and "$N(x)$" ("x is a natural number"):

(Meaning-Fixing Stipulation)** "0", "$S(x)$", and "$N(x)$" are to be assigned a meaning such that the boxed conjunction of the second-order Dedekind-Peano axioms, relativized to a natural number predicate "$N(x)$", is necessarily true.

According to the current proposal, this stipulation brings it about that "0", "$S(x)$", and "$N(x)$" get assigned a meaning such that the stipulation indeed turns out to be true. Moreover, one might think that, if the presuppositions of a stipulation of **(Meaning-Fixing Stipulation*)** are entitled, so are the presuppositions of **(Meaning-Fixing Stipulation**)**. Thus, so the thought goes, Hero* can come to possess internalistic apriori knowledge of the (necessary) truth of a version of the second-order Peano axioms by drawing the appropriate instance of the **(Sentential Stipulation Inference)**, and a disquotational step.

The described route thus promises internalistic knowledge of a version of the second-order Peano axioms, without the need to carry out Frege's proof. Would it not be much more convenient for Hero to follow Hero*?

First, note that although Hero* can acquire internalistic knowledge of a version of second-order arithmetic by the described route, he needs to do more in order to acquire something akin to the knowledge of Hero. Hero* does not yet possess internalistic knowledge of Frege Arithmetic, for making the command **(Meaning-Fixing Stipulation**)** does not fix the meaning of a number operator like "#". And although Hero can interpret FA in his theory — Boolos (1987) showed how to define "#" and how to proof a version **(HP)** in second-order arithmetic in the course of establishing the equi-consistency of second-order arithmetic and FA — the question remains whether Hero*'s knowledge is really akin to Hero's in the end, because it is not clear that the meanings of Hero*'s terms are the same as the meanings of Hero's terms.

This question is relevant. I have argued above that Hero's stipulation promises to meet hermeneutic demands, i.e. that the meaning fixed for "$\#F$" by Hero's stipulation is the same as the meaning of "the number of the Fs". If Hero*'s stipulation does meet this demand, we have strong reasons to prefer the route via abstraction principles.

The argument for Hero's stipulation meeting hermeneutic demands rests on the claim that a stipulation of **(HP)** — or a stipulation of the **(Hume Rules)** — meets **(Frege's Constraint)**. And, at least prima facie, a direct stipulation of the axioms of second-order arithmetic fails this constraint. It does not make the applications of numbers immediate. So we cannot use the same argument for Hero*'s stipulation. This issue deserves further research.

7.4.4 Against epistemic analyticity

The **(Implicit Definition Thesis)** is sometimes associated with the notion of *epistemic analyticity*. The notion has first been endorsed by Boghossian (1996). Boghossian argues that, although the classical notion of analyticity — truth in virtue of meaning[188] — is either unclear or uninstantiated[189], there is an interesting epistemological notion in the vicinity, namely:

> **(Epistemic Analyticity)** A sentence S is *epistemically analytic* if and only if understanding S suffices for being in a position to acquire a warrant for the proposition expressed by S.[190]

According to Boghossian, some interesting statements which have been claimed to be true in virtue of meaning are epistemically analytic. For example: statements expressing the validity of basic logical laws. Boghossian provides an argument template for establishing the validity of basic logical laws to explain why this is so. We have seen one instance of it above, and I have argued that it is defective (see 7.3.4). However, since my own template avoids the difficulties of Boghossian's template, one might think that the notion of epistemic analyticity can be vindicated after all.

So are validity claims and sentential stipulations like **(HP)** epistemically analytic? No. In fact, the considerations above show that one should not posit any direct link from understanding to being in a position to acquire a warrant.

[188] Boghossian dubs it *metaphysical analyticity*.
[189] See (Williamson 2007, chapter 3) for additional considerations in this regard.
[190] Every notion of analyticity should be construed as a property of sentences, as opposed to propositions (Russell 2008, p. 22).

Consider again the case of Frege's **(BLV)**. Although a stipulation of **(BLV)** suffices to generate an understanding of "ϵ", **(BLV)** cannot be true, because of Russell's paradox.[191] Thus, there is no guarantee that we can acquire knowledge of the truth of stipulations on the basis of knowledge of what our stipulations are.

And such cases also undermine the claim that knowledge of our stipulations already suffices for being in a position to acquire a warrant for their truth (or truth-preservation). This is why we conceived of the **(Implicit Definition Inference)** as an inference with *presuppositions*. One also needs to possess supporting warrants in order to acquire a warrant by virtue of drawing the inference. The following claim is false:

> **(Warrant by Understanding alone)** Knowing that one has just made a certain meta-linguistic stipulation to fix the meaning of t suffices for being in a position to acquire a warrant for the goodness of the definition.

Only the following, weaker principle holds:

> **(Warrant by Understanding AND warrants for success)** Knowing that one has just made a meta-linguistic stipulation to fix the meaning of t suffices for being in a position to acquire a warrant for the goodness of the definition, given one also possesses warrants for the relevant preconditions of semantic success.

This bears directly on the notion of epistemic analyticity. Ceteris paribus, that one knows that one has just made a meta-linguistic stipulation to fix the meaning of t entails that one understands t. If understanding was sufficient for being in a position to acquire a warrant, then knowing that one has just made a meta-linguistic stipulation would be sufficient as well. But we have just seen that it is not. Thus, we should also reject the claim that *understanding* entails being in a position to acquire a warrant.

7.4.5 A third notion of analyticity

However, my account motivates introducing a notion weaker than, but still close to the notion of epistemic analyticity. The notion of analyticity my proposal really motivates is simply the notion of a statement internalistically knowable on the basis of one of the simple inferences above, and logic. Such statements are epistemically interesting: they can

[191] I assume that our background logic is second-order logic with unrestricted comprehension.

be internalistically known just on the basis of apriori knowable stipulation facts. We may call such statements *Carnap analytic*:

(Carnap Analyticity) A sentence S is *Carnap analytic* if and only if it is in the deductive closure of statements that can be internalistically known on the basis of instances of the **(Sentential Stipulation Inference)** or the **(Implicit Definition Inference)**.[192]

There is another, more fundamental way to express the notion of Carnap analyticity, which assumes that warrants for the validity of basic logical rules (and the soundness of mathematical rules) can be acquired by virtue of the **(Implicit Definition Inference)**:

(Carnap Analyticity*) A sentence S is *Carnap analytic* if and only if it is in the R-closure of conclusions of instances of the **(Sentential Stipulation Inference)** and the **(Implicit Definition Inference)**, where R-closure is closure under rules whose validity or soundness can be internalistically known on the basis of the **(Implicit Definition Inference)**.

