E tu, Brute?
The Varieties of Caesar Problems

MSc Thesis *(Afstudeerscriptie)*

written by

Paul Talma
(born January 10th, 1997 in Paris, France)

under the supervision of **Dr. Thomas Schindler**, and submitted to the Examinations Board in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam.*

Date of the public defense:	**Members of the Thesis Committee:**
August 20th, 2024	Dr. Thomas Schindler
	Dr. Giorgio Sbardolini
	Dr. Jeroen Smid



Institute for Logic, Language and Computation

Abstract

This thesis discusses the Caesar Problem for Neo-Logicist philosophies of mathematics. After developing the philosophical and technical background for Neo-Logicism in chapter 1, I articulate three versions of the Caesar Problem in chapter 2 and critically examine two lines of response to the problem in chapter 3. I argue that the Caesar Problem splinters into an epistemic problem, a problem of semantic indeterminacy, and a problem of linguistic understanding. I also argue that the third problem is the most stringent one. I consider the most influential line of response to the problem, developed by Hale and Wright, and argue that while it resolves the problem of semantic indeterminacy satisfactorily, it offers an over-intellectualized response to the other two problems. I also consider a recent line of response to the Caesar Problem, which I dub indeterminism. Although indeterminism is promising, I highlight a number of areas which need to be developed before the solution can be considered satisfactory.

Acknowledgements

My deepest thanks go to Professor Thomas Schindler and to Jiarui Qu. This thesis was in many ways difficult to write. Without Thomas's relentless encouragement and apparently inexhaustible patience, it would have been impossible. Time after time, when it seemed to me that I had reached a philosophical or motivational dead-end, your generosity with my ideas showed me the way out. I am also lucky to have had the opportunity to learn from your deft touch in combining logical rigor and philosophical insight.

Jiarui was an ever-insightful partner in thought throughout the writing process, spurring refinements and suggesting new avenues to explore at every turn. I have lost count of the times your offhand remarks turned out, upon reflection, to contain the seeds of a powerful objection or a suggestive new perspective. My attraction to philosophy has always been mediated by philosophers capable of plucking the hardest, most abstract problems out of the third realm and turning them into issues of pressing human concern. No one has been more inspiring in this respect than you.

I would also like to thank my dear friends in and around the Master of Logic program for being endless sources of intellectual stimulation: Rodrigo Almeida, Orestis Dimou Belegratis, Josef Doyle, Merlin Krezmien, Ferréol Lavaud, Dean McHugh, Davide Perinti, Alyssa Renata, Fatima Scha, Jonas van der Schaaf, Lingyuan Ye, and Morgane Zeigler. I hope to continue learning from you for years to come.

I am also deeply grateful to Theodore Becker-Jacob, Iain Espey, Julia Goolsby, Emma Heath, Sebastian Martinez-Sanchez, and Jose Antonio de los Heros Montari for our continuing friendship. You each continue to inspire me in your pursuit of the good, the true, and the beautiful.

Finally, I want to thank my parents, Valéry and Véronique Talma, for their love and support—I am the lucky and grateful recipient of plenty of both.

Contents

Acknowledgements **i**

Introduction **1**

1 Philosophical Background **4**
1.1 Neo-Logicism 4
1.2 The Context Principle 8
1.3 Hume's Principle 13
1.3.1 The content of the principle 13
1.3.2 The significance of the principle 18

2 The Caesar Problem **30**
2.1 Some logical facts 35
2.2 The epistemic problem 37
2.3 The semantic problem: indeterminacy 39
2.4 The semantic problem: linguistic understanding 44
2.5 Summary 49

3 Solving the Caesar Problem **51**
3.1 Hale and Wright 51
3.1.1 Sortals 52
3.1.2 The Sortal Inclusion Principle 54
3.1.3 Category Ontology 59
3.1.4 Evaluation 63
3.2 Indeterminism 71
3.3 Summary 78

Conclusion **80**

Works Cited **82**

Introduction

This thesis is concerned with a problem faced by Neo-Logicism. Neo-Logicism is one the dominant philosophical accounts of our knowledge of arithmetic. According to Neo-Logicism, our knowledge of arithmetic is based on knowledge of Hume's Principle. Hume's Principle states that, for two concepts F and G, the number of Fs is the same as the number of Gs if and only if the Fs and Gs can be put into one-to-one correspondence: to each F we can assign a G in such a way that, when we are done, each G is assigned to some unique F. Remarkably, it is possible to prove just about any theorem of number theory on the basis of Hume's Principle. This result—known as Frege's Theorem, after the founder of Logicism, Gottlob Frege—suggests that Hume's Principle could serve as a foundation for arithmetical knowledge. But it raises the question, how is Hume's Principle itself known? The Neo-Logicist answer is that Hume's Principle defines the concept of number: numbers *just are* what two concepts have in common if they can be put into one-to-one correspondence.

This package of views holds great promise: if true, it would explain what might otherwise seem mysterious, namely that we have rich and secure knowledge of a realm of abstract objects—the numbers—entirely inaccessible to our senses. But while Frege's Theorem is a piece of mathematics, and hence not open to serious doubt, the claim that Hume's Principle is a definition of the concept of number faces a serious objection—one as old as Logicism itself. According to this objection, Hume's Principle cannot serve as a definition of number, for it does not set sharp boundaries for the concept. In particular, as Frege pithily observed, Hume's Principle does not settle whether Julius Caesar is a number. And, to put it mildly, if your definition of number does not rule out Roman emperors, you might reasonably wonder whether you have the right definition. If Hume's Principle does not

provide a definition of the concept of number, however, Frege's Theorem is robbed of its philosophical significance. Thus, the Caesar Problem—as it has come to be known—strikes at the very root of Neo-Logicism.

This thesis seeks to advance our understanding of the Caesar Problem. Its main contributions are, first, a careful taxonomy of the varieties of the Caesar Problem, and second, a critical examination of extant solutions to it.

The first chapter lays down the philosophical and technical preliminaries for tackling the Caesar Problem. This includes an overview of the Neo-Logicist approach to the foundations of arithmetic, a discussion of the proper interpretation of Hume's Principle within the Neo-Logicist framework, and an overview of the role player by the principle in the Neo-Logicist account of our knowledge of arithmetic.

The second chapter focuses on the Caesar Problem. After laying down some of the logical results upon which the problem depends, it develops three distinct versions of the problem. On the first version, the problem is that Hume's Principle is inadequate to explain the extent of our knowledge of numbers: there are things we know concerning numbers, such as that Caesar is not one, which the principle leaves unexplained. On the second version, the problem is that the concept defined by Hume's Principle is indeterminate. Since our concept of number is presumed to be determinate, this problem casts doubt on the claim that Hume's Principle is an adequate definition of our concept of number. On the third version, the problem is that Hume's Principle fails to meet a basic condition which a definition must meet to secure a grasp of a concept: it fails to convey what it takes for an object to fall under that concept. Although each version of the problem merits a response, I argue that the third version of the problem is the most severe: while the first two allow that Hume's Principle conveys an understanding of an incomplete or indeterminate concept, the third purports to show that Hume's Principle fails to convey *any* concept at all.

The third chapter critically surveys two of the most prominent solutions to the Caesar Problem. According to the first solution, Hume's Principle operates against the backdrop of a distinctive ontological picture, on which the world is carved up into distinct categories, such that each object belongs to exactly one category. Given a few assumptions about these categories, we can show that the numbers are distinct from the Roman emperors, and indeed from any kind of

object meeting certain weak conditions. According to the second solution, Hume's Principle requires supplementation by piecemeal stipulations: if Caesar is not to be a number, this must be stipulated directly. And likewise for all other questions of identity left undecided by Hume's Principle. It is possible to *stipulate* that Caesar is not a number because—according to this solution—there is no fact of the matter antecedent to our stipulation.

I evaluate each solution against the three versions of the Caesar Problem laid out in chapter 2. I do not conclude in favor of one or the other response, but simply highlight the strengths and weaknesses of each, pointing out where further work must be done to attain a satisfactory solution.

Chapter 1

Philosophical Background

1.1 Neo-Logicism

Neo-Logicism is a contemporary school in the philosophy of mathematics that takes inspiration from Gottlob Frege's Logicism. The latter was Frege's lifelong project of establishing that mathematical truths are ultimately logical truths. It is difficult to overstate the importance of this project to the history of analytic philosophy—it was endorsed, in some form or other, by Bertrand Russell, Rudolf Carnap, and the early Wittgenstein; that is, by the founding figures of the discipline—or to overstate its centrality to Frege's philosophy. As Richard Kimberly Heck has put it, "Almost all of [Frege's] intellectual work is devoted, in one way or another, to the attempt to establish logicism" (Heck 2011c: 1).

Frege's Logicism is characterized by two theses (Hale and Wright 2001b: 1): that numbers are objects existing independently of us, and that truths about numbers can be known on the basis of logical reflection alone. That numbers are objects is not meant to imply that they are much like Austin's middle-sized dried goods. The notion of object at play here is rather a *logical* one. Objects, in this sense, are intimately tied to our referential apparatus: to singular terms, quantifiers, and identity contexts (Parsons 2008, 1982).[1] So long as numbers can be quantified

[1]Cf. Linnebo (2018b §2.2), who distinguishes between a Fregean conception of logical object, on which objects are possible referents of singular terms, and a Quinean one, on which they are possible values of bound variables. Linnebo suggests that the Quinean criterion is broader: there may be objects over which it would seem possible to quantify, but which could not in principle be named. As an example, Linnebo cites the imaginary number i: mathematicians routinely quantify over it when they do complex analysis. However (argues Linnebo), there are reasons

over, referred to using singular terms, or meaningfully be said to be identical or distinct from one another, they count as objects in this logical sense. In particular, it is no bar to something's being an object in this sense that it is abstract.

Logicism also maintains that numbers are *objective*: not dependent on any conscious or representational activity for their being or (intrinsic) properties. That is, numbers exist, and have the properties they have, independently of what anyone or anything may think, feel, or say about them. It is of course non-trivial to characterize the kind of independence at issue. But it should be reasonably clear what kind of distinction we have in mind.

The second thesis characteristic of Logicism is that mathematical truths—specifically, truths of number—can be known on the basis of logical reflection upon definitions. Frege called truths knowable on such grounds "analytic" (Frege 1950 §3; Burge 2003: 200).

Why should it matter that mathematical truths be analytic in this sense? There is an obvious epistemic payoff: mathematics would be shown to be no less secure than logic itself. But Frege's logicism was not primarily an answer to skepticism about mathematics: his primary aim was to explain the *nature* of mathematical truths, not to increase his reader's confidence in them. A characteristic expression of this attitude reads as follows:

> The aim of proof is, in fact, not merely to place the truth of a proposition beyond all doubt, but also to afford us insight into the dependence of truths upon one another.
>
> Frege (1950 §2)

Exhibiting such dependencies would reveal "the ultimate ground upon which rests the justification for holding [a mathematical proposition] to be true" Frege (ibid. §3).

In addition to revealing the nature of mathematical truths, a logicist reduction of mathematics would help explain our knowledge of them. Such knowledge is prima facie puzzling. The truths of arithmetic seem to concern a domain of

---

to think that we could not, in principle, have in our language a term that refers to i (contrary to the impression this very sentence may give!). This is because the complex field admits of a non-trivial automorphism that maps i to $-i$. This being the case, it may be argued that there could be no explanation for our referring to i as opposed to $-i$.

abstract objects, the natural numbers. Indeed, this claim is one of the two core theses of Logicism. Yet numbers, being abstract, are utterly removed from our experience: in particular, we cannot interact with them causally, and hence cannot perceive them—they are not given to us in intuition.[2] To explain how knowledge of arithmetic could be grounded in logical reflection upon definitions, we therefore need to explain how logical reflection and definitions could allow us to think about numbers or make reference to them in language.

This is a daunting task. For according to many philosophers past and present (Kant included), logic provides "canons for thought": it systematizes the notion of a good inference, or lays down constitutive rules for thought—rules one must follow if one is to count as thinking at all. According to another prominent (and related) conception, logic is a *formal* discipline. It is concerned with the form of thought rather than with its content. Thus, the rules it lays down are supposed to be maximally general, applicable no matter what one is thinking about. What logic does not do, on either conception, is tell us what objects exist. In particular it does not tell us that numbers exist. Likewise, definitions record or encourage patterns of language use. They are concerned primarily with language, and only derivatively, if at all, with the world beyond it. It is therefore hard to see how logical reflection upon definitions could possibly appraise us of the existence of an immensely complex realm of abstract objects, allow us to refer to them, and ground our extensive knowledge of them.

Frege's own logical system comprised a number of "basic laws" forming what we would now call full second-order logic together with the following "Basic Law V":

$$ext(F) = ext(G) \leftrightarrow \forall x(Fx \leftrightarrow Gx).$$

Basic Law V states that any two concepts F and G have the same extension if and only if they are equivalent—iff the same things fall under F as fall under G. Using Basic Law V, Frege is able to derive the Dedekind-Peano axioms, from which in

[2]The notion of intuition at play here goes back to Kant, where it means something like *singular, immediate representation*. Frege takes up the notion in the *Grundlagen*, explicitly recalling Kant's usage. Although he does not attempt a definition, it is clear that the key feature of intuition—the one that sets the problem—is that nothing can be intuited which does not affect the senses.

turn a large and interesting class of arithmetical truths can be derived (enough to think that most of the interesting arithmetical truths, if discoverable at all, are derivable from the Dedekind-Peano axioms). If Basic Law V is indeed a law of logic, then (at least from a technical point of view), the reduction of arithmetic to logic is a success. Unfortunately, as Bertrand Russell pointed out to Frege as the second volume of his *Grundgesetze* went to press, far from being a logical law, Basic Law V is a logical falsehood. For consider the (conjunctive) concept *being the extension of a concept and not falling under that concept*. Call this concept R and its extension r. Do we have Rr? Suppose we do. Then by the definition of R, r is itself the extension of a concept under which it does not fall. But r is the extension of R. So r does not fall under R: we have $\neg Rr$, a contradiction. Thus we must rather have $\neg Rr$. But then, again by the definition of R, the extension r is *not* the extension of a concept under which it does not fall—that is, is it the extension of a concept under which it does fall. But it is the extension of, and hence falls under, R; that is, Rr, a contradiction.

Frege's Logicism died with the discovery that Basic Law V is inconsistent. It remained dead until the 1980s, when Crispin Wright drew attention to the fact that Frege's derivation of the Dedekind-Peano axioms only essentially relies on Basic Law V at one point: to derive another principle, Hume's Principle, from which the axioms follow.[3] And while Basic Law V is contradictory, Hume's Principle appears to be not only consistent but obviously true. The principle states that, for any concepts F and G, the number of Fs is the same as the number of Gs just in case the Fs and the Gs can be put into one-to-one correspondence. Wright showed that this principle entails, by standard reasoning in second-order logic, the Dedekind-Peano axioms, and hence a wide class of interesting arithmetical truths. This discovery marked the birth of Neo-Logicism.

Wright did not claim that Hume's Principle was a law of logic, as Frege had claimed for Basic Law V. But the principle did not seem like any old truth, either. Wright argued that it constituted an explanation of the concept of number. As

[3]Actually, the story is somewhat more complicated. Peter Geach and Charles Parsons independenly noted that the Dedekind-Peano axioms follow from Hume's Principle before Wright did. But neither attempted to prove this fact. The proof, it turned out, is non-trivial. Moreover, neither Geach nor Parsons drew any philosophical conclusions from this fact. See Heck (2011d) for details on the history of Hume's Principle.

such, it was supposed to enjoy a privileged epistemic status: it could be known simply by reflecting on our concept of number. More importantly, it could be used by someone who lacked the concept of number to acquire that concept. Thus, our cognitive access to numbers and our knowledge of arithmetic could be explained by logical reflection on something like a definition of the number concept. Hume's Principle is not a definition in Frege's sense—it does not allow for the elimination of its definiendum everywhere it occurs. So the letter of Frege's Logicism could not be revived. Nonetheless, Wright claimed, Hume's Principle was a kind of definition—in particular, it enjoys the same epistemic status as *bona fide* definitions. It could thus provide a compelling answer—at last—to Frege's question: "How, then, are numbers given to us, if we cannot have any ideas or intuitions of them?" (Frege 1950 §62).

In the remainder of this chapter, we will introduce the necessary background to appreciate in detail the shape of the Neo-Logicist acount of arithmetical knowledge. In the next chapter, we will introduce and discuss one of the most serious problems facing the Neo-Logicist program: the Caesar Problem. Finally, in the third chapter, we shall critically examine the extant solutions to this problem.

1.2 The Context Principle

The Context Principle plays a central role in Neo-Fregean philosophy of mathematics. Broadly speaking, the principle asserts that the meaning of complete sentences is—in a sense to be clarified presently—prior to that of sub-sentential expressions. This being the case, our understanding of some singular terms might be explained by our understanding of complete sentences in which they occur, and our understanding of *those* might be explained by our understanding of some other complete sentences not containing those singular terms. In this way, we could explain our understanding of number-terms in terms of our understanding of *prima facie* less puzzling expressions. At any rate, so goes the Neo-Logicist strategy for explaining how numbers are given to us.

To flesh out this story, however, we shall need a more precise understanding of the context principle. The principle was originally stated, in varying language, by Frege in the *Grundlagen*:

(i) Never ask for the meaning of a word in isolation, but only in the context of a sentence. (p. xxii)

(ii) Only in a sentence have the words really a meaning. [...] It is enough if the sentence taken as a whole has a sense; it is this that confers on its parts also their content. (§60)[4]

(iii) Since it is only in the context of a sentence that words have any meaning, our problem becomes this: To define the sense of a sentence in which a number word occurs. (§62)

(iv) We next laid down the fundamental principle that we must never try to define the meaning of a word in isolation, but only as it is used in the context of a sentence. (§106)

These are clearly various ways of taking these suggestive remarks. Linnebo and Burge agree in extracting three kinds of lessons from them:

- **A necessary condition:** a word only has a meaning if it occurs, or could occur, in the context of a (meaningful) sentence.
- **A sufficient condition:** if a word occurs (or could occur) in a meaningful sentence, then it has a meaning.
- **A methodological strategy:** it is only possible to explain the meaning of a word in terms of the meanings of sentences in which it occurs.

Note that "meaning" here translates Frege's "Bedeutung." Since Frege wrote the *Grundlagen* prior to developing his distinction between sense (*Sinn*) and denotation (*Bedeutung*), we should countenance two versions of each interpretation: one concerning a word's sense and the other concerning its denotation. Moreover, it is clear that the principles need to be qualified. For example, consider the sufficiency condition, pitched at the level of denotation. The sentence "Santa Claus does not exist" is meaningful, and has a denotation (a truth-value). Yet "Santa Claus" does not thereby refer. One option is to restrict the class of relevant sentences to

[4]Here and below, we follow Linnebo in translating Frege's "Satz" as "sentence" rather than as Austin's "proposition." See Linnebo (2018b 109, note 3) for justification.

true atomic sentences. This will rule out "Santa Claus does not exist" as logically complex. However, it will not rule out (plausibly true) sentences like "Santa Claus is a fictional character." But this sentence also attributes non-existence, in a way; perhaps we should also exclude these from the class of relevant sentences. In fact, as far as the denotation-level necessity and sufficiency criteria are concerned, it seems like we can get away with focusing solely on sentences of the form "t exists" or $\exists x(x = t)$—for t to have a denotation, it is both necessary and sufficient that it figure in a true sentence of this form. But as we will see, there are other classes of sentences which it may be reasonable to take as relevant.

Although there are other interesting exegetical and philosophical questions concerning the first and (especially) the third lesson, it is the second that it most directly relevant to the Neo-Logicist project. As Burge writes, it is the sufficiency condition (at the level of denotation) that "provides a philosophical basis for understanding theoretical reference to abstract entities and other theoretical entities of science" (Burge 2005: 15). This condition "locates ontology in the evaluation of theories. It implicitly rejects philosophical requirements on reference or denotation that go beyond standard methods for evaluating the truth of sentences or theories." (Burge 2005: 16; see also Quine 1948). In particular, it rejects the idea that for a singular term to refer, one has to stand in a certain causal relation to its referent, or be able to describe it uniquely. How does it do so?

Suppose that we were unsure whether some singular term t referred. How might we go about determining whether it does in fact refer? One possibility would be to ask, does "t" refer? To answer this question, we might then inquire into the properties of the expression "t": how was it introduced? How has its meaning evolved over time? Is it associated with a definition? And so on. Alternatively, we could simply ask, does t exist? And to answer that question, we would bring to bear all of our usual means for deciding the existence of things of the kind t purports to denote. If t is "Homer," for example, we might look at historical records, analyze rhythmic patterns in different parts of the *Odyssey*, and the like. If our evidence told in favor of the existence of Homer, we would conclude that Homer exists, and hence that "Homer" refers.

The former strategy involves semantic ascent: our inquiry is turned toward the term "t" itself. The latter strategy, by contrast, keeps its sights on the world

rather than the word—it uses, rather than mentions, the term "t" in determining whether it has a reference.

How does all of this bear on the Neo-Logicist project? The Neo-Logicist wishes to explain how singular reference to—and hence knowledge of—numbers is possible. This requires providing some conditions under which a numerical term, such as "the number of keys on my keyboard" successfully refers, and showing that such conditions obtain. By the context principle, it will suffice to show that a certain class of sentences in which the term occurs have a denotation, i.e. a truth-value. And here, the Neo-Logicist follows Frege, who after recalling his context principle writes that

> our problem becomes this: To define the sense of a sentence in which a number word occurs. That obviously leaves us still a very wide choice. But we have already settled that number words are to be understood as standing for self-subsistent objects. And that is enough to give us a class of propositions which must have a sense, namely those which express our recognition of a number as the same again. If we are to use the symbol a to signify an object, we must have a criterion for deciding in all cases whether b is the same as a, even if it is not always in our power to apply this criterion. In our present case, we have to define the sense of the sentence
>
> "the number which belongs to the concept F is the same as that which belongs to the concept G"
>
> Frege (1950 §62)

Frege is obviously building up to an application of the sufficiency criterion (though there are hints that he also has the necessity criterion in mind). What is the relevant class of sentences which must be shown to have a truth-value, if number words are to refer? Frege appears to reason as follows: if an expression is a singular term purporting to refer to a certain kind of object, then the relevant sentences are identity statements linking this term to other terms for objects of this kind. Thus, to show that number terms refer, it will suffice to show that statements of the form "the number of Fs is the number of Gs" all have a truth-value.

There are several questions to raise here. Is Frege right in his choice of sentences? Why must *all* sentences of this form have a truth-value? Why isn't it enough that for each number term, *some* sentence (of the right form) containing that term has a truth-value? Conversely, why *only* sentences of this form? Why not also require that other identity statements containing a number term have a truth-value? And anyway, why are identity statements the focus in the first place? Why not ask that some other atomic propositions involving number terms have a truth-value?

We cannot give a complete answer to these questions here. We will merely re-emphasize the connection between objecthood and identity that was identified as a core claim of (Neo-)Logicism in the previous section: objects are, by their nature, the kinds of things for which questions of identity and distinctness are well-posed; to conceive of something as an object, it is therefore sufficient to be sensitive to such questions concerning that (kind of) object; this sensitivity is modeled by a grasp (possibly implicit) of criteria of identity for objects of that kind.[5]

Let us summarize the import of the context principle to the question, how are numbers given to us? That question concerns our reference to numbers; it is puzzling, since some of the standard ways of explaining our reference to objects—in particular, causal connections—are unavailable. The context principle suggests another route: it will suffice for reference to numbers if a certain range of sentences containing number terms have a truth-value (for which it suffices that sentences in that range have truth-*conditions*). The connection between objecthood and identity suggests (though we did not conclusively establish) that the relevant range of sentences consists in identity statements between number terms. The task of explaining our grasp of numbers has therefore been transformed into the task of giving truth-conditions for statements of numerical identity. It is to carry out this task that the Neo-Fregean, following a lead which Frege abandons, turns to Hume's Principle.

[5]Linnebo (2018a §2.2) develops the connection between reference to objects and criteria of identity more thoroughly. See also section 3.1.1 below, which discusses the notion of a sortal concept, possession of which requires (according to Wright) grasp of a criterion of identity. If *number* is a sortal concept, this view constitutes another argument for the importance of numerical identity statements.

1.3 Hume's Principle

In this section, we review some of the basic issues surrounding the interpretation and significance of Hume's Principle, so as to ground our later discussion.

Hume's Principle is the (universal closure of the) following statement:

$$Nx : Fx = Nx : Gx \leftrightarrow F \sim G,$$

where $F \sim G$ abbreviates the claim—expressible in pure second-order logic—that the Fs and the Gs can be put into one-to-one correspondence.

1.3.1 The content of the principle

Hume's Principle asserts an equivalence of some sort between its left- and right-hand sides. But just what is this relation of equivalence supposed to be? In standard formulations of the principle, such as the one given above, this equivalence is simply material equivalence. However, Frege's own writings, as well as the Neo-Logicist strategy, suggest that something stronger should be true.

Recall that Frege's aim in the *Grundlagen*, as well as Wright's in *Frege's Conception*, is to explain how numbers could be "given to us" as objects. Moreover, as explained in section 1.1, this problem is to be understood as that of explaining how we could have singular thought about numbers, or how we could understand singular terms for numbers. And as we saw in the previous section, this question is in turn to be answered by stating truth-conditions for statements of numerical identity. Hume's Principle seems conspicuously poised to explain such statements.

But how exactly does Hume's Principle provide statements of numerical identity with truth-conditions? The Fregean and Neo-Fregean approach to this question is suggestively expressed by Frege: in an abstraction principle like Hume's Principle,

> we carve up the content [of the right-hand side] in a way different from the original way, and this yields us a new concept.
>
> Frege (1950 §64)

Thus, to understand the role played by Hume's Principle in the (Neo-)Fregean project, we must come to terms with this notion of "recarving."

It is clear from Frege's own remarks, as well as from the principle's role in the Neo-Fregean project, that what is being recarved are meanings, and that this recarving is supposed to preserve them. For Frege suggests, in discussing a similar principle concerning lines and directions, that

> The sentence
>
> "line a is parallel to line b"
>
> is to mean the same as
>
> "the direction of line a is identical with the direction of line b"
>
> (§65)

But what notion of meaning can support this recarving? Any candidate will have to perform a delicate balancing act. For, on the one hand, the idea that the left-hand side recarves the content of the right-hand side suggests that the two sides share a content which is divided up now one way, now another. But on the other hand, Frege insists that the recarving yields a *new* concept—indeed, this is the whole point. And if a concept is expressed by the left-hand side which is not expressed by the right, then the two sides must have distinct meanings after all.

To fix notation, let us write Hume's Principle as

$$Nx : Fx = Nx : Gx \equiv F \sim G.$$

We can read $\equiv$ as "recarves"; our task is to figure out what it could mean. It will be helpful to draw upon Frege's own distinction between *Sinn* and *Bedeutung* (Frege 1892). The *Bedeutung* of an expression is what we would now call its denotation or semantic value (in a context of use). Traditionally, the *Bedeutung* of a singular term is an object, that of a predicate is a concept or set (although important for some purposes, the difference does not matter in the present context).

Could Hume's Principle assert that $Nx : Fx = Nx : Gx$ and $F \sim G$ have the same *Bedeutung*? Recall that the *Bedeutungen* of sentences, in Frege's system, are truth values—true sentences denote the True, false ones the False. Thus, reading

Hume's Principle as asserting sameness of *Bedeutungen* is tantamount to interpreting $\equiv$ as a classical biconditional. Read this way, Hume's Princple very likely true.[6] But unfortunately, this reading is too weak to support the Neo-Logicist project. For clearly, materially equivalent sentences don't generally recarve one another's meaning—whatever recarving is supposed to mean exactly. Hume's Principle is supposed to be an *explanation* of the truth conditions of numerical identities. But, again, material equivalence is a much weaker relation than explanation.

It is safe to say that, although Hume's Principle should certainly entail the material equivalence of $Nx : Fx = Nx : Gx$ and $F \sim G$, it must go beyond this assertion to play the role asked of it by Neo-Fregeans. Can we then read the principle as asserting sameness of sense?

Fregean senses are notoriously difficult entities to explicate. Frege's own explanations point in two different directions. In the *Grundgesetze*, Frege articulates a strong principle of compositionality for senses:

> If a name is part of the name of a truth value [i.e. of a sentence], then the sense of the name is part of the thought expressed by the latter [i.e. by the sentence].
>
> Frege (2013 §32)

On the usual understanding, the compositionality of sense asserts that the sense of a complete sentence is *determined* by the senses of its component expressions (and by their mode of composition, a qualification I will suppress for brevity; see Westerstahl, Baltag, and Benthem 2021: 774). In the passage just quoted, Frege goes beyond this in suggesting that the sense of a complete sentence is *constituted* by that of its component expressions.[7] If we take this passage to heart, we are left with a very fine-grained notion of content. In particular, it is enough to rule out

[6] Though see Boolos (1997), who argues that Hume's Principle, by assigning a number to the universal concept, contradicts ZFC (in which no such cardinal number exists). It is not clear that there really is a conflict here. But if there is, and if one is inclined to side with ZFC, one ought to conclude that Hume's Principle is false after all, at least in its unrestricted form. See Wright (2001a: 313–316) for discussion.

[7] To see that this conception is stronger, suppose that sentence φ contains an expression with sense s while sentence ψ does not. On the compositionality-as-constitution interpretation, φ and ψ cannot have the same sense. For s is a part of the sense of φ, but not of the sense of ψ. However, on the compositionality-as-determination interpretation, φ and ψ may still share a sense. For it is compatible with that interpretation that different arrangements of different

sentences φ and ψ as having the same meaning if there is an expression occuring in one whose sense is not expressed by any expression ocurring in the other.

Indeed, this notion of sense is too fine-grained to be useful. For Hume's Principle is supposed to introduce us to a new concept: *number*. The expression Nx in the principle's left-hand side is supposed to express something new, something not expressible in the base vocabulary of (we may suppose) pure second-order logic. But then by Frege's criterion, the two sides of the principle *ipso facto* have different senses.

Dummett, in his *Frege: Philosophy of Mathematics*, raises a similar point. He considers a conception of sense on which two sentences have the same sense iff any thinker who understands one thereby understands the other (Dummett 1991: 171–172). Pure synonyms, like "ketchup" and "catsup" provide an illustration. It is not required, of course, that anyone with "ketchup" in their vocabulary also understand the word "catsup." But anyone who understands a sentence containing the word "ketchup"—who grasps the thought it expresses—thereby grasps the thought expressed by the corresponding sentence in which "catsup" everywhere substitutes for "ketchup"—even if they do not know the word "catsup." And as Dummett argues, Hume's Principle does not meet this condition: someone who understands (grasps the thought expressed by) $F \sim G$ need not understand (grasp the thought expressed by) $Nx : Fx = Nx : Gx$ (and vice-versa). At any rate, it is not obvious that anyone in possession of the concept of number is also in possession of the concept of a bijection.[8]

We can, however, extract a different notion of sense from Frege's writings. In "The Thought," Frege appeals to a criterion of identity for senses along the following lines (cf. Frege 1956: 138):

> Expressions a and b have the same sense iff no thinker who understands the two expressions could rationally take a different attitude towards the thoughts expressed by $\varphi(a)$ and $\varphi(b)$, where $\varphi(x)$ is an open formula with free variable x of the same type as a and b.

Dummett, focussing on complete sentences rather than arbitrary expressions, rephrases

sub-sentential senses yield the same sentential sense (i.e. that the function mapping sentences to senses not be injective).

[8]See Heck (2011a) for an elaboration of this point.

this principle thus:

> Sentences φ and ψ have the same sense iff any thinker who understands both knows immediately that they must have the same truth-value.
>
> Dummett (1991: 171)

The key question here is how to understand the kind of immediacy at issue. Bob Hale provides the "official" Neo-Logicist answer (though he calls senses "truth-conditions"):

> Two sentences have the same truth-conditions iff anyone who understands both of them can tell, without determining their truth-values individually, and by reasoning involving only compact entailments, that they have the same truth-value.
>
> Hale (2001a: 102)

The notion of *compact entailment* is meant to explicate Dummett's "immediately." Roughly, $\varphi_0, \varphi_1, \ldots, \varphi_n$ compactly entail ψ if the φ_i entail ψ and contain no "surplus" content irrelevant to the entailment. As a toy illustration, $a > b$ and $b > c$ compactly entail $a > c$, but $a > b$ and $b > c \wedge c > d$ do not.

On this notion of sense, it is arguable that that $Nx : Fx = Nx : Gx$ and $F \sim G$ have the same sense.[9] For there is a case to be made that if someone really understands all of the concepts involved, they can infer back and forth between the two directly. This, at any rate, is Dummett's judgement on the matter: "anyone who understands both $[Nx : Fx = Nx : Gx]$ and $[F \sim G]$, for example, must straightaway recognise them as equivalent" (Dummett 1991: 171).