There are a lot of Carnap analytic truths. For example:

- The axioms and theorems of Frege Arithmetic: these can be known on the basis of **(HP)** or the **(Hume Rules)**, and logical rules.
- Statements expressing the validity of MP and CP, and other basic logical laws. They are the conclusions of good instances of the **(Logicist Inference)**.
- Tautologies. They can be obtained on the basis of proofs without premises, using only logical rules.

Moreover, if the proposal for arithmetic can be combined with a weak hermeneutic claim (and I have explained in 7.1.8 how such a claim may be established), then Hero will be able to justify apriori that, necessarily, a statement of ordinary arithmetic is true if and only if the corresponding reconstructed statement is true. This motivates the following notion:

[192] I am grateful to Thomas Brouwer for making me aware of a deficiency in an earlier definition of this notion.

(Metasemantic Analyticity) A sentence S is metasemantically analytic if one can acquire apriori knowledge to the effect that necessarily, the proposition expressed by S is true if the proposition expressed by a certain Carnap analytic statement is true.

All metasemantically analytic truths are knowable apriori. If an appropriate hermeneutic claim for abstractionism can be defended apriori (see 7.1.8), many truths of ordinary arithmetic will turn out to be metasemantically analytic, and thus knowable apriori. I hope that most of what has traditionally been deemed analytic turns out to be metasemantically analytic. This merits further research.

7.4.6 Ebert's proposal

In his PhD thesis (Ebert 2005*a*), Ebert presents his own approach to precisifying Hale's and Wright's idea of knowledge by stipulation. At a first glance, his approach might look very similar to mine. However, there are major differences between our approaches, which need to be pointed out.

According to Ebert, the act of stipulation is best described as a direct, primitive acceptance of a sentential matrix containing an undefined term (Ebert 2005*a*, pp. 221f).[193] Ebert identifies some features of primitive acceptance, which correspond to features of belief. For example, the acceptance of a statement S gives rise to assertions of S (Ebert 2005*a*, p. 252). However, Ebert repeatedly stresses that primitive acceptance is a state different from belief, and considers it to be an important objection to his proposal that acceptance might be a doxastic state too weak to underlie knowledge (Ebert 2005*a*, p. 252).[194]

In any case, according to Ebert, stipulations qua primitive acceptances admit of three dimensions of success that are all in place in the paradigmatic cases:[195]

- Effectiveness: If the stipulation meets certain conditions — in Ebert's terms: if the

[193]It seems that Ebert wants to allow for the stipulation of rules as well, but he does not discuss this case in any detail.

[194]Although Ebert seems to conceive of stipulations as primitive acceptances of statements — linguistic entities — it does not always become clear whether it is really a statement that is accepted, rather than the proposition expressed by it. In fact, construing acceptance as a relation between a subject and a linguistic entity is very implausible. For example: how could warranted true acceptance constitue knowledge, as Ebert contends, without acceptance being a relation between a subject and a *proposition*? Presumably, the thought is that the acceptance of a statement brings with it an (additional) acceptance of a proposition, once the statement becomes understood.

[195]The names for the three dimensions of achievement are taken from (Ebert 2011).

stipulation is *effective* — the subject is "directly confronted with the content of what is stipulated" (Ebert 2005*a*, p. 224). Thus, acceptance immediately leads to an understanding of the content expressed by the stipulated matrix. This raises the question what exactly the conditions for effectiveness are. For Ebert, a minimal condition is Hale's and Wright's Generality condition which I discussed in 2.3 and 7.2 (Ebert 2005*a*, p. 124).

- Success: A fortiori, in case the stipulation meets some further preconditions, the stipulation brings it about that the undefined term gets assigned a meaning such that the matrix is true.

- Epistemic effectiveness: according to Ebert, if certain preconditions are in place — Ebert's calls them "(epistemic) presuppositions" — the accepted pattern is not only true, but becomes an item of (non-inferential) *externalistic* knowledge (Ebert 2005*a*, p. 124). Moreover, if the subject meets some additional conditions the subject can also *claim* this knowledge. Let me explain.

The (epistemic) presuppositions Ebert has in mind are similar to the conditions Hale and Wright impose in their (2000). First and foremost, they include consistency and conservativeness (Ebert 2005*a*, p. 226).[196][197] The relevant presuppositions are met by relevant abstraction principles. For example: **(HP)**. Thus, if Hero accepts **(HP)** to fix the meaning of "#", "#" becomes understood by Hero and gets assigned a meaning that renders **(HP)** true. Moreover, Hero thereby comes to possess (non-inferential) externalistic knowledge of **(HP)**, or so the thought goes.[198]

Ebert does not draw the distinction between hermeneutic and revolutionary programmes. However, I take it that he assumes that the meaning fixed for "#" is such that Hero can come to possess externalistic knowledge of the existence and the properties of the numbers we all know and love.

The story does not end here. Ebert is dissatisfied with the outlook of merely establishing

[196] Ebert does not impose a non-arrogance condition. However, he thinks he can avoid certain easy knowledge worries by referring to the special form of abstraction principles (Ebert 2005*a*, pp. 253f).

[197] Ebert also draws an analogy to the external-world case. He agrees with Wright in that, when it comes to perceptual belief-forming methods, there are preconditions such as the non-obtaining of sceptical scenarios (Ebert 2005*a*, pp. 225f). As soon as these conditions are met, so the thought goes, a subject can acquire externalistic knowledge by virtue of carrying out perceptual belief-forming methods.

[198] Ebert does not endorse the Hero metaphor. I use it to render clearer the connections between my proposal and Ebert's proposal.

that it is *possible* to possess arithmetical knowledge, externalistically conceived. He also wants to account for the possibility to claim such knowledge (Ebert 2005*a*, p. 226).[199]

Claiming knowledge of **(HP)** requires some kind of reflectivity. In particular, it cannot possibly suffice that the epistemic presuppositions are true. However, Ebert also rightly observes that it is better not required that Hero *proves* that the preconditions are met. For this leads to insurmountable regress problems (Ebert 2005*a*, p. 231).

This dilemma[200] is resolved by appealing to entitlements (of cognitive project). According to Ebert, "for a subject to claim to know p on the basis of a primitive acceptance, he has to make sure that he is entitled to do so" (Ebert 2005*a*, p. 248). And to be entitled to claim knowledge of p just *consists in* possessing entitlements for the epistemic presuppositions, or so the thought goes (Ebert 2005*a*, p. 248). Thus, the condition for properly asserting knowledge of a primitively accepted **(HP)** is that Hero *has made sure* that he possesses entitlements for all the epistemic presuppositions.

Of course, Ebert also holds that Hero possesses relevant entitlements, and can make sure that he possesses them. For example, Hero possesses an entitlement for **(HP)**'s consistency, and Hero can make sure that he possesses it (Ebert 2005*a*, p. 247).

What does it involve to *make sure* that one is entitled? Does Hero need to *prove* that the conditions for entitlement are met, or is some looser kind of access enough? Ebert gives a somewhat vague answer for the case of **(HP)**'s consistency (Ebert 2005*a*, p. 249). As to condition (i), Ebert thinks that Hero needs to *show* that none of the *known* paradoxes applies to **(HP)**. As to condition (ii), Ebert thinks that Hero needs to *show* that any attempt to justify the consistency of **(HP)** involves further presuppositions in no more secure a prior epistemic standing. In this context, "to show" sounds just like "to prove". This suggests that it is not easy to claim knowledge of **(HP)**. For example, if "known paradoxes" refers to the paradoxes known to the experts, quite a bit of logical knowledge is required on Hero's side. And if one needs to prove that one cannot justify **(HP)**'s consistency without regress, one needs to know of Gödel's results.

Be that as it may, there is a further unclarity: Ebert does not explain why Hero can claim knowledge once he has made sure that he possesses entitlements for all the preconditions. Maybe Ebert thinks that this is because some additional kind of reasoning

[199]Thus, Ebert seems to share the Wrightean intuition that I discussed in chapter 3.
[200]This is, of course, just another version of the Shapiro-Ebert dilemma I discussed in 2.3.

on Hero's side can yield an inferential entitlement for the claim that he knows, and that this is sufficient for responsibly claiming knowledge. This adoption of (my interpretation of) Wright's suggestion would make Ebert proposal look similar to mine. What are the most relevant differences, if we assume that this is how we should complete the picture?

The most relevant differences are that Ebert thinks (i) that knowledge of abstraction principles in *non-inferential*, and (ii) that it is based on stipulations conceived of as *primitive acceptances* of sentences of the object language. My own proposal is incompatible with both claims. As to (i), I have argued that knowledge of abstraction principles is inferential. It is acquired by a primitive *inferential* belief-forming method on the basis of maker's knowledge about one's own stipulations. As to (ii), I conceive of stipulations as meaning-fixing *meta-linguistic commands*. So both the type of knowledge generated, and the act of stipulation are fundamentally different.

I believe that there are things to say in favour of my own choices, and that the reasons Ebert gives in favour of his choices are not convincing.

Ebert offers two reasons for the claim that we should construe knowledge by stipulation as non-inferential and based on primitive acceptance. Firstly, Ebert wants to avoid appeal to a disquotational step. That is: he wants to avoid conceiving of the structure of justification in such a way that Hero first acquires knowledge of the *truth* of **(HP)** — a meta-language claim about a sentence of an object language — and then acquires knowledge of **(HP)** by a disquotational step. His aim is to allow for Hero acquiring knowledge of the **(HP)** directly. Although I think that I could also construe the **(Abstractionist Inference)** such that it directly delivers object language claims, the way I laid down my proposal above (see 7.4.2) requires a disquotational step after a sentential stipulation of **(HP)** has been made, and the truth of **(HP)** has been inferred by virtue of the **(Abstractionist Inference)**.

The disquotational step has first been problematized by Ebert (2005*b*). Jenkins (2008*a*) shares the underlying worry. Both Ebert and Jenkins observe that being able to disquote presupposes an understanding of the sentence to be disquoted, and contend that the sort of understanding required is so substantial that it requires the possession of a warrant for the proposition the disquoted sentence expresses, in cases where the sentence to be disquoted is the device of a meaning-fixing stipulation. If this was true, then the argument involving the disquotational step could not be used to acquire a first warrant for **(HP)**, because one of its inferential steps would already presuppose the possession of a warrant for **(HP)**. In fact, the argument would display transmission-failure, because it would exhibit the **(Information Dependence Template)**.

I reject the claim that understanding the matrix of a meaning-fixing stipulation requires the possession of a warrant for the proposition expressed by it. One can understand "ϵ" — and sentences including "ϵ", such as **(BLV)** — on the basis of a meaning-fixing stipulation of **(BLV)**, without possessing a warrant for it. We can even know that **(BLV)** is inconsistent, and nevertheless understand it and use it in our reasoning. On my view, all that is required for understanding is that one has the right inferential dispositions (which can be overridden).

So I do not think that the envisaged problem arises because of the conditions for understanding. However, a full discussion of this worry certainly requires investigating the conditions for *concept possession*, since the argument might be stated, and is probably is intended to apply, at the level of thought. Does possessing the concept of naive extension require the possession of a warrant for **(BLV)**, if the concept is introduced by a stipulation of the truth of **(BLV)**? I cannot do more here than noting that I do not see why this should be the case. I grasp the concept of naive extension, and I know which concept is assigned to "ϵ" by a stipulation of the truth of **(BLV)**, without possessing a warrant for **(BLV)**.

Maybe the point is that a specific step from $T('HP')$ to HP can only transmit warrant if one already possesses a warrant for HP, *even if* one already understands HP. I do not see why this should be the case.[201] I conclude that it is not a good reason to prefer Ebert's proposal that a disquotational step must be avoided.