However, there are reasons to doubt that even this—more promising—construal of sense will support a recarving claim. An example due to Fine (2002) shows

[9] In fact, both Hale (2001a) and Linnebo (2018b) claim that the two sides of an abstraction principle compactly entail one another. But I have not been able to understand why. No abstraction principle is a truth of logic. So the two sides do not logically entail one another. *A fortiori*, they do not compactly entail one another. Hale mentions that in his sense, φ logically entails ψ if it is logically, analytically, or conceptually impossible for φ to be true and ψ to be false. This suggests that we shoud perhaps take Hume's Principle to be part of the background logic (since it is supposed to be a conceptual or analytic truth). Of course, the two sides of an abstraction principle entail one another against the backdrop of the abstraction principle.

that sense-as-truth-conditions is too coarse-grained to support an assignment of reference to the cardinality operator.

Given a concept F, let F^+ be the concept that applies to the Fs and to the least natural number (if any) that is not F. Then consider the following:

(i) $Nx : Fx = Nx : Gx$

(ii) $F \sim G$

(iii) $Nx : F^+x = Nx : G^+x$

(iv) $F^+ \sim G^+$

By Hume's Principle, (i) and (ii) have the same sense, as do (iii) and (iv). But note that (ii) and (iv) also have the same sense: for each follows compactly from the other.[10] Thus, (i) and (iii) have the same sense. But then a modicum of compositionality for senses yields that $Nx : Fx$ and $Nx : F^+x$ have the same sense, and hence the same reference. And yet there is one more object falling under F^+ than under F, so that the number of Fs and the number of F^+s are distinct. Thus, if we adopt Hale's notion of sense, Hume's Principle fails to assign to $Nx : Fx$ its intended meaning as "the number of Fs."

Thus, if the idea that the left-hand side of an abstraction principle recarves the content of its right-hand side is to be defended, we must find a notion of content that is coarser-grained than Frege's strict notion, on which the sense of a sentence is in part constituted by the senses of its component expressions, yet finer-grained than Hale's sense-as-truth-conditions. Unfortunately, no such notion seems forthcoming.

1.3.2 The significance of the principle

As we saw in the previous section, Hume's Principle is a key ingredient in the Neo-Logicist account of how we refer to numbers. But the Neo-Logicist wishes not only to explain our reference to numbers, but also our knowledge of arithmetical truths. Here too, Hume's Principle is of central importance. This importance

[10] That they follow from one another is clear. That they follow compactly is easily established given the formal definition of compact entailment, which need not detain us here.

stems from two sources. The first is what has come to be known as Frege's Theorem.[11] Roughly, Frege's Theorem states that pure second-order logic plus Hume's Principle interprets second-order Peano Arithmetic. The second is the (alleged) epistemic status of Hume's Principle. Hume's Principle is supposed to be analytic, or otherwise in excellent epistemic standing. Neo-Logicists take these two features of Hume's Principe to combine into an explanation of our arithmetical knowledge. Let us take a closer look at each step.

Frege's Theorem

In this section, we give a sktech of Frege's Theorem and discuss a related philosophical question. Our presentation of the theorem draws from those of Heck (2011f), Heck (2011b), and Wright (1983).

The technical core of (Neo-)Logicism is a proof that arithmetical truths follow from Hume's Principle. This result is achieved, not by proving each arithmetical truth from Hume's Principle piecemeal, but by showing that a small set of fundamental truths about the natural numbers can be proven from Hume's Principle. These truths are known as the Dedeking-Peano axioms for arithmetic, and are as follows:

1. $\mathbb{N}0$
2. $\forall x \forall y(\mathbb{N}x \wedge Pxy \rightarrow \mathbb{N}y)$
3. $\forall x \forall y \forall z(\mathbb{N}x \wedge Pxy \wedge Pxz \rightarrow y = z)$
4. $\forall x \forall y \forall z(\mathbb{N}x \wedge \mathbb{N}y \wedge Pxz \wedge Pyz \rightarrow x = y)$
5. $\neg \exists x(\mathbb{N}x \wedge Px0)$
6. $\forall x(\mathbb{N}x \rightarrow \exists y(Pxy))$
7. $\forall F(F0 \wedge \forall x \forall y(\mathbb{N}x \wedge Fx \wedge Pxy \rightarrow Fy) \rightarrow \forall x(\mathbb{N}x \rightarrow Fx))$

[11] See Heck (2011d) for the history of the name.

Pxy is read as "x precedes y." $\mathbb{N}x$ is read as "x is a (natural) number." We call this system PA2.[12]

The derivation of PA2 is to take place in the theory whose sole non-logical axiom is Hume's Principle. We call this theory Frege Arithmetic, abbreviated FA.

Note that both PA2 and FA have full-second order logic as the background logic. In particular, their logical axioms include the comprehension axioms:

$$\exists F \forall x (Fx \leftrightarrow \varphi(x))$$

for any formula φ not containing F free.[13]

We also recall the definition of a (proof-theoretic) interpretation:

Definition 1.3.1 (Interpretation of a language)**.** Let $\mathcal{L}$ be a language and T^* an $\mathcal{L}^*$-theory. An interpretation I of $\mathcal{L}$ into T^* is given by:

(i) an $\mathcal{L}^*$-formula $D^I(x)$ such that $T^* \vdash \exists x D^I(x)$

(ii) for every $\mathcal{L}$-constant symbol c, an $\mathcal{L}^*$-formula $C^I(x)$ such that $T^* \vdash \exists! x (C^I(x) \wedge D^I(x))$

(iii) for every n-place $\mathcal{L}$-relation symbol R, an n-place $\mathcal{L}^*$-formula $R^I(\overline{x})$ such that $T^* \vdash \forall \overline{x} (R^I(\overline{x}) \rightarrow \bigwedge_{i=1}^n D^I(x_i))$

(iv) for every n-place $\mathcal{L}$-function symbol f, an $n+1$-place $\mathcal{L}^*$-formula $F^I(\overline{x}, y)$ such that both $T^* \vdash \forall \overline{x} \forall y (F^I(\overline{x}, y) \rightarrow (\bigwedge_{i=1}^n D^I(x_i) \wedge D^I(y)))$, and $T^* \vdash \forall \overline{x} (\bigwedge_{x=1}^n D^I(x_i) \exists! y (F^I(\overline{x}, y) \wedge D^I(y)))$

We define a map $\varphi \mapsto \varphi^i$ of $\mathcal{L}$-formulas to $\mathcal{L}^*$-formulas as follows for the atomic clauses:

$$(x = y)^I := x = y \qquad (x = c)^I := C^I(x)$$
$$(R(\overline{x}))^I := R^I(\overline{x}) \quad (f(\overline{x}) = y)^I := F^I(\overline{x}, y)$$

[12] The conventional name for these axioms is Peano Arithmetic after their original publisher, even though Dedekind is credited with their original formulation. See Dedekind (1963) and Peano (1967).

[13] For a full treatment of the deductive system for second-order logic, see for example Button and Walsh (2018 §1.C). For background on the technical and philosophical issues surrounding second-order logic, see Shapiro (1991).

I is required to commute with the sentential connectives, and quantifiers are relativized to D^I (so that $(\forall x\varphi(x))^I := \forall x(D^I(x) \to \varphi^I(x))$).

Definition 1.3.2 (Interpretation of a theory)**.** Let T be an $\mathcal{L}$-theory and T^* be an $\mathcal{L}^*$ theory. Then T is interpretable in T^* iff there is an interpretation I of $\mathcal{L}$ into T^* such that, for all $\mathcal{L}$-sentences φ, if $T \vdash \varphi$, then $T^* \vdash \varphi^I$.

Intuitively, the function I provides a "translation" of the language $\mathcal{L}$ of T into that of T^*. To interpret T, T^* must first provide definitions for the vocabulary of T (this is the point of (i)–(iv) above), which determines the translation of the atomic sentences of $\mathcal{L}$. Logically complex sentences are required to be translated compositionally. Fnially, we require that theorems of T be translated as theorems of T^*.

In practice, we interpret T_1 in T_0 by defining the basic vocabulary of T_1 in T_0 and then proving, in T_0, the translation of the axioms of T_1.

We can now state and prove

Theorem 1.3.3 (Frege's Theorem)**.** FA interprets PA2. Moreover, Frege's own definition of number, successor, and 0 induces such an interpretation.

Readers familiar with the proof may wish to skim the next few pages.

Proof. We will sketch the high points of the proof, leaving some of the minor lemmas stated but unproved.

Let us start with Frege's definitions of the basic vocabulary of PA2. Here is the definition of zero:

$$0 =_{\text{def}} Nx : x \neq x$$

and his definition of predecessor:

$$Pmn =_{\text{def}} \exists F \exists x(Fx \wedge n = Ny : Fy \wedge m = Ny : (Fy \wedge x \neq y))$$

In words, m precedes n if n is the number of Fs, and for some F thing x, m is the number of Fs that aren't x. (Although we have given a definition of the predecessor relation, we shall sometimes talk of the successor relation, with the obvious intended interpretation.)

It remains to define *natural number*. This is somewhat more involved. The definition makes use of the notion of the *ancestral* of a relation. Call a concept F *hereditary* with respect to a relation R if whenever x is F, anything in the R-image of x is also F; that is,

$$\forall x(Fx \rightarrow \forall y(Rxy \rightarrow Fy))$$

We then define the strong ancestral of R as follows:

$$R^*ab \leftrightarrow_{\text{def}} \forall F\big(\forall x(Rax \rightarrow Fx) \wedge \forall x\forall y(Fx \wedge Rxy \rightarrow Fy) \rightarrow Fb\big)$$

In words, a stands in the strong ancestral of R to b iff b falls under every concept F which is such that (i) every R-successor of a falls under F and (ii) F is hereditary with respect to R. If R^*ab, we shall say that b is in the R-series beginning with a.

Finally, we define the *weak ancestral* of R by

$$R^{*=}ab \leftrightarrow_{\text{def}} R^*ab \vee a = b$$

Intuitively, a stands in the (weak) ancestral of R to b if there is an R-chain from a to b.

We are now in a position to define *natural number*:

$$\mathbb{N}x \leftrightarrow_{\text{def}} P^{*=}0x.$$

With these definitions on the table, we can begin to derive some of the axioms of PA2 in FA. For ease of reference, we will recopy the axioms before proving them.

Axiom 1: $\mathbb{N}0$.

Axiom 1 follows immediately from the definition of zero and of ancestral, since we have $P^{*=}00$.

Axiom 2: $\forall x\forall y(\mathbb{N}x \wedge Pxy \rightarrow \mathbb{N}y)$.

For axiom 2, we must show that the successor of any number is itself a number.[14] So suppose $\mathbb{N}n$ and Pnm. By the definition of $\mathbb{N}$, from $\mathbb{N}n$ we have $P^{*=}0n$. Now, it is an easy lemma concerning the ancestral of a relation R that if $R^{*=}xy$ and

[14]Note that we haven't yet justified the use of the definite article for successor. But we will show later that successors exist and are unique.

Ryz, then $R^{*=}xz$. Hence we have that $P^{*=}0m$, and thus $\mathbb{N}m$.

Axiom 3: $\forall x \forall y \forall z(\mathbb{N}x \wedge Pxy \wedge Pxz \rightarrow y = z)$.

Axiom 3 states that successors of numbers are unique. Thus, suppose that $\mathbb{N}n$ and Pna and Pnb. By the definition of P, we have, for some F, G, c and d,

$$a = Nz : Fz \wedge Fc \wedge n = Nz : (Fz \wedge z \neq c)$$

and

$$b = Nz : Gz \wedge Gd \wedge n = Nz : (Gz \wedge z \neq d)$$

from which we have

$$Nz : (Fz \wedge z \neq c) = Nz : (Gz \wedge z \neq d).$$

Now by Hume's Principle,

$$\lambda z.(Fz \wedge z \neq c) \sim \lambda z.(Gz \wedge z \neq d)$$

Another easy lemma shows that "adjoining" a new item to two equinumerous concepts results in two equinumerous concepts. Thus we have

$$F \sim G$$

and hence, by Hume's Principle again,

$$Nz : Fz = Nz : Gz$$

from which we conclude that $a = b$, as required.

Axiom 4: $\forall x \forall y \forall z(\mathbb{N}x \wedge \mathbb{N}y \wedge Pxz \wedge Pyx \rightarrow x = y)$.

Axioms 4 states that (numerical) predecessors are unique. So suppose that $\mathbb{N}a$, $\mathbb{N}b$, Pac, and Pbc. Unwinding the definition of predecessor yields, for some F, G, x, and y,

$$c = Nz : Fz \wedge Fx \wedge a = Nz : (Fz \wedge z \neq x)$$

and

$$c = Nz : Gz \wedge Gy \wedge b = Nz : (Gz \wedge z \neq y).$$

By the first conjunct of each sentence, we have

$$Nz : Fz = Nz : Gz$$

so that, by Hume's Principle,

$$F \sim G.$$

Now, as above, it is an easy lemma to show that if F and G are equinumerous, "removing" an item from each results in equinumerous concepts. Hence,

$$\lambda z.(Fz \wedge z \neq x) \sim \lambda z.(Gz \wedge z \neq y)$$

so that by Hume's Principle,

$$Nz : (Fz \wedge z \neq x) = Nz : (Gz \wedge z \neq y)$$

from which we have $a = b$, which suffices to prove axiom 4.

Axiom 5: $\neg\exists x(\mathbb{N}x \wedge Px0)$.

Axiom 5 asserts that 0 has no predecessor. This follows almost immediately from the definition of 0 and predecessor. For recall that for x to precede y, there must be a concept F that y numbers and such that something falls under F. But this requires that F not be the empty concept. Yet by definition, 0 numbers the empty concept.

Axiom 7: $\forall F(F0 \wedge \forall x\forall y(\mathbb{N}x \wedge Fx \wedge Pxy \rightarrow Fy) \rightarrow \forall x(\mathbb{N}x \rightarrow Fx))$.

Axiom 7 is simply the induction axiom, stating that any concept closed under successor and under which falls 0 comprises all the natural numbers. We shall prove it before Axiom 6, since the proof of the latter appeals to induction.

To prove Axiom 7, we shall need a series of lemmas.

First, note that

$$R^*ab \wedge Fa \wedge \forall x\forall y(Fx \wedge Rxy \rightarrow Fy) \rightarrow Fb.$$

That is, if Fa and if b is in the R-series beginning with a and if F is closed under R-successors, then Fb. This follows almost immediately from the definition of the ancestral.

Next, note that we can weaken the condition R^*ab in the antecedent to $R^{*=}ab$. For if $a = b$, then we immediately have Fb.

We can now instantiate the concept $Fx \wedge R^{*=}ax$ for F in this fact to yield

$$\big[R^{*=}ab \wedge (Fa \wedge R^{*=}aa) \wedge \forall x \forall y((Fx \wedge R^{*=}ax) \wedge Rxy \rightarrow (Fy \wedge R^{*=}ay))\big] \rightarrow (Fb \wedge R^{*=}ab),$$

which implies

$$R^{*=}ab \wedge Fa \wedge \forall x \forall y(R^{*=}ax \wedge Fx \wedge Rxy \rightarrow Fy) \rightarrow Fb.$$

In words, if F is hereditary with respect to R "from a onwards," and if Fa and $R^{*=}ab$, then Fb. We can rewrite this fact as

$$Fa \wedge \forall x \forall y(R^{*=}ax \wedge Fx \wedge Rxy \rightarrow Fy) \rightarrow (R^{*=}ab \rightarrow Fb).$$

We can now derive Axiom 7 as an instance of this lemma by instantiating P for R and 0 for a and generalizing over b:

$$F0 \wedge \forall x \forall y(P^{*=}0x \wedge Fx \wedge Pxy \rightarrow Fy) \rightarrow \forall x(P^{*=}0x \rightarrow Fx).$$

Substituting in definitions for $\mathbb{N}$ yields

$$F0 \wedge \forall x \forall y(\mathbb{N}x \wedge Fx \wedge Pxy \rightarrow Fy) \rightarrow \forall x(\mathbb{N}x \rightarrow Fx),$$

which is the induction axiom.

Axiom 6: $\forall x(\mathbb{N}x \rightarrow \exists y(Pxy))$.

Axiom 6 asserts that every number precedes some thing; in other words, every natural number has a successor. As mentioned above, the proof is by induction. The induction concept is $\lambda x.P(x, Ny : P^{*=}yx)$. The embedded concept $\lambda y.P^{*=}yx$ is the concept *weak predecessor of* x (where the "weak" indicates that x itself is to fall under this concept). Thus, x falls under the induction concept just in case its successor is the number of weak predecessors of x; that is, the successor of x is the number of numbers less than or equal to x. (Note that the concept makes

no restriction to predecessors that are numbers. But it can be shown that the predecessors of any number "bottom out" in 0, which has no predecessor.)

The proof of this axiom is by far the most involved. It is not especially difficult, but it draws upon a large number of lemmas. We shall therefore just sketch the proof of the base and successor cases of the induction, and we refer the reader to Heck (2011f §2.5) for details.

For the base case, we must show that

$$P(0, Nx : P^{*=}0x),$$

i.e. that the number of things identical to or preceding zero is the successor of 0. We first show that nothing precedes zero, from which it follows that the only thing satisfying $P^{*=}x0$ is 0 itself. It follows that 0 is the number of the concept $\lambda x.x \neq 0 \wedge P^{*=}x0$. Hence, by the definition of successor, we see that the number of the concept $\lambda x.P^{*=}x0$ is the successor of 0. This establishes the base case.

For the inductive step, we must show that

$$\forall y\big(\mathbb{N}y \wedge P(y, Nx : P^{*=}xy) \rightarrow \forall z(Pyz \rightarrow P(z, Nx : P^{*=}xz))\big),$$

that is, for any number y, if the successor of y is the number of objects preceding or identical to y, then for any successor z of y, the successor of z is the number of objects preceding or identical to z. In other words, we must show that the property *being the number of numbers preceding or indentical to oneself* is hereditary with respect to P.

The proof requires two lemmas, which we state without proof. First, we prove that no number precedes itself, that is

$$\mathbb{N}a \rightarrow \neg P^{*}aa.$$

The proof is by induction. Using this result, we show that

$$\mathbb{N}b \wedge Pab \rightarrow Nx : P^{*=}xa = Nx : (P^{*=}xb \wedge x \neq b).$$

That is, for any natural number b, if a precedes b, then the number of numbers

preceding or equal to a is the number of numbers preceding b.

We are now ready to derive the successor case of the induction. Suppose that its antecedent is true, i.e. (working with some particular a)

$$\mathbb{N}a \wedge P(a, Nx : P^{*=}xa).$$

Suppose also that Pab. We must show that $P(b, Nx : P^{*=}xb)$. By our second lemma we have that

$$Nx : P^{*=}xa = Nx : (P^{*=}xb \wedge x \neq b)$$

and hence

$$P(a, Nx : P^{*=}xb \wedge x \neq b).$$

Now, Axiom 3 asserts the uniqueness of successors (of natural numbers). Thus, from Pab and this last result, we have

$$b = Nx : (P^{*=}xb \wedge x \neq b).$$

Thus, it suffices to show that

$$P(Nx : (P^{*=}xb \wedge x \neq b), Nx : P^{*=}xb).$$

This will follow from the fact that "adjoining" a new object a to a concept results in a concept whose number is the successor of the number of the original concept. Formally,

$$\neg Fa \rightarrow P(Nx : Fx, Nx : (Fx \vee x = a)).$$

For this, we look to the definition of successor, which we recall to be

$$P(a, b) \leftrightarrow_{\text{def}} \exists F \exists x (Fx \wedge a = Ny : Fy \wedge \wedge b = Ny : Fy \wedge x \neq y).$$

Our statement about adjunction follows easily using Hume's Principle and some elementary facts of pure second-order logic. Now, our earlier statement that the number of numbers preceding but not equal to b precedes the number of numbers

preceding or equal to b follows by taking F to be $\lambda x.P^{*=}xb \wedge x \neq b$. For this yields

$$P(Nx : (P^{*=}xb \wedge x \neq b), Nx : ((P^{*=}xb \wedge x \neq b) \vee x = b)).$$

But using Hume's Principle together with the fact that $P^{*=}bb$, we can show that $Nx : ((P^{*=}xb \wedge x \neq b) \vee x = b)) = Nx : P^{*=}xb$, as required.

This completes the proof. □

We have shown that FA interprets PA2. We mentioned above that this result partly explains the philosophical interest of Hume's Principle. How so? The idea is that, by knowing Hume's Principle (and hence FA), one could come to know (a wide class of) arithmetical truths—in particular, the truths derivable from PA2. But these truths are expressed in the language of arithmetic. What Frege's Theorem shows is that, on the strength of Hume's Principle, we can come to know whatever is expressed by their translation into the language of pure second-order logic together with the *number of* operator. Evidently, then, the philosophical interest of Frege's Theorem depends on the quality of our translation. And as Sean Walsh has argued (to my mind, convincingly), some translations fare very poorly indeed when it comes to preserving or transferring epistemic properties such as knowability or warrant (Walsh 2014 see). We therefore note that the Neo-Fregean must explain why his definitions are good ones if Frege's Theorem is to have any implications for our knowledge of arithmetic.

The epistemic status of Hume's Principle

We mentioned above that there are two sources for the philosophical interest of Hume's Principle. The first is that it interprets PA2. The second is that the principle is supposed to have a special epistemic status. Let us briefly review what that status is supposed to be.

The "official" Neo-Logicist line is that Hume's Principle is an *implicit definition* of the "number of" operator $Nx : Fx$ (Hale and Wright 2001a). The principle is a definition in that it is an attempt to "explain the meaning" of the cardinality operator. The definition is implicit in that it does not allow one to eliminate occurrences of the operator everywhere. Only in specific contexts—namely, those

of the form $Nx : Fx = Nx : Gx$, where F and G do not themselves contain occurrences of the operator—does Hume's Principle allow us to paraphrase away Nx.

The definitional status of Hume's Principle, in turn, is supposed to ground a favorable epistemic status: analyticity. If the principle is a definition, then it can be known apriori, simply by stipulation, in the way that "apples are red" can be known apriori if we stipulate that "apple" is to mean "red thing." What underlies this picture of apriori knowledge is what Hale and Wright call the "traditional conception." On this conception, some principles are constitutive of the meanings of (some of) the expressions occurring in them. The meanings thereby constituted are such as to make the principles true.[15]

This traditional conception and the attendant notion of analyticity have of course been at the center of some of the most heated disputes in 20th century analytic philosophy.[16]

In early presentations of Neo-Logicism, Hume's Principle was advanced as an analytic truth—as a definition or explanation of the concept of number. The sense of analyticity at issue was epistemic, not metaphysical.[17] Nonetheless, it soon became clear that the real issue was not about whether Hume's Principle is analytic, but instead about whether it is knowable on the basis of pure logic.[18]

[15]Of course, the relevant "principles" do not have to be propositions—one prominent version of this idea traces back to Gentzen's implicit definition of the logical constants (Gentzen 1969), and Gentzen's "principles" were rules of inference, not propositions. In this case, the meanings conferred onto the logical constants are such as to make the rules valid.

[16]Beginning with Quine (1951, 1954).

[17]See Boghossian (1996) for the distinciton. Roughly, a truth is epistemically analytic if it can be known solely on the basis of a grasp of the concepts it involves. A truth is metaphysically analytic if it is true solely in virtue of the meanings of the words occurring in its statement.

[18]Wright (2001a) concedes as much.

Chapter 2

The Caesar Problem

Frege introduces the method of abstraction by means of an example. Directions are abstract objects, and hence not (directly) given to us in intuition. How, then, can we think about or refer to directions? In §64, Frege suggests the following: directions are given to us as the directions of lines. Specifically, the following *Direction Principle* holds:

$$dir(a) = dir(b) \leftrightarrow a//b.$$

Here, $dir(a)$ is to be read as "the direction of a" and $a//b$ as "line a is parallel to line b."

Lines, we may suppose, *are* given to us in intuition. And likewise, facts concerning parallelism of lines can be determined by appeal to intuition. Thus, we may treat the Direction Principle as an implicit definition of the "direction of" operator and use it to explain our cognitive access to directions, in much the same way that Neo-Fregeans wish to use Hume's Principle to explain our cognitive access to numbers.

Despite its apparent virtues, Frege ditches this abstractionist account of direction. Having disposed of two objections, he writes

> But there is still a third doubt that may make us suspicious of our proposed definition. In the proposition
>
> "the direction of a is identical with the direction of b"

> the direction of a plays the part of an object, and our definition affords us the means of recognizing this object as the same again, in case it should happen to crop up in some other guise, say as the direction of b. But this means does not provide for all cases. It will not, for instance, decide for us whether England is the same as the direction of the Earth's axis—if I may be forgiven an example which looks nonsensical. Naturally, no one is going to confuse England with the direction of the Earth's axis; but that is no thanks to our definition of direction. That says nothing as to whether the proposition
>
> "the direction of a is identical with q"
>
> should be affirmed or denied, except for the one case where q is given in the form of "the direction of b." What we lack is the concept of direction; for if we had that, then we could lay it down that, if q is not a direction, our proposition is to be denied, while if it is a direction, our original definition will decide whether it is to be denied or affirmed.
>
> (§66)

The abstractionist definition of directions given in §64 settles the identity of directions when these are given as such—as the direction of this or that line. But it remains silent on whether a direction is identical with some object *not* given as a direction. And an exactly parallel objection vitiates the purported definition of numbers by Hume's Principle: "we can never—to take a crude example— decide by means of our definitions whether any concept has the number Julius Caesar belonging to it, or whether that same familiar conqueror of Gaul is a number or is not" (§56).[1]

[1]In fact, the quote in the text is raised as an objection to a different proposal. On that proposal, "the number of Fs is n" is defined inductively as follows (writing $\#F$ for "the number of Fs"):

$$\begin{array}{rcl} \#F = 0 & \leftrightarrow & \forall x \neg Fx \\ \#F = n+1 & \leftrightarrow & \exists x(Fx \wedge \#[Fy \wedge y \neq x] = n) \end{array}$$

As Frege notes, this definition gives a sense to identities of the form $\#F = n$, where n is a numerical term. But it leaves open whether, for example, $\#F =$ Julius Caesar for some F. This is precisely the issue affecting the definitions based on the Directions principle and Hume's Principle.

Frege concludes that "we cannot by these methods [of abstractive definition] obtain any concept of direction with sharp limits to its application, nor therefore, for the same reasons, any satisfactory concept of Number either" (§68). With Hume's Principle thus discarded as a basis for defining Number, Frege goes on to identify numbers with extensions of concepts.[2]

But what, exactly, is the problem? Frege's own complaint is that the implicit definition of number by means of Hume's Principle does not decide whether Julius Caesar is the number 0, or indeed whether he is a number at all. But Frege does not explain why this should be a failing on the part of Hume's Principle. After all, the non-identity of Caesar and 0 does not seem to be a mathematical fact—it is not taught in schools, and I have yet to see such a result appear in the Journal of the AMS. Why think that it must admit of a proof from within a foundational theory of arithmetic? Absent a convincing answer to this question, the Caesar Problem will seem to be a mere distraction, not a matter deserving the attention of serious philosophers and logicians.

I will argue tha the Caesar Problem is in fact a deep problem that strikes at the very heart of Neo-Logicism; it has exactly as much philosophical interest as that program itself. But seeing this takes a bit of setup. One of the goals of this section is therefore to articulate precisely why the Caesar Problem is properly so-called. As we will see, there are several ways of spelling out the problem; a subsidiary goal of this section is to bring attention to the differences between these articulations and highlight their respective assumptions.

Before attempting a theoretical account of the problem, it will be worth our while to display some of the examples that motivate the problem. We shall avail ourselves of somewhat fine-grained terminology in this section: we shall let "0" stand for our ordinary term for the number 0 (the one that you and I employ), which we shall distinguish from "0_{HP}," which will abbreviate $Nx : x \neq x$. We shall use "0_{FHP}" to designate the corresponding abstract of the *Finite Hume's Principle*:

$$Bx : Fx = Bx : Gx \leftrightarrow (F \sim G \vee F \text{ and } G \text{ are both infinite})^3$$

[2]As is well-known, that sad story ends in contradiction. And in any case, this "solution" to the Caesar Problem merely kicks the problem up a floor, since extensions are themselves abstract objects, and Frege's account of them via Basic Law V also invites the Caesar Problem.

[3]The B is for Boolos, who introduced this principle into the literature.

(Infinitude can be defined in pure second-order logic; see e.g. Shapiro (1991).) We shall also use "0_{RA}" to denote an abstraction term for 0 in the context of a suitable abstraction principle for real numbers (see for example Hale 2001b; Shapiro 2000). Finally, recall that $dir(a)$ denotes the direction of line a under the principle of Direction Abstraction.

Now consider the following statements:

(i) $\text{Caesar} = 0_{\text{HP}}$

(ii) $0 = 0_{\text{HP}}$

(iii) $0 = 0_{\text{FHP}}$

(iv) $0_{\text{HP}} = 0_{\text{RA}}$

(v) $0_{\text{HP}} = 0_{\text{FHP}}$

(vi) $dir(a) = 0_{\text{HP}}$

(vii) 0_{HP} is Roman

Each of (i)–(vii) gives rise to a Caesar Problem, insofar as each is a sentence undecided by the abstraction principles governing the terms involved. Each brings out a different aspect of the Caesar Problem. (i) asks whether 0_{HP} is identical to a concrete particular—Caesar—not given to us via an abstraction operator. (ii) and (iii) ask whether $0_{\text{HP/FHP}}$ is identical to an abstract (as in *non-concrete*) object—namely, 0—though one that is also not given to us via an abstraction operator. (iv), (v), and (vi) ask about an identity between two abstracts subject to different abstraction principles (see Cook and Ebert 2005 for an overview of such cross-abstract identifications). Finally (vii) is an atomic sentence involving an abstract that is not an identity statement. (vii) reminds us that it is not only identity statements that Hume's Principle fails to decide. As we will see in the next subsection, the only atomic statements decided by Hume' Principle are numerical identities.

It is important to lay down a variety of examples because our verdicts about the cases are not uniform. (i), (vi), and (vii) are uncontroversially false, and known

to be so.[4] By contrast, the Neo-Logicist project presupposes that (ii) is true: the point of the project is to explain our knowlege of numbers, and this it could only do if Hume's Principle gave us access to *bona fide* numbers, a situation to which (ii) gives (partial) expression. (iv) and (v) are somehwat divisive: some philosophers think they are false, while others think they are true (see e.g. Studd 2023). Finally, by the transitivity of identity (and assuming the truth of (ii)), one's stance on (iii) should be the same as one's stance towards (v).

As stated above, Hume's Principle (together with other relevant abstraction principles) does not decide any of these statements. Should it? And if so, which way should statements like (iv) and (v) be decided? The fact that we judge some of (i)–(vii) true and some false should make the most naive response to the Caesar Problem—that of decreeing all mixed identities to be false and calling it a day—unattractive. For doing so would run roughshod over the key Neo-Logicist principle (ii) and over the attractive-sounding (iv) and (v).

In just about everybody's opinion, (i), (vi), and (vii) should come out as false, and moreover, this should be knowable on the basis of a grasp of the concepts involved. Thus, if Neo-Logicism is to be safeguarded, it must be shown that (i), (vi) and (vii) can be refuted on the basis of Hume's Principle, perhaps toghether with other principles of similar epistemic status. Matters are less clear regarding the other sentences. Whether they ought to be decided, and how, will depend on what precisely one takes the Caesar Problem to consist in. To elucidate this question, we will carefully present various interpretations of the Caesar Problem. As we will see, different versions of the problem make different demands on the Neo-Logicist program: some require all of (i)–(vii) to be settled, while others require only some. And some require them to be answered in a specific way, whereas others require only that they be answered. Before explaining the various versions of the problem, however, it will be helpful to record some elementary logical facts that form the technical backbone of all of the interpretations we will consider.

[4]MacBride (2006 §2.3) argues that we could only know these facts if our common-sense understanding of numbers and emperors contained, perhaps "implicitly," an argument for them. I disagree, but pursuing the matter would take us too far afield. Suffice it to say that I do not believe that all knowledge is grounded in argument.

2.1 Some logical facts

Following Wright, let us imagine that we are adding the "number of" operator to a language containing only vocabulary in good philosophical standing, and with no means of reference to number. This somewhat idealized scenario is justified in light of Neo-Logicism's goals: to explain how someone priorly innocent of number could acquire a concept of number and knowledge of arithmetic on the basis of Hume's Principle. Let $\mathcal{L}$ denote the base language and $\mathcal{L}^+$ denote the extension of $\mathcal{L}$ with the cardinality operator Nx. An interpretation of Nx over a given domain M is a function from $\mathcal{P}(M)$ to M. In general, we shall be concerned with expanding a given $\mathcal{L}$-structure $\mathcal{M}$ to an $\mathcal{L}^+$-structure $\mathcal{M}^+$ that preserves the interpretation of the vocabulary of $\mathcal{L}$. All of this is simply to model the addition of Nx and Hume's Principle to a base language.