[201] Imagine Hero commands that "#" is to be assigned a meaning such that **(HP)** is true. According to my proposal, Hero can acquire a full internalistic warrant for the truth of **(HP)** by virtue of drawing the **(Sentential Stipulation Inference)**. Moreover, by virtue of having made the meaning-fixing command, Hero understands "#" and **(HP)**. Why should Hero not be able to acquire a full internalistic warrant for **(HP)** in this case, by virtue of applying a disquotational step, without already possessing a warrant for **(HP)**?

I suspect that at least Jenkins (2008*a*) thinks that the real worry is Bonjour's worry about implicit definition (Bonjour 1998, §2.5), i.e. the worry that one already needs to possess a warrant for **(HP)** in order to acquire a warrant for the claim that the stipulation has been *successful* and that **(HP)** is *true*. According to my proposal, Hero can acquire a warrant for the truth of **(HP)** if he possesses warrants for the presuppositions of the **(Sentential Stipulation Inference)**. So the real worry might be that **(HP)** is among the presuppositions. Why? In general, the worry might be that the stipulations themselves are presuppositions of the suggested primitive inferences, maybe because of the phenomenon of *presupposition expansion* (see 6.1.6). This is an interesting objection. Although I think it can be rebutted, it merits further research.

Ebert offers a second reason for conceiving of the structure of justification as he does. According to Ebert, his account of primitive acceptance "nicely captures the phenomenology of knowledge acquisition in mathematics" (Ebert 2005*a*, p.222). According to Ebert, axioms are just accepted, and no reasoning is involved.

However, firstly, the phenomenology of actual mathematics cannot be decisive in a reconstructive epistemological project. We are not looking for actual belief-forming methods, but for canonical belief-forming methods. Consider the following fact: although the gold standard of mathematical reasoning is (formal) proof, many mathematical beliefs are not formed on this basis. Mathematicians often say: "I see that this follows from that". But this does not show that we should not assign a special role to formal proofs.

Secondly, mathematicians often explicitly take the axioms to define their terms, and mention this to justify the axioms, as e.g. Hilbert in the *Foundations of Geometry* (Hilbert 1903). I could just as well offer *this* fact about actual mathematical practice to support my own proposal.

Note that my proposal can account for something that looks just like primitive acceptance. Often the premises of arguments are not explicitly mentioned (or thought of). A lot of times, belief-formation happens implicitly, and (partly) unconsciously. We come to believe that a place is dangerous without explicitly or consciously basing this belief on anything, or even without being able to articulate the basis for this belief, although the belief has a basis, such as other beliefs and experiential states. The same might be said about the belief-forming methods for mathematical axioms. The belief-forming process underlying our beliefs in axioms might be based on propositions about meaning-fixing commands, but this premise might be suppressed and not become explicit in many cases.

I now turn to a critique of Ebert's proposal. I think that his non-inferential conception cannot easily account for the fact that axioms are *believed*. For Ebert, stipulation is acceptance, and acceptance is not belief. What underlies this concession seems to be Wright's thought that genuine belief requires the possession of evidence.[202] However, according to Ebert's proposal, Hero just accepts **(HP)**. This acceptance is not *based on* any evidence.

This is problematic for two reasons. Firstly, we clearly *believe* that the axioms are true,

[202] This is why Wright (2004*b*) construes entitlements as warrants to *trust*, and not as warrants to *believe* (see 4.4).

and we need to explain how we come to believe that axioms are true. Secondly, Ebert admits that, even assuming that acceptance is a belief-like state, substantial argument is needed to establish that knowledge can be based on mere acceptance (Ebert 2005*a*, p. 252).

I think that accounting for knowledge will be very difficult, since the possession of knowledge — even in the externalist sense — requires the possession of evidence. If there is no evidence, there cannot be knowledge. Ebert's proposal construes the structure of justification in a wrong way. We need to account for there being evidence for (the truth of) axioms.

Ebert might respond in at least two ways. Either he accepts that there isn't any evidence for (the truth of) abstraction principles and argues that this does not lead to undesired consequences, or he accounts for there being evidence for them after all. I cannot see how the first response can be made good. As to the second response, the envisaged evidence could either be a (warranted) proposition, or an experiential state. If it is the latter, the proposal will look dangerously similar to rational intuition proposals. If it is the former, then the proposal will collapse into something very similar to my proposal. For what could the proposition be that Hero uses to justify **(HP)**? The best choice is the proposition, knowable by introspection, that he has made a certain stipulation.

7.5 Intermediate conclusion

In the first part of this chapter, I presented an account of the epistemic working of the neo-Fregean implicit definition proposal, and applied it to mathematics. In particular, I argued that there is a non-deductive inferential route to justifying the good standing of explicit meaning-fixing stipulations. I sketched how our non-defective epistemic agent — Hero — could come to possess internalistic knowledge of arithmetical truths. The demands on Hero were relatively moderate: the reader of this thesis could be Hero. This vindicates the **(Traditional Epistemic Project)** for mathematics.

The limits of the proposal come with transmission-failure. Only a thorough investigation of the presuppositions of semantic success can reveal these limits. To this end, I have examined these presuppositions in more detail. There are interesting connections between three classical objections to neo-Fregeanism and the question what the presuppositions for semantic success are. The proposal sheds new light on these worries.

In the third part of this chapter, I argued that my proposal generalizes, and considered various objections and rejoinders. I then applied the proposal to logic, and sketched how it might be applied to implicit definitions in general. On this basis, I argued that my proposal promises to vindicate two epistemologically interesting notions of analyticity: **(Carnap analyticity)** and **(Metasemantic Analyticity)**.

8 Conclusion

To conclude, I sketch what I take to be the most important achievements of this thesis, and make some specific remarks as to where I think further research is needed.

8.1 Insights

The following points I regard as insights:

- **(Frege's Constraint)** has an important role to play in any philosophy of mathematics that bases mathematical knowledge on logic and definitions. Among other things, it is a means to ensure that we engage in a hermeneutic reconstructive project, as opposed to a revolutionary one (1.2, 2.3.5, 7.1.8).

- Some epistemological issues that have only been examined in the context of external-world scepticism are relevant to the epistemology of mathematics and logic. For example, logico-mathematical belief-forming methods have characteristic presuppositions (4.2, 7.2, 7.3.3); Wright's Instability Argument for **(Relevance Internalism)** can be applied to the mathematical case (3.5.2).

- There are at least two ways to argue for the idea that belief-forming methods with non-evidentially warranted presuppositions confer or transmit full internalistic warrant (and internalistic knowledge). One of the strategies endorses an externalistic notion of knowledge. Thus, we can meet the demands of **(Relevance Internalism)** regardless of whether the ordinary notion of knowledge is an externalistic notion (3.3.3, 4.5).

- In 5.3, I argued that rule-circular arguments fail to transmit warrant because they exhibit the **(Information Dependence Template)**.

- There might be arguments which fail to transmit warrant, but not obviously so. This is because an entitled presupposition of an inferential steps entails the conclusion of the argument, and the entailment is such that the conclusion also has to count as an entitled presupposition. I called this phenomenon *presupposition expansion* (6.1.6).

- In chapter 6, I have shown that it would not be devastating to our epistemology if all of mathematics and logic was a mere entitlement. Directly applying the entitlement proposal to mathematics remains an interesting option.

- However, as we have seen in chapter 7, we can do better by embedding the neo-Fregean **(Implicit Definition Thesis)** into a general anti-sceptical internalistic epistemological framework.

- The framework enables us to determine the boundaries of claimable knowledge in logico-mathematical cases. One result is that, although the Peano axioms can be claimed to be known, we cannot claim to know their consistency (7.1.7).

- Boghossian's notion of epistemic analyticity is flawed, but something close to it can be vindicated (7.4.5).

8.2 Open questions

Among other things, the following issues merit further research:

- The question of whether there is a viable form of epistemic consequentialism which implies that entitlements have epistemic value (4.6).

- The question of what exactly the epistemic value of the possession of evidence and knowledge is, as opposed to the epistemic value of mere entitlement (5.1).

- The question under what consequence relation entitled presuppositions are closed. Answering this question is crucial to understand the phenomenon of presupposition expansion mentioned above (6.1.6).

- The question of whether particular issues with implicit definitions in the logical case can be solved. For example: the problem of hyper-circularity (7.3.5).

As to the **(Implicit Definition Thesis)**, the following more general questions deserve further examination:

- Regarding its interaction with epistemology: what is the extent of the **(Implicit Definition Thesis)** beyond logic and mathematics? Can it be applied to the wider apriori? For example: conceptual truths of ordinary language ("everything coloured is extended"), the apriori content of scientific theories (e.g. their Carnap conditionals), and philosophical theories and frameworks.

- Regarding its interaction with philosophy of language: how does the proposal work together with different *meta*-semantic theories? What is presupposed for a meaning-giving stipulation to be successful depends on what meaning is. It would be interesting to examine the consequences of adopting proof-theoretic vs. reference-theoretic accounts of meaning in logico-mathematical cases in more detail. This investigation promises to reveal (i) how much the account can really achieve on the epistemological side — the more substantial the presuppositions are the less epistemic progress we make; (ii) whether the mathematical and the logical case can be treated uniformly.

- Regarding its interaction with metaphysics: my account of knowledge-by-stipulation stands ready to explain our knowledge of any subject-matter that can be treated via abstraction principles in the neo-Fregean way. It would be interesting to take up the question of how far this can take us: can abstraction-based accounts be given of non-concrete entities appealed to in science and everyday life: linguistic types, biological kinds, nation-states, and musical works?

References

Alston, W. P. (2005), *Beyond Justification*, Cornell University Press.

Altschul, J. (2011), Epistemic Entitlement, *in* 'The Internet Encyclopedia of Philosophy'. Version of 25/07/2011.
URL: *http://www.iep.utm.edu/ep-en/*

Anscombe, G. (1957), *Intention*, Basil Blackwell.

Arntzenius, F. & Dorr, C. (forthcoming), Calculus As Geometry, *in* 'Space, Time and Stuff', Oxford University Press, chapter 8.

Ashwell, L. (2010), 'Superficial Dispositionalism', *Australasian Journal of Philosophy* **88**(4), 635–653.

Baker, A. (2005), 'Are there Genuine Mathematical Explanations of Physical Phenomena?', *Mind* **114**, 223–238.

Bangu, S. (2008), 'Inference to the best explanation and mathematical realism', *Synthese* **160**, 12–20.

Bealer, G. (2000), 'A Theory of the A Priori', *Pacific Philosophical Quarterly* **81**, 1–30.

Benacerraf, P. (1973), 'Mathematical Truth', *The Journal of Philosophy* **70**, 661–680.

Bergmann, M. (2000), 'Externalism and Skepticism', *The Philosophical Review* **109**(2), 159–194.

Boghossian, P. (1996), 'Analyticity Reconsidered', *Noûs* **30**, 360–391.

Boghossian, P. (2001), 'How are objective epistemic reasons possible?', *Philosophical Studies* **106**, 1–40.

Bonjour, L. (1998), *In Defense of Pure Reason*, Cambridge: Cambridge University Press.

Boolos, G. (1987), The Consistency of Frege's Foundations of Arithmetic, *in* J. J. Thomson, ed., 'On Being and Saying: Essays in Honor of Richard Cartwright', MIT Press, pp. 3–20.

Boolos, G. (1990), The Standard of Equality of Numbers, *in* G. Boolos, ed., 'Meaning and Method: Essays in Honor of Hilary Putnam', Cambridge University Press, pp. 261–277. Reprinted in: Boolos 1998, pp. 202-219.

Boolos, G. (1997), Is Hume's Principle Analytic?, *in* R. J. Heck, ed., 'Logic, Language, and Thought', Oxford University Press. Reprinted in Boolos 1998, pp. 301-314.

Boolos, G. (1998), *Logic, Logic and Logic*, Harvard University Press.

Bostock, D. (1997), *Intermediate Logic*, Clarendon Press.

Bueno, O. & Colyvan, M. (2011), 'An Inferential Conception of the Application of Mathematics', *Noûs* **45**(2), 345–374.