The relevant facts are as follows. First, we note that models of Hume's Principle are easy to come by.

Proposition 2.1.1. Let $\mathcal{M}$ be an infinite model of some $\mathcal{L}$-theory T. Then $\mathcal{M}$ can be expanded to a model $\mathcal{M}^+$ of $T + \mathsf{HP}$.

Proof. Suppose that $\mathcal{M}$ has cardinality $\kappa \geq \aleph_0$. Let $f : \kappa \to M$ be an injection. Now let the interpretation of Nx in $\mathcal{M}^+$ be g, defined as follows:

$$g(F) = \begin{cases} f(0) & \text{if } |F| = \kappa \\ f(|F|) & \text{if } \aleph_0 \leq |F| < \kappa \\ f(|F|+1) & \text{if } |F| < \aleph_0 \end{cases}$$

Clearly, this interpretation of Nx satisfies HP. Thus, it remains to check that T is satisfied in $\mathcal{M}^+$. But this too is clear: we have expanded the language but left the interpretation of the symbols of $\mathcal{L}$ unchanged. □

A rather more involved proof (essentially Frege's proof of the infinitude of the number series, given above) shows that Hume's Principle is only satisfiable on infinite domains. We record these two facts in the following proposition:

Proposition 2.1.2. For any theory T, the theory $\mathsf{HP} + T$ is satisfiable on all and only the infinite models of T.

Equivalently, Hume's Principle is consistent with a theory T iff T has an infinite model.

As an easy corollary with T the empty theory, we have that any infinite set M can serve as the domain of a model of Hume's Principle. Thus, Hume's Principle imposes no cardinality requirements on its models besides that they be infinite.

An important stronger result is that the same holds even if a base domain is fixed.

Proposition 2.1.3. Let $\mathcal{M}$ be an $\mathcal{L}$-structure of cardinality κ and K an infinite set of cardinality $\geq \kappa$ disjoint from M, the underlying domain of $\mathcal{M}$. Then there is a structure $\mathcal{M}'$ whose domain is the union of M and K satisfying HP.

In other words, starting with a collection of κ objects, we can add at least κ new objects—intuitively, the numbers—and end up with a model of HP. (In fact, we do not need to add *any* new objects as long as κ is already infinite, since we can interpret the numbers as being amongst the old objects. What the present proposition claims, however, is that the result holds even if we insist that numbers be "new." That is, we can add as many numbers as we like to an existing domain, whether finite or infinite, and still satisfy HP.) The proof is similar to that of Proposition 2.1.1.

Next, we note that Hume's Principle imposes no significant structure on the underlying domain.

Proposition 2.1.4. Let $\mathcal{M}$ be a model of HP interpreting Nx as $f : \mathcal{P}(M) \to M$ and let g be an injection whose domain includes the abstracts of HP. Let $\mathcal{M}'$ be like $\mathcal{M}$ expect that Nx is interpreted by the following function:

$$h(F) := g(f(F))$$

Then $\mathcal{M}' \models \mathsf{HP}$.

This proposition signals that the identity of the abstracts does not matter, so far as the satisfiability of Hume's Principle is concerned: any assignment of objects to concepts that respects cardinality will do.

In fact, matters are worse: even if a base domain is fixed, the model may be expanded to models of Hume's Principle of different cardinalities.

Proposition 2.1.5. Let $\mathcal{M}$ be a model of an $\mathcal{L}$-theory T with infinite models and let K be a set of any infinite cardinality. Then there is a model $\mathcal{M}'$ of $T + \mathsf{HP}$ whose domain is M—the underlying domain of $\mathcal{M}$—together with K.

With these facts on the table, we are now in a position to explain the various guises of the Caesar Problem.

2.2 The epistemic problem

On one line of thinking, the problem is that (i) (and the like) are not just false, but *obviously* false. More precisely, (i) is known to be false on the basis of a grasp of the concept of number. If Hume's Principle does not allow us to know that (i) is false, then it cannot be what underlies our grasp of number. Here is Heck raising the point:

> […] it is rarely mentioned, because it is not thought important, that Frege takes for granted that we *do* recognize that Caesar is not a number: Frege's objection is *not* that HP does not decide whether the singleton of the null set is the number zero (which, on Frege's explicit definition, it happens to be). The example Frege chooses is one about which he takes us to have strong intuitions: Whatever numbers may be, Caesar is not among them. Thus, there must be *more* to our apprehension of numbers than a mere recognition that they are objects that satisfy HP. Something explains *why* "no one is going to confuse [Caesar] with the [number zero]."
>
> Heck (2011e: 121)

Note that the problem is not that Hume's Principle does not suffice to establish everything we know about number. For I know what my favorite number is, but good luck figuring out what it is on the basis of Hume's Principle and pure second-order logic. Rather, (Heck argues) there are some numerical truths that we know purely on the basis of our "apprehension of numbers"; such truths include the negation of (i) (but obviously exclude the truth about my favorite number), and such truths should follows from any good explanation of our grasp of numbers

(recall that (i) was "Caesar = 0"). Since the negation of (i) does not follow from Hume's Principle, Hume's Principle is deficient.

An interesting consequence of this line of thought is that Hume's Principle is under no obligation to decide sentences which we ourselves cannot decide (on the basis of the concept of number). If sentences like (ii), (iii), (iv)[5] are undecided by our grasp of the number concept, then it is no objection to Hume's Principle that it does not decide them.[6] In fact, in such a situation, it might be desirable *not* to decide such statements. If it did, Hume's Principle might be guilty of epistemic overreach, forcing conclusions about matters of ontology that go beyond anything implicit in our mathematical practice.[7] As we shall see, this construal of the Caesar Problem conflicts with some of the ones to be canvassed below, on which leaving any such identity undecided would be fatal to the Neo-Logicist project.

What should we make of this version of the Caesar Problem? The answer depends very much on its target. If the Neo-Logicist project is to explain *our* knowledge of arithmetic and *our* capacity to refer to numbers, then the objection might hit its mark. For (unless it can be answered), it shows that our numerical knowledge outstrips that which Hume's Principle supports. On the other hand, if the project is to provide a route—any route—to arithmetical knowledge, then this epistemic objection loses much of its force. For all that is required of Hume's Principle, on this conception of the Neo-Logicist project, is that it deliver the *arithmetical* truths we know. Of course, the Neo-Logicist must then distinguish arithmetical truths from other truths involving numbers, and he must argue that all Caesar-style questions fall outside this range. A plausible first pass at drawing this distinction would be to deem a truth arithmetical if it can be expressed in the language of second-order Peano Arithmetic. Of course, Hume's Principle does not decide all such sentences. But perhaps it decides most or all of those we might one day come to know.

We conclude that the epistemic Caesar Problem poses at best a limited challenge to Neo-Logicism.

---

[5]Recall that (ii) is $0 = 0_{\mathrm{HP}}$, (iii) is $0 = 0_{\mathrm{FHP}}$, and (iv) is $0_{\mathrm{HP}} = 0_{\mathrm{RA}}$.

[6]Heck (2011e: 121–122) makes the same point.

[7]Wright (2001b: 250) makes a similar point.

2.3 The semantic problem: indeterminacy

Fine (2002) presses a model-theoretic version of the Caesar Problem inspired by the facts rehearsed section 2.1. The problem revealed by the facts above, according to Fine, is that Hume's Principle does not sustain a suitably *determinate* conception of the natural numbers.

Intuitively, our concept of natural number is *determinate* if it provides answers to all the questions we can formulate concerning the natural numbers. What is it for a concept to "provide answers" to arithmetical questions? It is not required that, from axioms capturing that concept, we should be able to decide all arithmetical sentences—for Gödel's incompleteness theorems dash that hope. Rather, what determinacy requires is that our conception single out a class of structures, all of which agree on arithmetical statements. In other words, if any two structures disagree on an arithmetical claim, then at least one of these must be an unacceptable way of interpreting our arithmetical concepts.[8]

Note that many perfectly good concepts are not determinate in this sense. For example, there are presumably several distinct acceptable ways of interpreting the concept *chair*. On some of these ways, borderline cases of chairs (sculptures of chairs, for example) will count as chairs, while on others, they will not. The concept *chair* is therefore not determinate. But this in no way impugns our capacity to

[8]In view of Gödel's completeness theorem, does this requirement not collapse into the previous one, that an axiomatic theory capturing our conception be complete? Gödel's completeness theorem asserts that first-order semantic consequence and first-order syntactic consequence coincide. It follows that if a sentence is true in all models of a theory, then that sentence is provable from that theory. Hence, it might seem as though the second characterization of determinacy entails the first: if every admissible interpretation of our arithmetical concepts agree on every arithmetical sentence, then every arithmetical sentence should be decided by a decent axiomatization of our arithmetical concepts.

This line of reasoning, however, assumes that any "decent axiomatization" of our arithmetical concepts will be a first-order one. However, a number of recent trends in the philosophy of logic and mathematics push back on the assumption that first-order logic is the only acceptable way of formalizing our mathematical knowledge. For example, a number of authors view second-order logic (with its standard semantics) as the right way to axiomatize our mathematical theories (see Shapiro 1991). Others (McGee 1997; Murzi and Topey 2021; Warren 2020b) hold that our use of mathematical concepts is open-ended, in a way that resists capture by a single first- (or second-)order theory. If any such view is correct, then our mathematical competence might be able to pin down structures giving unequivocal answers to arithmetical questions, even if the complete theory of these structures remains beyond our (computable) deductive grasp.

refer to or theorize about chairs (whatever good that may do us). The problem of determinacy is meant to apply specifically to the case of number. *Natural number*, unlike *chair*, but perhaps like *electron*, is supposed to be determinate. It is, to many, simply implausible that our natural number concept fails to single out a single structure. Moreover, it might be held that the scientific credentials of arithmetic—its reputation as an unassailably secure body of knowledge, one of the few to command rational certainty—depend (in part) on its vocabulary being rigorously and unambiguously defined. Any vagueness in the basic concepts of arithmetic could threaten the status of arithmetic as a science.

Thus, it is taken as a desideratum on an account of the concept *natural number* that it explains how we could grasp a determinate natural number concept, or, equivalently, that the account explains how the concept which, upon theoretical reflection, it ascribes to us could be determinate. The results of the previous section, insofar as they show that Hume's Principle has many non-equivalent admissible interpretations, suggest that the Neo-Logicist account of our concept of number does not meet this desideratum.

Indeed, Hume's Principle badly fails to constrain the interpretation of the "number of" operator, and fails in two ways. First, by the corollary to Proposition 2.1.2 above, Hume's Principle has models of every infinite cardinality. Since such models cannot be isomorphic, Hume's Principle is not categorical. Worse still, the same holds relativized to a base domain of objects: by Proposition 2.1.3, if we start with a model $\mathcal{M}$ not containing numbers, we can extend it to a model $\mathcal{M}'$ of Hume's Principle containing κ numbers, for any infinite κ.

Second, even if we fix an (infinite) cardinality for the domain, by Proposition 2.1.3, *any* assignment of concepts to objects that respects equinumerosity will satisfy Hume's Principle. So for all that Hume's Principle says Caesar could be the number 0, my copy of the *Grundgesetze* could be the number 1, and 2 could be a translantic cable. But notice that different assignments will, in general, result in different truths. For example, Caesar $= 0$ is true on the assignment just mentioned, but false on the intended assignment. If all these interpretations are admissible, we have rampant indeterminacy. Fine (2002) calls these two problems the "external" and "internal" Caesar Problems, respectively, though I think the terms "vertical" and "horizontal" are more fitting, and will use these.

Thus, if Hume's Principle was meant to determine a reference for numerical terms by constraining the admissible interpretations of our mathematical language, the Caesar Problem shows that Hume's Principle has failed its purpose. Hume's Principle does not secure a determinate conception of natural number.

How serious is this problem? I will argue that the horizontal problem is indeed serious, but that the vertical problem is not.

Hume's Principle is on the hook only for as much determinacy as our concept of number actually enjoys. It ought not be a requirement on Hume's Principle that it decide how many numbers there are, if our actual concept of number also fails to do so. The vertical Caesar Problem exploits the facts of section 2.1 to argue that our concept *natural number*, which is determinate, is not captured by Hume's Principle, which does not yield a determinate concept. Now, our concept *number* is plausibly determinate enough to rule that there are infinitely many numbers. It is much less plausible that it is determinate enough to rule that there exist κ-many numbers for some measurable cardinal κ.[9] And indeed, the (vertical) indeterminacy of Hume's Principle affects the higher infinite—not the natural numbers. Indeed, in every model of Hume's Principle, the first ω abstracts form a structure isomorphic to the natural numbers. Moreover, this initial segment is definable within the theory.[10] So the fact that Hume's Principle leaves the number of numbers undetermined signals an indeterminacy only in the general concept of cardinality it provides, and not in the concept of *finite* number it makes available to us.[11] Absent some

[9] A measurable cardinal is a very large cardinal. The existence of measurable cardinals is independent of ZFC.

[10] Essentially in the same way that the set of natural numbers is characterized in ZFC, as the smallest inductive set: x is a natural number $\leftrightarrow \forall F((F0 \wedge \forall y(Fy \rightarrow FSy)) \rightarrow Fx)$.

[11] How does the general concept of cardinality provided by Hume's Principle relate to the concept of cardinality articulated in modern axiomatic set theory? The most salient difference is that Hume's Principle assigns a number to every concept (including the universal concept), whereas ZFC does not assign cardinalities to many concepts (of course, ZFC assigns cardinalities in the first instance to sets, not to concepts). Is this a problem? George Boolos, for one, thinks so: "one who seriously believes that [HP is an analytic truth] has to be bothered by the incompatibility of the consequence of Frege arithmetic that there is such a number as [the number of everything] with the claim made by ZF plus standard definitions …that there is no such number" (Boolos 1997: 260). There are two lines of response to this objection. First, one can deny that numbers, as characterized by Hume's Principle, are sets. Since Hume's Principle and ZFC have different subject matters, there is no conflict. Second, one can try to motivate some restrictions on Hume's Principle. For example, one might want to restrict its applicability to *extensionally definite* concepts, where such concepts contrast with *indefinitely extensible* concepts.

argument that our concept number enjoys more (vertical) determinacy than this, the Neo-Logicist can respond to the vertical Caesar Problem that Hume's Principle confers on the number concept all the determinacy it in fact has.

I conclude that the vertical Caesar Problem need not unduly worry the proponent of Hume's Principle. What about the horizontal problem? Matters here are a bit more subtle. Recall that this version of the problem takes its start by noting that, even if we fix both a base domain and a cardinality for the whole universe (numbers included), we are still free to interpret "natural number" as just about any countable set. Why should this threaten determinacy? Because on some interpretations, (i) above—Caesar = 0—will be true, and on others false. Thus, it will be indeterminate whether 0 is Caesar.

One immediate line of response to the horizontal problem is to urge a modicum of structuralism. For *no* theory is immune from this kind of problem: given a model $\mathcal{M}$ of any T (even a second-order theory T), and a set D of the same cardinality as $\mathcal{M}$, we can turn D into a model of T by "pushing" the structure of $\mathcal{M}$ onto D via a bijection (whose existence is guaranteed by the fact that $\mathcal{M}$ and D have the same cardinality).

In more detail, suppose $\mathcal{M} \models T$ is an $\mathcal{L}$-structure, and let f be a bijection between M, the underlying domain of $\mathcal{M}$, and D, which could be any set. We construct a model $\mathcal{D}$ of T based on D as follows. For each symbol of $\mathcal{L}$, we let its interpretation in $\mathcal{D}$ be the image under f of its interpretation in $\mathcal{M}$. That is, we have

$$c^{\mathcal{D}} = f(c^{\mathcal{M}})$$
$$R^{\mathcal{D}} = f(R^{\mathcal{M}})$$
$$g^{\mathcal{D}} = f(g^{\mathcal{M}})$$

where $f(A) := \{f(a) : a \in A\}$. It is easy to verify that $\mathcal{D} \models T$.

In light of the universal availablility of this "push-through" construction—first brought to philosophical attention by Putnam (1977, 1980); see Button and Walsh (2018 ch. 2)—we might wonder whether there is a special problem for Hume's

How precisely to characterize the concept of indefinite extensibility is of course a vexed question (see Linnebo 2018a).