Burge, T. (1993), 'Content Preservation', *Philosophical Review* **102**(4), 457–488.

Burgess, J. P. (2005), *Fixing Frege*, Princeton University Press.

Burgess, J. & Rosen, G. (1997), *A Subject with no Object*, Clarendon Press.

Carnap, R. (1950), 'Empiricism, Semantics, and Ontology', *Revue Internationale De Philosophie* **4**(2), 20–40.

Carnap, R. (1966), *Philosophical Foundations of Physics*, Basic Books.

Carroll, L. (1895), 'What the Tortoise Said to Achilles', *Mind* **4**(14), 278–280.

Chisholm, R. (1977), *Theory of Knowledge*, Prentice-Hall.

Choi, S. & Fara, M. (2012), Dispositions, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', spring 2012 edn.

Colyvan, M. (2002), 'Mathematics and Aesthetic Considerations in Science', *Mind* **111**(441), 69–74.

Conee, E. & Feldman, R. (2001), 'Internalism Defended', *American Philosophical Quarterly* **38**(1), 1–18.

Cook, R. (2009), 'Hume's Big Brother: Counting Concepts and the Bad Company Objection', *Synthese* **170**, 249–369.

Cook, R. & Ebert, P. (2005), 'Abstracts and Identity', *Dialectica* **59**(2), 121–139.

Davidson, D. (1973), 'Radical Interpretation', *Dialectica* **27**, 214–228.

Divers, J. & Miller, A. (1999), 'Mathematical Platonism: Reliability and Judgement-Dependence', *Philosophical Studies* **95**(3), 277–310.

Dogramaci, S. (2010), 'Knowledge of Validity', *Noûs* **44**(3), 403–432.

Dretske, F. (2000), 'Entitlement: Epistemic Rights without Epistemic Duties?', *Philosophy and Phenomenological Research* **60**(3), 591–606.

Dummett, M. (1973), *Frege: Philosophy of Language*, Harvard University Press.

Dummett, M. (1991), *Frege: Philosophy of Mathematics*, Harvard University Press.

Ebert, P. (2005*a*), The Context Principle and Implicit Definitions, PhD thesis, School of Philosophical, Anthropological and Film Studies, University of St Andrews.

Ebert, P. (2005*b*), 'Transmission of warrant-failure and the notion of epistemic analyticity', *Australasian Journal of Philosophy* **83**(4), 505–521.

Ebert, P. (2011), A Framework for Implicit Definitions. Draft. Forthcoming in: Abstractionism. Oxford University Press, 2011.

Ebert, P. & Shapiro, S. (2009), 'The Good, the Bad and the Ugly', *Synthese* **170**(3), 415–441.

Eklund, M. (2006), 'Neo-Fregean Ontology', *Philosophical Perspectives* **20**(1), 95–121.

Enoch, D. & Schechter, J. (2008), 'How Are Basic Belief-Forming Methods Justified?', *Philosophy and Phenomenological Research* **76**(3), 547–579.

Evans, G. (1982), *The Varieties of Reference*, Oxford University Press.

Feferman, S. (1960), 'Arithmetization of metamathematics in a general setting', *Fundamenta Mathematicae* **49**, 35–92.

Feferman, S. e. a., ed. (1990), *Kurt Gödel: Collected works, Volume II: Publications 1938-1974*, Oxford University Press.

Field, H. (1980), *Science Without Numbers*, Princeton University Press.

Field, H. (1982), 'Realism and anti-realism about mathematics', *Philosophical Topics* **13**, 45–69. Reprinted in: Field 1989.

Field, H. (1984*a*), Can We Dispense with Space-Time?, *in* 'PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association', Vol. 2, pp. 33–90.

Field, H. (1984*b*), 'Critical Notice of Crispin Wright: "Frege's Conception"', *Canadian Journal of Philosophy* **14**(4), 637–662.

Field, H. (1984*c*), 'Is Mathematical Knowledge Just Logical Knowledge?', *The Philosophical Review* **93**(4), 509–552.

Field, H. (1985), 'On Conservatives and Incompleteness', *The Journal of Philosophy* **82**(5), 239–260.

Field, H. (1989), *Realism, Mathematics, and Modality*, Blackwell.

Field, H. (2005), Recent Debates about the Apriori, *in* T. Szabo, ed., 'Oxford Studies in Epistemology', Vol. 1, Oxford University Press.

Fine, K. (2002), *The Limits of Abstraction*, Oxford University Press.

Frege, G. (1884), *Die Grundlagen der Arithmetik. Eine logisch mathematische Untersuchung über den Begriff der Zahl*, Meiner Verlag.

Frege, G. (1893), *Grundgesetze der Arithmetik*, Vol. 1, Verlag Hermann Pohle.

Frege, G. (1979*a*), *Gottlob Frege: Posthumous Writings*, Basil Blackwell.

Frege, G. (1979*b*), Numbers and Arithmetic. In: Frege 1979*a*, pp. 275-277.

Frege, G. (1979*c*), Sources of Knowledge of Mathematics and the mathematical natural Sciences. In Frege 1979*a*, pp. 267-274.

Frege, G. (1980), *Philosophical and Mathematical Correspondence*, Basil Blackwell.

Gabbay, D. M. (2007), We can be in harmony with Classical Logic. Available online: http://www.kcl.ac.uk/artshums/depts/philosophy/people/staff/associates/research/gabbay/files/WeCanBeinHarmonywithClassicalLogic.pdf.

Gödel, K. (1964), Cantor's continuum problem: revised edition, *in* P. H. Benacerraf, P., ed., 'Philosophy of Mathematics', Englewood cliffs. Reprinted in: Feferman 1990, pp. 254-270.

Gentzen, G. (1934), 'Untersuchungen über das logische Schliessen', *Mathematische Zeitschrift* **39**, 405–435.

Gettier, E. L. (1963), 'Is Justified True Belief knowledge?', *Analysis* **23**, 121–123.

Goldman, A. (1967), 'A Causal Theory of Knowing', *The Journal of Philosophy* **64**(12), 357–372.

Goldman, A. (1979), What Is Justified Belief?, *in* G. S. Pappas, ed., 'Justification and Knowledge', G. Reidel Publishing Company, pp. 1–23.

Goldman, A. (2009), 'Internalism, Externalism, and the Architecture of Justification', *Journal of Philosophy* **106**(6), 309–338.

Gupta, A. (2009), Definitions, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy'.
URL: *http://plato.stanford.edu/archives/spr2009/entries/definitions/*

Hale, B. (1994*a*), 'Is Platonism Epistemologically Bankrupt?', *The Philosophical Review* **103**(2), 299–325.

Hale, B. (1994*b*), Singular Terms (2), *in* O. McGuinness, Brian Gianluigi, ed., 'The Philosophy of Michael Dummett', Kluwer Academic Publishers, pp. 17–44.

Hale, B. (1996), Singular Terms (1), *in* M. Schirn, ed., 'Frege: Importance and Legacy', Walter de Gruyter & Co, pp. 438–357.

Hale, B. (1997), 'Grundlagen §64', *Proceedings of the Aristotelian Society* **97**, 243–261. Reprinted in: Hale & Wright 2001*a*, pp. 91-116.

Hale, B. & Wright, C. (2000), Implicit Definition and the A Priori, *in* C. Boghossian, Paul Peacocke, ed., 'New Essays on the A Priori', Clarendon Press, pp. 286–319.

Hale, B. & Wright, C. (2001*a*), *The Reasons Proper Study. Essays towards a Neo-Fregean Philosophy of Mathematics*, Clarendon Press.

Hale, B. & Wright, C. (2001*b*), To Bury Caesar..., *in* 'The Reasons Proper Study', Clarendon Press.

Hale, B. & Wright, C. (2005), Logicim in the twenty-first century, *in* 'The Oxford Handbook of Philosophy of Mathematics and Logic', Oxford Scholarship Online Monographs.

Hale, B. & Wright, C. (2007), 'Focus restored: Comments on John MacFarlane', *Synthese* **170**(3), 457–482.

Hale, B. & Wright, C. (2009), The Metaontology of Abstraction, *in* 'Metametaphysics', Oxford University Press.

Handfield, T. & Bird, A. (2008), 'Dispositions, Rules and Finks', *Philosophical Studies* **140**(2), 285–298.

Heck, R. G. J. (1992), 'On The Consistency of Second-Order Contextual Definitions', *Noûs* **26**(4), 491–294.

Heck, R. G. J. (1996), 'The Consistency of Predicative Fragments of Frege's Grundgesetze der Arithmetik', *History and Philosophy of Logic* **17**, 209–220.

Heck, R. G. J. (1997*a*), 'Finitude and Hume's Principle', *Journal of Philosophical Logic* **26**(6), 589–617.

Heck, R. G. J. (1997*b*), The Julius Caesar Objection, *in* 'Language, Thought and Logic: Essays in Honour of Michael Dummett', Oxford University Press.

Heck, R. G. J. (2000), 'Cardinality, Counting, And Equinumerosity', *Notre Dame Journal of Formal Logic* **41**(3), 187–209.

Heck, R. G. J. (2005), 'Julius Caesar and Basic Law V', *Dialectica* **59**(2), 161–178.

Heck, R. G. J. (2006), Reason and Language, *in* C. MacDonald & G. F. MacDonald, eds, 'McDowell and his Critics', Wiley-Blackwell, pp. 22–49.

Hellman, G. (1989), *Mathematics without Numbers*, Clarendon Press.

Hilbert, D. (1903), *Grundlagen der Geometrie*, Teubner.

Hjortland, O. T. (2009), The Structure of Logical Consequence: Proof-Theoretic Conceptions, PhD thesis, School of Philosophical, Anthropological and Film Studies: University of St Andrews.

Holton, R. (2003), 'David Lewis's Philosophy of Language', *Mind and Language* **18**(3), 283–295.

Horwich, P. (1998), *Meaning*, Oxford University Press.

Jenkins, C. S. (2007), 'Entitlement and Rationality', *Synthese* **157**, 25–45.

Jenkins, C. S. (2008*a*), 'Boghossian and Epistemic Analyticity', *The Croatian Journal of Philosophy* **8**(22), 113–127. Available online at: http://www.st-andrews.ac.uk/ csj6/details/BoghossianFinal.pdf.

Jenkins, C. S. (2008*b*), *Grounding Concepts: An Empirical Basis for Arithmetical Knowledge*, Oxford University Press.

Leite, A. (2005), 'Epistemological Externalism and the Project of Traditional Epistemology', *Philosophy and Phenomenological Research* **70**(3), 505–533.

Lewis, D. (1970), 'How to Define Theoretical Terms', *The Journal of Philosophy* **67**(13), 427–446.

Lewis, D. (1974), 'Radical interpretation', *Synthese* **27**, 331–344.

Lewis, D. (1984), 'Putnam's Paradox', *Australasian Journal of Philosophy* **62**(3), 221–236.

Linnebo, Ø. (2004), 'Predicative fragments of Frege Arithmetic', *Predicative fragments of Frege Arithmetic* **10**(2), 153–174.

Linnebo, Ø. (2007), Bad Company Tamed. To appear in Synthese.

Linnebo, Ø. & Uzquiano, G. (2009), 'Which Abstraction Principles Are Acceptable? Some Limitative Results.', *British Journal for the Philosophy of Science* **60**(2), 239–252.

MacBride, F. (1999), 'A Study of Fieldian Nominalism', *British Journal for the Philosophy of Science* **50**, 431–455.

MacBride, F. (2003), 'Speaking with Shadows: A Study of Neo-Logicism', *British Journal for the Philosophy of Science* **54**, 103–163.

McGee, V. (1985), 'A counterexample to Modus Ponens', *The Journal of Philosophy* **82**(9), 462–471.

Melia, J. (2000), 'Weaseling Away the Indispensability Argument', *Mind* **109**(435), 455–479.

Moore, G. (1959), Proof of an External World, *in* 'Philosophical Papers', Allen and Unwin, pp. 127–150.

Nelson, E. (2011), Elements. Unpublished draft.

Nozick, R. (1988), Knowledge and Scepticism, *in* 'Perceptual Knowledge', Oxford University Press.