Principle here. It might seem, instead, that this kind of referential indeterminacy is a problem for *any* attempt to explain reference in terms of a theory's narrowing down the space of admissible interpretations beyond isomorphism. For example, suppose we were to take the Dedekind-Peano axioms, instead of Hume's Principle, as an implicit definition of the number concept (a suggestion raised by MacFarlane 2009). If these axioms were embedded in a broader language capable of referring to Roman emperors (as they evidently are), the same problem would arise: there would be no telling whether Caesar is the number zero—some interpretations of the axioms would interpret 0 as Caesar, and other would not.

A standard response to Putnam's push-through construction in the philosophy of mathematics is thus to take on board this much structuralism: we ought not ask our theories to pin down their models up to more than isomorphism (see Button and Walsh 2016). Now, as was pointed out above, Hume's Principle is *not* categorical: it has (many) non-isomorphic models. But insofar as the models differ with respect to height, this is just the vertical Caesar Problem. And if we fix both the size of a model and its non-numerical population, then all remaining interpretations of Hume's Principle over this domain are isomorphic (as can be seen by taking the obvious bijection from one model's numbers to the other).[12] Does the failure of Hume's Principle to provide more than this really signal a flaw in the Neo-Logicist program?

Once again, the answer depends on the ambitions of that program. If Hume's Principle is meant to *recover* our ordinary concept of number, then any indeterminacy in the recovered concept not found in the original will show that it has failed its purpose. If our natural number concept really is determinate enough to decide whether Caesar is zero, and if Hume's Principle fails to decide this same fact, then the concept provided by the latter simply cannot be our concept of number. This will be true regardless of whether one thinks that asking for theories to pin down interpretations up to more than isomorphism is "asking for too much." On the other hand, if the principle is simply meant to exhibit a route to knowledge of arithmetical truths, then—so long as one thinks that the distinctness of Caesar and zero is not an arithmetical truth, in the relevant sense—it will not matter

[12]For details on this limited kind of categoricity for Hume's Principle, see Walsh and Ebels-Duggan (2015).

whether the principle decides the distinctness of Caesar and zero.

2.4 The semantic problem: linguistic understanding

Recall one of the ambitions of Neo-Logicism: to explain how someone without prior grasp of the concept of number could achieve arithmetical knowledge. The basic shape of the Neo-Logicist answer is that Hume's Principle, should the thinker be so lucky as to come across it, would both endow the thinker with a grasp of the number concept and constitute a known truth from which the rest of arithmetic could be derived. The Caesar Problem poses a challenge to the first component of this solution: it puts pressure on the claim that Hume's Principle can explain a thinker's posession of the concept of number. Let us see why.

As we saw in section 1.3, Hume's Principle is supposed to provide an understanding of the "number of" operator by explaining, in a language that does not contain this operator, what it takes for a statement of numerical identity to hold. More specifically, Hume's Principle grants this understanding because the right-hand side of the principle gives the truth-conditions of the left-hand side: knowledge of the truth conditions of $F \sim G$, together with knowledge of Hume's Principle (understood as a statement of identity of truth-conditions) thus suffices for knowledge of the truth-condition of $Nx : Fx = Nx : Gx$.

By contrast, Hume's Principle does not appear to provide truth-conditions for mixed identities in previously understood terms—such explanations, at any rate, are not derivable from Hume's Principle. If understanding a sentence is a matter of knowing its truth-conditions, then it would seem that the speaker does not understand a wide range of sentences containing Nx. But he understands the rest of his vocabulary (the "number of" being the only linguistic innovation at issue), and he has a general grasp of the mechanisms of predication. Hence, the most (and only?) plausible explanation for his lack of understanding is that he does not understand the "number of" operator. Hence, knowing Hume's Principle and using the expression Nx in accordance with it does not suffice to secure an understanding of the number concept.

More specifically, note that sentences of the form $c = Nx : Fx$ are undecided by Hume's Principle whenever c is not of the form $Nx : Gx$. Now, if a speaker

does not understand such sentences, then he does not understand the predicate $\lambda y.y = Nx : Fx$. For understanding a predicate requires a grasp of that predicate's contribution to the truth-conditions of sentences in which it occurs (as argued for example by Davidson 1967; see also Wright 1983: 141–142; Hale and Wright 2001a: 134). If a speaker has such a grasp, then provided they understand the other expressions used in an atomic sentence containing that predicate, they understand the whole sentence. Conversely, if a speaker does not understand the sentence "Caesar is zero," yet understands the term "Caesar" and has a grasp of predication, then he does not understand the predicate "x is zero."

This follows from a general principle of compositionality for linguistic understanding put forward by Gareth Evans.[13] According to Evans, if a thinker is capable of thinking the thoughts Fa and Gb, then, as a matter of metaphysical necessity, they are capable of thinking the thought Fb.[14] More generally, Evans' insight is that thoughts consist of the structured exercise of basic capacities. Thought contents mark the exercise of these capacities. And these basic capacities are, by their very nature, freely combinable (a fleshed out version of the point would have to incorporate a range of qualifications, but the basic idea should be clear). Transposed from the realm of thought to that of linguistic understanding, Evans' point becomes that someone who can understand the sentences Fa and Gb can also, and necessarily, understand the sentence Fb. The underlying thought is again that understanding the meaning of a complex sentence is a matter of composing one's understanding of the parts of the sentence according to systematic rules, and that, within the constraints of syntax, the basic understandings are freely recombinable.

Why should it matter, as far as the goals of Neo-Logicist are concerned, whether a speaker understands the predicate "x is zero"? Why must someone have a general grasp of the conditions under which an object falls under the concept expressed by "x is zero" in order to come to know arithmetical truths?

The answer is that an understanding of this predicate is required to understand many aspects of the Neo-Logicist reconstruction of arithmetic. Most importantly,

[13] Hale and Wright (2001c: 342) also cite Evans in motivating their construal of the Caesar Problem.

[14] Evans (1982: 100).

such an understanding is required for the proof that every number has a successor. Recall that the key idea of the proof is to show that the number belonging to the following concept[15] is the successor of n:

$$x = n \vee P^*xn.$$

To understand this concept, one needs to understand its constituent $x = n$. But, according to the present construal of the Caesar Problem, this is precisely the kind of concept which Hume's Principle fails to provide a grasp of. For this concept is really an abbreviation for

$$x = Ny : Fy$$

for some suitable F (see section 1.3.2 for details). If Hume's Principle does not ground an understanding of concepts expressed by predicates of the form $x = Ny : Fy$, then Frege's (and our) proof of the infinity of the number series will be, in Heck's phrase, nothing more than "a series of uninterpreted squiggles" Heck (2011d: 20).

Thus, the Caesar Problem threatens the Neo-Logicist claim that knowledge of Hume's Principle can support the acquisition of concepts of the form $\lambda x.x = Ny : Fy$, and thereby the claim that knowledge of arithmetic can be grounded in knowledge of Hume's Principle.

Before moving on, we must clarify this problem somewhat. First, note that the issue is not that Hume's Principle does not allow one to know whether Caesar is zero, or more generally whether any number is identical with any non-abstract. Rather, the issue is that Hume's Principle does not even allow the speaker to understand such sentences. As Heck writes, "the Caesar problem is less about deciding the truth-values of certain statements than it is about guaranteeing that those statements have a sense. …The worry is not that [on the strength of Hume's Principle] we do not know whether Caesar is 0. It is that we do not so much as understand the question whether he is" (ibid.: 17)

Second, we must address the following (eminently reasonable) response:[16] someone who knows Hume's Principle *does* know the truth-conditions of sentences of

[15] We are being loose with the distinction between concepts and predicates here.

[16] Thanks to Thomas Schindler for pressing me to think about this line of response.

the form $c = Nx : Fx$. On the standard Tarskian theory, such a sentence is true just in case "c" and "$Nx : Fx$" refer to that same object (perhaps with the qualification that both do refer to an object). Likewise, "Caesar is a number" is true just in case what "Caesar" refers to falls under the concept expressed by "x is a number."[17] Or in a more disquotational key, "$c = Nx : Fx$" is true just in case c is $Nx : Fx$.[18] Either way, nothing has been said so far to rule out knowledge of the truth-conditions of mixed identities based on such theories (Tarskian or disquotational). And such theories seem available to the numerical neophyte no less than to the average speaker of English. If such theories are good accounts of the semantic competence of the latter, why do they not show that the former attains a similar competence with his vocabulary?

I do not think this line of response can finesse the Caesar Problem, however. For this line proves too much. Take the Tarskian theory first. Note that our speaker can give such a Tarskian theory for any language whatsoever, including one which he does not speak (provided he knows its syntax). Indeed, (given the right theory of syntax) I myself could give a Tarskian theory of truth for the Korean language (which I do not speak). But of course, I remain ignorant of the truth-conditions of Korean sentences. Likewise, our speaker's ability to give Tarskian truth-conditions for sentences containing $Nx : Fx$ does not show that he understands those sentences. In particular, knowing that "Caesar is 0" is true just in case "Caesar" and "0" co-refer does not suffice for a grasp of the truth-conditions of that sentence—not if grasping the truth-conditions of a sentence suffices for understanding it.

What about the disquotational response? Well, I could also give a disquotational theory of truth for Korean. Or at any rate, I could "mouth" such a theory, by prounouncing again the sentence whose truth-conditions I am trying to explain (actually *giving* such a theory probably requires that I know the language in which the theory is formulated). But my ability to do so clearly does not endow me with an understanding of the Korean language. Likewise, the numerical neophyte's ability to utter "'Caesar is 0' is true just in case Caesar is 0" betokens no understanding of that sentence.

[17] Tarski (1956); Davidson (1967).
[18] Ramsey (1927).

This response might be too quick, however. The Neo-Logicist claim, recall, is that to fix a meaning for a singular term, it suffices to give truth-conditions to identity statements containing it. The Neo-Logicist further insists that to do so, we do not need to give truth-conditions for *all* identity statements involving them term. We only need to do so for certain canonical identity statements. Having done so, we can use the understanding thereby acquired to understand occurrences of the singular term in other contexts—including in other identity contexts. So while it is true that a disquotational truth theory is of no help to someone who does not already understand the terms it contains, that is not the situation we are considering. Instead, we should ask whether, having already mastered Hume's Principle, a speaker is in a position to understand the occurrence of "zero" in the right-hand side of the following disquotation instance: "Caesar = 0" is true iff Caesar is zero.

Whether this is the case, it seems, depends on whether the Neo-Logicist is correct to claim that understanding a singular term only requires understanding a certain privileged subset of identity statements involving it (and on his having correctly identified this subset as the identity statements having the form of the left-hand side of Hume's Principle).

In closing, let us note that this version of the Caesar Problem is more stringent than the previous two. For while a solution to the previous two problems requires securing only as much determinacy, or knowledge of mixed identities, as we in fact enjoy, a solution to this Caesar Problem requires truth-conditions to be provided for *all* sentences containing number terms. Such is the import of Evans' generality constraint: understanding an expression anywhere requires understanding it everywhere. Conversely, a failure to understand "Caesar is zero" belies a failure to understand even "zero is zero." If this version of the Cesar Problem has merit, then, it threatens the entire viability of the Neo-Logicist project. In contrast to the situation with the previous two versions of the problem, I do not see a way to weaken the Neo-Logicist position so as to escape, or mitigate, this version of the problem. If I am right, either it is answered convincingly, or the Neo-Logicist program collapses.

2.5 Summary

We have presented several versions of the Caesar Problem. In this subsection, we review and summarize the situation.

We began with the epistemic Caesar Problem. According to this problem, we know that Caesar is not a number on the basis of a grasp of the relevant concepts, but knowledge of Hume's Principle does not suffice for knowing that Caesar is not a number. It follows that there is more to our grasp of the concept of number than knowledge of Hume's Principle. A prominent advocate of this formulation of the Caesar Problem is Heck (2011e).

We found the import of this problem to vary with the goals of the Neo-Logicist. We found this form of the problem to be of limited import. The stated ambitions of the Neo-Fregean program is to exhibit a route to arithmetical knowledge. If the Neo-Logicist wishes to be faithful to *our* route, then the epistemic problem is indeed a problem. But it might be open to him to settle for exhibiting *some* such route. That Hume's Principle fails to deliver some non-arithmetical knowledge is no mark against its contribution to *that* project.

We then considered construals of the Caesar Problem motivated by model-theoretic facts concerning Hume's Principle. The model-theoretic facts in question all reveal that Hume's Principle fails to constrain the interpretation of the "number of" operator, beyond requiring it to respect cardinality. As a result, the determinacy of the concept of number supported by Hume's Principle is called into question. Following Fine (2002), we distinguished between two dimensions of interpretation which Hume's Principle underdetermines. According to the *vertical* Caesar Problem, Hume's Principle fails to determine the number of numbers. According to the *horizontal* Caesar Problem, it fails to determine the *identity* of the numbers.

We argued that the Neo-Logicist need not lose sleep over the vertical Caesar Problem. Determinacy intuitions are strongest concerning arithmetical truths—truths concerning the first ω numbers. When it comes to determining the number of (infinite) numbers, our intuitions that there is a determinate truth of the matter are both weaker and less weighty. And as we saw, any vertical indeterminacy affecting Hume's Principle must take hold in the higher infinite. Truths of pure

arithmetic are unaffected by this indeterminacy.

By contrast, the horizontal Caesar Problem seems to pose more of a challenge. For it threatens a determinacy which our concept of number plausibly *does* enjoy. We pointed out that no purely "theory-first" method of reference fixing could secure this kind of determinacy. But this fact might be taken either to show that our number concept is indeterminate after all—this would be the way of a die-hard theory-first advocate—or to show that all theory-first views, Neo-Logicism included, fail in their intended purpose. As in the case of the epistemic problem, it is open to the Neo-Logicist to retreat to a weaker claim: that Hume's Principle provides a concept that supports knowledge of *some* arithmetical truths, even if it does not support knowledge of everything we take ourselves to know about numbers.

Finally, we examined construals of the Caesar Problem relating to linguistic understanding. We saw that Hume's Principle does not explain the truth-conditions of many sentences involving numerical terms. But this can only be due to the fact that Hume's Principle fails to explain those numerical terms. And this failure threatens to jeopardize the Neo-Logicist's understanding of the number concept itself, and of the proof of the infinity of the natural numbers—it threatens to defang the Neo-Logicist program entirely.

In the next chapter, we shall consider two prominent attempts to answer the Caesar Problem and evaluate each against our various versions of the problem.

Chapter 3

Solving the Caesar Problem

In this chapter, we critically examine two of the most prominent contemporary responses to the Caesar Problem. The first is due to Hale and Wright. The basic strategy is to appeal to a background metaphysical picture to decide mixed identities; Hume's Principle, against this backdrop, decides the problematic identities after all. The second is gestured at by Dummett, briefly discussed by Linnebo, and given full articulation by James Studd. The basic strategy there is to embrace the indeterminacy of Hume's Principle and find ways to live with it.

We will examine how each solution fares against the various Caesar Problems discussed in the previous chapter. Although our discussion is not decisive, we highlight costs associated with each solution and challenges each must must overcome.

3.1 Hale and Wright

Hale and Wright's favored solution to the Caesar Problem has its roots in Wright (1983). there, Wright lays down the *Grundgedanke* of this line of reply: Caesar and the number 0 can be distinguished by their identity criteria. 0, being a number, is individuated by Hume's Principle, the criterion of identity appropriate for numbers. Caesar, being a person, is individuated in some other way—presumably by reference to conditions of psychological or physical continuity.[1] Since 0 and Caesar are individuated differently, they must, by Leibniz's Law, be different. More generally, the fact that persons and numbers have different criteria of individuation

[1]See Perry (1975) for a survey of criteria of personal identity.

shows that they are two fundamentally different kinds of things—and hence (goes the argument) that no instance of the one is an instance of the other.

All this sounds reasonable enough. Let us unpack it more carefully. Hale and Wright's strategy has developed in two stages. The original version of the solution makes heavy use of a "Sortal Inclusion Principle" to be discussed shortly. Its successor relies unstead on a "category ontology." We will present both strategies in the next few subsections, before considering objections to them. Since both responses make heavy use of the notion of a sortal concept, however, we devote the next subsection to an examination of this concept.