Parsons, C. (1965), Frege's Theory of Number, *in* M. Black, ed., 'Philosophy in America', Cornell University Press, pp. 180–203.

Parsons, C. (1990), 'The Structuralist View of Mathematical Objects', *Synthese* **84**, 303–346.

Parsons, C. (1995), 'Platonism and Mathematical Intuition In Kurt Gödel's Thought', *The Bulletin of Symbolic Logic* **1**(1), 44–74.

Peacocke, C. (2000), Explaining the A Priori: The Programme of Moderate Rationalism, *in* 'New Essays on the A Priori', Clarendon Press, pp. 255–285.

Pedersen, N. J. L. (2006), Mathematical realists have Moore to learn from scepticism than you might think. Unpublished draft.

Pedersen, N. J. L. (2007), Mathematical scepticism and entitlement. Unpublished draft.

Pedersen, N. J. L. (2009*a*), 'Entitlement, value and rationality', *Synthese* **171**(3), 443–457.

Pedersen, N. J. L. (2009*b*), Hume's Principle and entitlement. Unpublished draft. Forthcoming in: Ebert, P. and Rossberg, M.: Abstractionism. Oxford University Press.

Prior, A. (1960), 'The Runabout Inference-Ticket', *Analysis* **21**(2), 38–39.

Pritchard, D. (2005), *Epistemic Luck*, Oxford University Press.

Pryor, J. (2004), 'What is wrong with Moore's Argument?', *Philosophical Issues* **14**, 349–378.

Putnam, H. (1979), What is Mathematical Truth?, *in* 'Mathematics, Matter, and Method: Philosophical Papers', 2 edn, Vol. 1, Cambridge University Press, pp. 60–78.

Putnam, H. (1981), *Reason, Truth, and History*, Cambridge University Press.

Quine, W. V. O. (1951), 'Two Dogmas of Empiricism', *The Philosophical Review* **60**, 20–43.

Quine, W. V. O. (1953), On What There Is, *in* 'From a Logical Point of View', Harvard University Press, pp. 1–19.

Rosen, G. (1993), 'The Refutation of Nominalism', *Philosophical Topics* **21**, 149–186.

Russell, B. & Whitehead, A. N. (1910-1913), *Principia Mathematica*, Cambridge University Press.

Russell, G. (2008), *Truth in virtue of meaning*, Oxford University Press.

Schirn, M. (2003), 'Fregean Abstraction, Referential Indeterminacy and the Logical Foundations of Arithmetic', *Erkenntnis* **59**, 203–232.

Shapiro, S. (1983), 'Conservativeness and Incompleteness', *The Journal of Philosophy* **80**(9), 521–531.

Shapiro, S. (1991), *Foundations without Foundationalism*, Oxford University Press.

Silins, N. (2005), 'Transmission Failure Failure', *Philosophical Studies* **126**(1), 71–102.

Sosa, E. (2007), *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume 1*, Oxford University Press.

Tarski, A. (1935), 'Der Wahrheitsbegriff in den formalen Sprachen', *Studia Philosophica* **1**, 261–405.

Uzquiano, G. (2009), 'Bad company generalized', *Synthese* **170**(3), 331–347.

Wagner, S. (1981), 'Tonk', *Notre Dame Journal of Formal Logic* **22**(4), 289–300.

Walsh, S. (2010), Logicism, Interpretability & Knowledge of Arithmetic. Draft.

Weir, A. (2003), 'Neo-Fregeanism: An Embarrassment of Riches', *Notre Dame Journal of Formal Logic* **44**(1), 13–48.

Williams, J. (2007), 'Eligibility and Inscrutability', *Philosophical Review* **116**(3), 361–399.

Williams, J. (2008), 'The Price of Inscrutability', *Noûs* **42**(4), 600–641.

Williams, J. (2012), Davidson on reference. Unpublished typescript.

Williamson, T. (2000), *Knowledge and Its Limits*, Oxford University Press.

Williamson, T. (2007), *The Philosophy of Philosophy*, Wiley-Blackwell.

Wright, C. (1983), *Frege's Conception of Numbers as Objects*, Aberdeen University Press.

Wright, C. (1990), Field and Fregean Platonism, *in* A. Irvine, ed., 'Physicalism in Mathematics', Kluwer Academic Publishers, pp. 73–93.

Wright, C. (1991), 'Scepticism and Dreaming: Imploding the Demon', *Mind* **100**, 87–116.

Wright, C. (1999), 'Is Hume's Principle Analytic?', *Notre Dame Journal of Formal Logic* **40**(1), 6–30. Reprinted in: Hale & Wright 2001*a*, pp. 307-332.

Wright, C. (2000), 'Neo-Fregean foundations for real analysis: some reflections on Frege's Constraint', *Notre Dame Journal of Formal Logic* **41**, 317–334.

Wright, C. (2001), 'On Basic Logical Knowledge', *Philosophical Studies* **106**, 41–85.

Wright, C. (2003), Some Reflections on the Acquisition of Warrant by Inference, *in* 'New Essays on Semantic Externalism, Skepticism and Self-Knowledge', MIT Press, pp. 57–77.

Wright, C. (2004*a*), 'Intuition, Entitlement, and the Epistemology of Basic Logical Laws', *Dialectica* **58**(1), 155–175.

Wright, C. (2004*b*), 'Warrant for Nothing (And Foundations for Free?)', *Aristotelian Society Supplementary Volume* **78**(1), 167–212.

Wright, C. (2007*a*), On Quantifying into Predicate Position: Steps towards a New(tralist) Perspective, *in* M. Potter, ed., 'Mathematical Knowledge', Oxford University Press.

Wright, C. (2007*b*), The Perils of Dogmatism, *in* S. Nuccetelli & G. Seay, eds, 'Themes from G.E. Moore: New Essays in Epistemology and Ethics', Oxford University Press, pp. 25–48.

Wright, C. (2008), 'Internal-External: Doxastic Norms and the Defusion of Sceptical Paradox', *Journal of Philosophy* **105**(9), 501–517.

Wright, C. (2009), Abstraction and Epistemic Entitlement. Typescript.