3.1.1 Sortals

The notion of *sortal concept* plays a prominent role in Hale and Wright's solutions to the Caesar Problem (and indeed, in their construal of the Neo-Logicist project). Wright (1983: 2) characterizes sortal concepts as concepts, an understanding of which requires grasp of a *criterion of identity*. As Wright explains, grasp of any concept requires a grasp of that concept's *criterion of application*: "a conception of the differences between things to which it applies and things to which it does not" (ibid.: 2). What sets sortal concepts apart is that grasp of such concepts requires in addition a grasp of what is it for two instances of the concept to be identical or distinct.

Wright adduces some examples to motivate this distinction between sortal and non-sortal concepts. The concept *brown* is not a sortal concept: to possess that concept, it suffices to manifest an implicit understanding of the conditions under which something falls under the concept—paradigmatically, it suffices to be able to tell brown things apart from non-brown things. In particular, it is not required that one in addition be able to tell (or otherwise implicitly recognize) when one brown thing is identical with another. By contrast, *person* is a sortal concept, argues Wright:

> understanding *person* involves more than being able to distinguish persons from things of other sorts; it is essential, in addition, to have a conception of what it is to encounter the *same* person again, and of what constitutes the survival of a person through a given period of

> time. Fully to understand what it is for something to be a person is to have a grasp of the kinds of state of affairs which make contexts of the form 'a is the same person as b true.'
>
> Wright (1983: 2)

Now, one might wonder if this really is a good way to think about concepts like *person*. For, presumably, the criteria of identity for persons are quite complex—as a glance at Perry (1975)'s survey of the options should convince you. It is far from clear that ordinary thinkers possessing the concept *person* know—even implicitly—this criterion. Or consider the concept *electron*. It is a concept you and I certainly possess. Yet I would be hard pressed to articulate conditions under which electron a and electron b are the same electron. Perhaps two electrons are the same iff they occupy the same point in space—but in fact, I have a dim recollection that electrons do not occupy positions in space: they are rather to be thought of as probability distributions over the whole of space-time. Perhaps this is entirely wrong; the point is simply that it is exceedingly implausible that, beneath all of this, I somehow grasp what it is for two electrons to be the same or distinct. And yet, I retain the concept *electron*, and this concept is plausibly sortal in Wright's sense. So there is reason to question Wright's characterization of sortal concepts as those possession of which requires grasp of a criterion of identity.

Wright's characterization may be weakened to: sortal concepts are those *full understanding* of which requires a grasp of a criterion of identity. The idea is that although I may not need to grasp the criterion of identity for persons or electrons in order to think thoughts containing these concepts, my understanding of these concepts is somehow lacking until I do grasp such a criterion. This is more plausible, although developing this line of thought would require specifying what it is to *fully* understand a concept.[2]

Be that as it may, it is in fact not essential to Wright's project that a sharp distinction be drawn between sortal and other kinds of concepts. What matters is that each object falls under at least one sortal concept, and that each sortal be associated with a unique criterion of identity. Defending these two assumptions

[2]Is there a sharp boundary between better understanding the concept *electron* and better understanding electrons? Arguably, one of the lessons of the analytic/synthetic debate is that there is not. But I digress.

would take us too far afield, so we simply record them here. They will play a prominent role in section 3.1.3. To fix terminology, we shall say that an object is *governed* by a criterion of identity if it falls under a sortal with that criterion of identity (with no presumption at this stage that objects are governed by unique criteria of identity).

3.1.2 The Sortal Inclusion Principle

The principle at the heart of Wright's original line of reply to the Caesar Problem is the following:

> (N^d): G is a sortal concept under which numbers fall (if? and) only if there are, or could be, singular terms a and b purporting to denote instances of G such that the truth-conditions of $a = b$ could adequately be explained as those of some statement to the effect that a $1-1$ correlation obtains between a pair of concepts.[3]
>
> Wright (1983: 117–118)

Here is a (hopefully cleaner) restatement, with "G-terms" abbreviating "singular terms purporting to denote instances of G":

> (N^D): if G is a sortal concept that applies to a number, then some identity between G-terms has the same truth-conditions as a statement of equinumerosity between concepts.[4]

Note that N^D requires that G be a sortal concept. We shall explain and discuss the significance of this requirement below.

The "root idea" behind N^D is

> that if instances of one sortal concept [e.g. *person*] are *in fact* instances of another [e.g. *number*], then the truth-values of certain identity statements linking terms for instances of the latter sortal must be

[3] I have made inessential adjustments to Wright's formalism.

[4] The astute reader will have noticed that N^D differs from N^d in two respects: first N^d only requires the relevant G-terms to *potentially* exist. Second, N^d carries the suggestion, absent from N^D, that the truth-conditions of an equinumerosity statement *explain* those of the relevant identity statement. I trust that omitting these aspects of N^d will not bias our discussion of Hale and Wright's solution.

> constituted by, or owed to, the very same facts as determine the truth-values of appropriate identity statements linking terms for instances of the former.
>
> Hale and Wright (2001c: 270)

This explanation of the basic idea suggests the following interpretation: if some numbers are persons, then facts about the identity of those numbers are identical to facts about the identity of those persons; and this identity of *facts* is mirrored in the identity of the truth-conditions of sentences stating those facts.

Now, how does N^D help solve the Caesar Problem? We are invited to consider whether some persons are numbers. Thus we substitute "x is a person" for G. N^D then tells us that some number is a person only if there are person-terms a and b such that $a = b$ has the same truth-conditions as $F \sim H$ for some F and H—that is, only if

$$a = b \equiv F \sim H.$$

But if a and b are names of persons, then the truth-conditions of $a = b$ are given by "a bears the right kind of psychological and physical continuity to b" or something of the sort. That is, we have

$$a = b \equiv a \cong_\Psi b$$

where $\cong_\Psi$ is a placeholder for the correct account of personal identity. Since the truth-conditions of $a \cong_\Psi b$ are manifestly not identical to those of $F \sim G$ for any F and G—that is, since $a \cong_\Psi b \not\equiv F \sim G$—we have

$$a = b \not\equiv F \sim G$$

for any F and G. Hence, no identity between person-terms has the same truth conditions as a statement of equinumerosity between concepts. By N^D, then no number is a person. Moreover, this argument clearly generalizes: if it works at all, it will work for any concept whose identity criterion differs from that for numbers.

What should we make of this response to the Caesar Problem? Besides N^D itself, the argument clearly relies on two additional premises. The first is that $\equiv$ is an equivalence relation, so that if $a = b \equiv a \cong_\Psi b$ and $a \cong_\Psi b \not\equiv F \sim G$, then

$a = b \not\equiv F \sim G$. The second is that, indeed, $a \cong_\Psi b \not\equiv F \sim G$—that is, that $a \cong_\Psi b$ and $F \sim G$ have different truth-conditions, for any F and G.

The first of these assumptions is clearly reasonable, at least within the context of a formal language. The second is more questionable. Must $a \cong_\Psi b$ and $F \sim G$ really differ in truth-conditions? They certainly differ in cognitive significance. But cognitive significance cuts more finely than truth-conditions, as the Neo-Logicist conceives of them. For, as we have seen, $\#F = \#G$ and $F \sim G$ are not cognitively equivalent. Yet for the Neo-Logicist, this is no bar to their having the same truth-conditions. Might not $a \cong_\Psi b$ and $F \sim G$ likewise share truth-conditions? Absent a more precise characterization of truth-conditions, there may remain room for doubt on this score.

Let us suppose, however, that $a \cong_\Psi b$ and $F \sim G$ have different truth-conditions. We must still inquire into N^D itself. Is it true? And if so, is it knowable on the basis of the kind of information available to the Neo-Logicist neophyte? Let us take these questions in turn.

N^D would seem to be unexceptionable. For suppose G is a sortal concept, say *person*—that applies to a number—say to 0. Then 0 is a person. But then "0" is a person-term. And by Hume's Principle, $0 = 0 \equiv \lambda x.x \neq x \sim \lambda x.x \neq x$. Thus, some identity between person-terms has the same truth-conditions as a statement of equinumerosity. Since this reasoning can clearly be generalized to arbitrary sortal concepts G, N^D follows.

This smooth argument, however, is mistaken. It slides illegitimately from the claim that 0 is a person to the claim that "0" is a person-term. This may sounds like like a hair-splitting complaint. But it is not, as we will see by considering a counter-example to a related principle: the Sortal Inclusion Principle.

The Sortal Inclusion Principle is a general metaphysical principle, from which N^D follows as a special case. Here is how Hale and Wright explain the principle:

> (SI): No objects falling under a given sortal concept F can be reputable candidates for identity with objects falling under another sortal concept G, unless […]: to every identity statement "a = b," where a and b purport to have reference among the relevant instances of F, corresponds at least one statement "A = B," where A and B purport

to denote instances of G, which has the same truth-conditions.

Hale and Wright (2001c: 370)

We restate the Sortal Inclusion Principle as follows (and call it SIP to distinguish our formulation from Hale and Wright's):

> (SIP): if sortal concepts F and G overlap, then for any F-terms a and b whose denotations are also Gs, there are G-terms c and d such that $a = b$ and $c = d$ have the same truth-conditions.

Hale and Wright are explicit that SIP is supposed to provide the grounds for N^D. And indeed, N^D follows from SIP together with Hume's Principle. For if we substitute *number* for G in SIP, we get that if a sortal concept F applies to numbers and a and b are F-terms denoting numbers in the overlap, then there are number-terms c and d such that $a = b$ has the same truth-conditions as $c = d$. But by Hume's Principle, the truth-conditions of $c = b$ are those of $G \sim H$ for some G and H. Thus, some identity between F-terms has the same truth conditions as a statement of equinumerosity, as N^D states.

Hale and Wright's raise the following counter-example to SIP (I have changed the specifics, but not the substance, of the example). Consider the following identity statements:

(a) The celestial body last visible in the morning = the celestial body first visible in the evening

(b) The planet last visible in the morning = the planet first visible in the evening

Celestial body and *planet* are overlapping concepts, and both are plausibly sortal. Moreover, as per philosophical folklore going back to Frege (1892), all terms occurring in (a) and (b) denote the planet Venus. Hence, by SIP, there are *planet*-terms c and d such that (a) has the same truth-conditions as $c = d$. Hale and Wright's contention, transposed to our example, is that there is no better candidate for c and d than the terms occurring in (b)—if (b) does not have the same truth-conditions as (a), then no planet-identity statement does. But (b) does *not* have the same truth-conditions as (a), for it cannot be known apriori that they have the same truth-value. Indeed, for all we know apriori, the celestial body last visible

in the morning may be a star rather than a planet and may be different from the first celestial body visible in the evening, even though the corresponding planets are identical—that is, for all we know apriori, (b) could be true and (a) false.

Thus, if this line of objection is sound, SIP is false. And since N^D was supposed to be a consequence of SIP together with Hume's Principle, N^D is robbed of its support.

Can we construct a similar "counter-example" to N^D directly? This is more difficult, since the abstract realm seems less rife with sortal overlap than the empirical world. But if *abstract object* is a sortal concept, the following example will do:

> The abstract object most-often referred to in Newton's *Principia* = the abstract object most-often referred to in Russell's *Principia*

Suppose that, as a matter of fact, the abstract object most-often referred to in these two works is the number 2. Then by N^D there is a statement of numerical identity with the same truth-conditions. The best candidate would seem to be either

> The number most-often referred to in Newton's *Principia* = the number most-often referred to in Russell's *Principia*

or

> $2 = 2$

The former recopies the syntactic form of the original statement but does not use canonical numerical terms (of the form "the number of Fs"), The latter does use canonical numerical terms, but only shares a superficial syntactic form with the original statement. Pretty clearly, however, neither statement has the same truth-conditions as the original statement.

So SIP and N^D are probably false. Returning to our original argument for N^D, we can diagnose its misstep. The issue, as mentioned, is that the argument slides from the claim that 0 is a number to the claim that "0" is a number-term. Recall, however, that a person-term is a term *purporting* to refer to a person, and for this it is neither necessary not sufficient to *in fact* refer to a person. Against

necessity, note that "Homer" is a person-term but does not refer to a person. Against sufficiency: "object in the middle of the room" may refer to a person (in the right context, and so on), but is not a person-term. Whether a term purports to refer to an object of a given kind seems to have little to do with what it in fact refers to. Perhaps it is a matter of the sense of the term, or of the speaker's referential intentions in using it. In any case, the flaw in the argument for N^D given above should now be clear: 0 is simply not a person-term, even if it happens to refer to a person.

There is a further worry concerning this early line of response by Wright. However, since this worry also affects Hale and Wright's updated response, we will present it after expounding that later response.

3.1.3 Category Ontology

In response to the kind of objection canvassed in the previous section, Hale and Wright have moved away from N^D and toward a different, though related, metaphysical picture. On that picture, objects belong to *categories*. A catgory is a "maximally inclusive sortal" (Hale and Wright 2001c: 389). One sortal concept F is a sub-sortal of another concept G if F and G have the same criterion of identity, and every F is G. For example, *tiger* is a sub-sortal of *cat*, which is in turn a sub-sortal of *mammal*, and so on. The relevant category to which all of these concepts belong is perhaps *organism*, or material object. With these concepts in place,

> we can see the outlines of a philosophical ontology in which …all objects belong to one or another of a smallish range of very general categories, each of these subdividing into its own respective more or less general pure sorts [i.e. sortal concepts] …Within a category, all distinctions between objects are accountable by reference to the criterion of identity distinctive of it, while across categories, objects are distinguished by just that—the fact that they belong to different categories.
>
> Hale and Wright (ibid.: 389)

Nikolaj Jang Pedersen helpfully reconstructs Hale and Wright's solution as follows:

> All objects of the ontology fall under sortal concepts, each of which is assumed to have a unique criterion of identity. Among the sortals we find categories, which are maximally extensive sortals in the sense that (i) all sub-sortals of a given category F share their criterion of identity and (ii) for any object x, if x is not F, then any sortal G which x falls under does not share its criterion of identity with F. Internally, objects are unified in a category by falling under sortals that share their criterion of identity, while objects of distinct categories differ in exactly that respect, *viz.* that any sortals they fall under do not share their criterion of identity. Once this picture is in place, it can be shown that any two categories F and G are either co-extensional or have no object in common. …Since Caesar and numbers belong to different categories, Caesar cannot be a number.
>
> Pedersen (2009: 144)

Let us codify the key definitions and assumptions of this solution.[5]

Recall first that a sortal concept is one that is associated with a criterion of identity: F is a sortal concept iff there is an equivalence relation $\sim_F$ such that, for any Fs a and b,

$$a = b \leftrightarrow a \sim_F b$$

Sortal concepts are assumed to have a unique criterion of identity. A *pure* sortal is one whose instances are essentially of that sort. *Number*, of course, is supposed to be a pure sortal: it is of the essence of any given number that it be a number. Now, Hume's Principle does give a criterion of identity for numbers, though not quite of the form above, since the relevant equivalence relation holds not between numbers but between the concepts that the given numbers number. However, we can derive a principle of this form from Hume's Principle. Let $\mathbb{N}$ be a definition of the numbers in FA—for example, $\exists X(x = Nz : Xz)$. Then we have, for any a and b such that $\mathbb{N}a$ and $\mathbb{N}b$, that

$$a = b \leftrightarrow \exists F \exists G(Nx : Fx = a \wedge Nx : Gx = b \wedge F \sim G)$$

[5]My reconstruction of the category ontology below is inspired by Pedersen (2009)'s, though we differ on some of the details.

Note that the relation defined in the right-hand side of this biconditional is an equivalence relation on numbers.

We are interested in different sortals having the same criterion of identity. Sortals F and G have the same criterion of identity, written $\sim_F{=}\sim_G$, iff $\forall x \forall y(x \sim_F y \leftrightarrow x \sim_G y)$—that is, if whenever x and y stand in the equivalence relation for F, they also stand in that for G, and vice-versa.[6] Note that the relation *having the same criterion of identity* is an equivalence relation on (sortal) concepts.

For two sortal concepts F and G, we will say that F is a sub-sortal of G, written $F \sqsubseteq G$, if (i) every F is G, and (ii) $\sim_F{=}\sim_G$. We will write $F \sqsubset G$ for $F \sqsubseteq G \wedge \exists x(Gx \wedge \neg Fx)$—i.e. G is a strictly "wider" sortal than F with the same criterion of identity. We can now define a *category* as a $\sqsubseteq$-maximal sortal. That is, a sortal F is a category if there is no G such that $F \sqsubset G$, or in other words, if any wider sortal G does not share its criterion of identity with F.

With an extra assumption, we can show that a category comprises all and only the objects governed by a given criterion of identity. The assumption is that if a is governed by the criterion of identity $\sim_F$, then the concept G defined by $Fx \vee x = a$ is also a sortal concept and its criterion of identity is $\sim_F$. In other words, whenever a is governed by $\sim_F$, we have $F \sqsubseteq \lambda x.Fx \vee x = a$.

With this assumption in place, we can show that if F is a category, then Fa iff a is governed by $\sim_F$. Suppose F is a category. The left-to-right direction of the biconditional is clear. For the right-to-left direction, suppose toward contradiction that a is governed by $\sim_F$ but does not fall under F. Consider the concept G defined by $Fx \vee x = a$. By the above assumption, we have $F \sqsubset G$. This contradicts the maximality of F under $\sqsubseteq$.

The converse result, that if F comprises all and only the objects governed by $\sim_F$, then F is a category, is readily shown. For consider any sortal G wider than F, i.e. such that all Fs are Gs but $\neg Fa$ for some a. Since Fa iff a is governed by $\sim_F$, a is not governed by $\sim_F$. But then we do not have $F \sqsubset G$. Hence, F is $\sqsubseteq$-maximal, i.e. it is a category.

This characterization of a category as a sortal F whose instances are all and

[6] In a language containing other non-logical vocabulary, it would be necessary to modalize this criterion: $\Box\forall x \forall y(x \sim_F y \leftrightarrow x \sim_G y)$. As long as we remain in the extensional realm of pure mathematics, however, we can dispense with this nicety.

only the objects governed by $\sim_F$ will be helpful below.

Hale and Wright's solution also depends on the following assumption: no object is subject to two distinct criteria of identity. That is, if Fa and Ga for two sortals F and G, then $\sim_F = \sim_G$. This assumption justifies us in speaking of "the" criterion of identity governing some object.

With these assumptions in place, we can argue our way out of the Caesar Problem as follows. First, note that if $\sim_F = \sim_G$ for some sortals F and G, then a is governed by $\sim_F$ iff a is governed by $\sim_G$. For if a is governed by $\sim_F$, then there is some H such that Ha and $\sim_H = \sim_F$. But since $\sim_F = \sim_G$ and identity of crierion of identity is an equivalence relation, we also have $\sim_H = \sim_G$. Thus, a is governed by $\sim_G$. The argument for the converse is exactly symmetrical.

Next, note that categories are either coextensive or disjoint. For let F and G be categories and suppose they are not disjoint. Then there is some a such that Fa and Ga. By our assumption that no object is subject to distinct criteria of identity, we have that $\sim_F = \sim_G$. Now, since F is a category, it comprises all and only the objects governed by $\sim_F$. And since $\sim_F = \sim_G$, F also comprises all and only the objects governed by $\sim_G$. But since G is a category, it too comprises all and only the objects governed by $\sim_G$. Thus, F and G are coextensive.

Next, note that if a and b are subject to different criteria of identity, they belong to disjoint categories. For let a belong to category F and b to category G. Since categories comprise all and only the objects with a given criterion of identity, and since a and b are subject to distinct criteria of identity, we have $\neg Fb$ and $\neg Ga$. Thus F and G are not coextensive. But then by our previous remark, they are disjoint.

It follows that if a and b are subject to different criteria of identity, then they are distinct.

To complete our solution to the Caesar Problem, it remains only to note that Caesar and zero have different criteria of identity. Zero's is given by Hume's Principle, while Caesar's is given by (something like) psychological continuity. Hence, Caesar $\neq 0$.

3.1.4 Evaluation

What should we make of Hale and Wright's sortal-based responses to the Caesar Problem? In considering this question, we will focus on the later, category-based version of the response. But much of what we say will carry over to the earlier, SIP-based solution.

Let us start by considering how Hale and Wright's solution fares in responding to each version of the problem we identified in Chapter 2.

Recall that the epistemic Caesar Problem is the problem of explaining our knowledge that Caesar and zero are distinct. Hale and Wright's solution, if otherwise sound, provides *one* way of coming to know that Caesar and zero are distinct. But it does not thereby explain *our* knowledge that Caesar and zero are distinct. To do so, the solution would have to be, at some level, accessible to ordinary knowers of that truth. And, contrary to what Hale and Wright sometimes suggest,[7] it is doubtful that our knowledge of the distinctness of Caesar and zero is grounded in a grasp (perhaps implicit) of their solution. For even fairly young children know that Caesar and zero are distinct. Yet such children are years, if not decades, away from the conceptual sophistication to frame—let alone knowledgeably believe—Hale and Wright's category ontology. After all, as Hale and Wright themselves state, the Sortal Inclusion Principle "amounts to a substantial metaphysical claim about when the extension of one sortal may properly be regarded as included in—or as overlapping with—that of another" (Hale and Wright 2001c: 370). Likewise, the category-based ontology expounded in the previous section is also a piece of "substantial metaphysics"—nontrivial and conceptually very sophisticated. We thus have every reason to think that Hale and Wright's solution is not what grounds children's knowledge that Caesar is not a number. And even for practicing mathematicians—people who could, in principle, understand the intricacies of Hale and Wright's category ontology—it seems highly unlikely that such an understanding actually plays any role in their knowledge that Caesar is not zero. Indeed, the only people to whom we can confidently ascribe knowledge of the category ontology are people like Hale and Wright themselves—trained

[7]For example: "It is surely because we already incohately think in terms such as these [in terms of a category ontology] that it strikes us as just obvious that Caesar is no number" Hale and Wright (2001a: 390).

metaphysicians! In a slogan, the present criticism of Hale and Wright's solution is that it requires every mathematician to be a metaphysician. This is too strong a requirement on arithmetical knowledge.

How could Hale and Wright respond to this objection? It seems to me that two lines of response are open. First, they could insist that, even though it is only *explicitly* known (or at any rate, believed) by a select few metaphysicians, the category ontology is nontheless *implicitly* grasped by many people, and hence does underlie our everyday knowledge that Caesar and zero are distinct. Second, they could instead give up on the attempt to characterize the grounds of *our* arithmetical knowledge and instead aim to exhibit *a* possible route to such knowledge.[8]

The main challenge facing the first option is to articulate the operative notion of implicit representation or implicit grasp. What exactly is it for someone to implicitly grasp the category ontology?

One possiblity is that the category ontology is known, but that the corresponding beliefs are not conscious, in the sense that they are not available to introspection. On this charaterization, however, grasping something implicitly requires exactly as much conceptual sophistication as grasping it explicitly—the difference is in whether the grasp is accessible to conscious reflection, not in the sophistication of the concepts employed. The same goes for a related proposal on which to believe implicitly is to be disposed to believe explicitly. Unless the notion of a disposition is stretched beyond recognition, you cannot be disposed to believe that which you lack the conceptual resources to entertain.

Another possibility is to cash out implicit belief as *dispositional* belief. The idea is that to implicitly believe that p, one need not represent p—consciously or not—or even be disposed to represent p. Instead, one must be disposed to act, roughly, *as though* one believed that p.[9] Thus, perhaps laypeople implicitly grasp the category

[8]There is a third option. Pedersen (2009) has shown that the full strength of the category ontology is not required to implement a solution to the Caesar Problem in the spirit of Hale and Wright. The key feature of categories used in Hale and Wright's solution is that they are equivalence classes under the relation *sharing a criterion of identity*. Thus, we can do away with the assumption that categories exist, and simply use these equivalence classes. While the resulting theory is somewhat weaker (and hence more secure), its conceptual sophistication remains far higher than is accessible to a child. So the argument in the text is untouched by Pedersen's improvement.

[9]Whether the apparent circularity in this formulation can be finessed is a serious question. See Schwitzgebel (2013) and Warren (2020a: 396–397) for discussion.

ontology by being disposed to act as though it were true (whatever exactly this means). This interpretation also faces difficulties, however. In particular, an individual's knowledge of Hale and Wright's category ontology is supposed to be part of an explanation of their arithmetical knowledge. Knowing the ontological theory is supposed to provide warrant for the belief that Caesar is not a number. But it is entirely unclear how *acting as though* one believed that p could possibly generate the warrant one would have if one (warrantedly) believed that p.

I conclude that, Hale and Wright's assertion to the contrary notwithstanding, the category ontology is not implicitly believed by most knowers of arithmetic. *A fortiori*, it is not a good account of how such knowers know that Caesar is not a number. Let us therefore turn to the second line of reply: What if the Neo-Logicists lower their sights and settle for showing only that someone *could* come to know that Caesar is not a number on the basis of their ontology?

It seems to me that such a result would be of limited significance. For it would boil down to the following: knowing that the concepts *person* and *number* are associated with different criteria of identity and that concepts associated with different criteria of identity do not overlap puts one in a position to know that no number is a person. This is surely correct, but amounts to little more than *modus ponens*.

There is a distinct but related worry lying nearby. Recall that one of the basic tenets of Neo-Logicism is that arithmetical truths are analytic, at least in the sense of being knowable without "collateral epistemic work" (Hale and Wright 2009 459, see also chapter 1, section 1.3.2 above). How should we square this with the apparent substantiality of Hale and Wright's category ontology?

We are faced with a dilemma. Either (i) knowledge of the category ontology is not required for knowledge of arithmetic, or (ii) the category ontology is knowable without collateral epistemic work.

So, which, if any, of (i) and (ii) holds? Let us consider each in turn. To evaluate (i), we must first clarify a point of terminology. The Neo-Logicists hold that *arithmetical* truths are analytic. Do such truths include Caesar identities? Or are they instead confined, roughly, to what the mathematician would recognize as properly arithmetical statements? If the latter, then the objection we are considering does not get off the ground: the machinery of categories is wheeled in specifically to

decide Caesar questions. If these statements are not supposed to be analytic, then it is no problem if deciding them requires heavy-duty metaphysics.

This line of response may be appropriate when tackling the first two variants of the Caesar Problem. However, it is unavailable when facing the third. Recall that the distinctive feature of that version of the Caesar Problem is that it is a global, rather than local, malady: Hume's Principle's inability to provide truth-conditions for mixed identities jeopardizes its claim to endow speakers with an understanding of the "number of" operator entirely. On that construal, a solution to the Caesar Problem is required to have *any* arithmetical knowledge on the basis of Hume's Principle. In answering this third version of the Caesar Problem, then, Hale and Wright cannot excuse the use of the category ontology as merely peripheral or otherwise unnecessary to the main explanatory task. Knowledge of the category ontology will be required for knowledge arithmetic. (i) is therefore false.

Let us consider (ii). Is the category ontology knowable without collateral epistemic work? Here, the answer seems to be a clear "no." Substantial metaphysical theses are in general not knowable (for most are false), let alone knowable without collateral epistemic work. On the contrary, painstaking care is needed if one is not to fall into error. We can therefore dismiss (ii).

Could Hale and Wright reply by distinguishing between, on the one hand, what someone must *know* (in addition to Hume's Principle) in order to know arithmetic truths, and, on the other, what must be *true* if someone is to gain knowledge of arithmetic in this way? Perhaps Hale and Wright's category ontology falls into this latter category: it serves as backdrop for the speaker's use of sortal concepts, something he implicitly relies on in his cognitive practice but which does not need to be cognitively available to him. After all, the fact that our knowledge depends—depends *epistemically*, not (merely) causally—on factors outside of our psychology has been one of the central insights of externalist epistemology (see e.g. Burge 2020). And Wright himself is one of the leading voices in this style of epistemology (Wright 2004).

Well, suppose that this is the case: that Hale and Wright's category ontology must be true, but needn't be known, for someone to know that Caesar and zero are distinct. One problem with this idea is that it leaves the Neo-Logicist subject without a route to this knowledge. *Ex hypothesi*, he does not have an argument

for the conclusion that Caesar and zero are distinct. But not just any way of stumbling into the belief that Caesar and zero are distinct will result in knowledge (forming the belief as a result of a blow to the head does not count). One's belief must be minimally sensitive to the truth for it to count as knowledge. Whence this sensitivity?

Let us turn to the problem of semantic indeterminacy. Recall that this version of the Caesar Problem calls upon the Neo-Logicist to explain how one could achieve a grasp of a determinate concept of number on the basis of Hume's Principle—one no less determinate than our actual concept of number, at any rate. Hume's Principle is indeterminate along two dimensions: horizontal and vertical. The vertical indeterminacy concerns the number of numbers: how high up do the cardinalities furnished by Hume's Principle go?[10] The horizontal indeterminacy concerns the identity of numbers: are they a *sui generis* kind of entity, or are some of them identical to already-familiar entities—roman emperors, real numbers, etc.? We argued in the previous section that vertical indeterminacy poses no threat to the Neo-Logicist. By contrast, horizontal indeterminacy plausibly does: Hume's Principle by itself delivers a concept of number less determinate than our own. So how does Hale and Wright's category ontology fare as a solution to *that* problem?

The answer is: all too well. The category ontology renders the number concept delivered by Hume's Principle highly (horizontally) determinate. In particular, it likely renders it at least as determinate as our concept of number. But there is a risk that the concept is thereby rendered *too* determinate: the category ontology decides questions that are arguably left open (i.e. indeterminate) by our actual concept of number, such as the identity or distinctness of 0_{HP} and $0_{\mathbb{R}}$. We shall examine below whether this is an issue for the Neo-Logicist. First, let us explain how the category ontology yields a determinate concept of number.

Given the characterization of categories given in the previous section, the following holds: whenever we have Fa and $\sim_F \neq \sim_{\mathbb{N}}$ for some sortal F, we will be able to conclude that $\neg \mathbb{N}a$. Thus, whenever it is determinate that $\sim_F \neq \sim_{\mathbb{N}}$, it will be determinate that $\forall x(Fx \rightarrow \neg \mathbb{N}x)$. Conversely, if it is determinate that $\sim_F = \sim_{\mathbb{N}}$, then for each a amongst the Fs and each number n, it will be determi-

[10] Recall that we are officially agnostic about whether the numbers to which Hume's Principle purports to give us access are *bona fide* cardinalities. See chapter 1, footnote 4.

nate whether $a = n$ (by reference to the criterion of identity). The question thus becomes, for which F is it determinate whether $\sim_F = \sim_{\mathbb{N}}$? If the answer is: for all sortals F, then the resulting concept of number will be perfectly (horizontally) determinate. However, if for some sortal F, it is indeterminate whether $\sim_F = \sim_{\mathbb{N}}$, then it will likewise be indeterminate whether $\exists x(Fx \wedge \mathbb{N}x)$. So, is identity of criteria of identity generally determinate?

Recall that $\sim_F = \sim_G$ abbreviates $\forall x \forall y(x \sim_F y \leftrightarrow x \sim_G y)$, which holds just in case $\sim_F$ and $\sim_G$ have the same equivalence classes. Thus, a counterexample would be an F such that, on certain interpretations, $G \sim_F H$ just in case G and H are equinumerous, while on others $G \sim_F H$ for some non-equinumerous G and H. Unfortunately, if anything is indeterminate, then such concepts are plentiful. For let φ be any indeterminate sentence (such as "Terry is bald", for a suitable Terry) and consider the following abstraction principle

$$Bx : Fx = Bx : Gx \leftrightarrow (F \sim G \wedge \varphi) \vee (\bot \wedge \neg\varphi).$$

Call the relation on concepts defined by the right-hand side $\sim_B$. Observe that on interpretations making φ true, $\sim_B$ coincides with $\sim_{\mathbb{N}}$, while on interpretations making φ false, $\sim_B$ is the empty relation. By our assumption that φ is indeterminate, interpretations of both kinds exist. Thus, it is indeterminate whether $\sim_B = \sim_{\mathbb{N}}$. What's more, this indeterminacy trickles down to indeterminacy of objects. For let b be $Bx : \lambda x.x \neq x$. Do we have $b = 0$? According to the general strategy under consideration, if $\sim_B \neq \sim_{\mathbb{N}}$, then $b \neq 0$, while if $\sim_B = \sim_{\mathbb{N}}$, then $b = 0 \leftrightarrow \lambda x.x \neq x \sim_{\mathbb{N}} \lambda x.x \neq x$, and so indeed $b = 0$. But it is indeterminate whether $\sim_B = \sim_{\mathbb{N}}$! Hence, it is indeterminate whether $b = 0$.

Thus, so long as there is some indeterminacy in the language (as surely there is), the category ontology cannot deliver a perfectly determinate concept of number. Is this a problem? It seems to me that it is not. The problem raised by the version of the Caesar Problem under consideration is not that Hume's Principle delivers a concept of number with some indeterminacy. The problem is that it delivers a *wildly* indeterminate concept of number—one on which it is not even determinate whether Caesar is a number! Solving this problem does not require dispelling all possible sources of indeterminacy in the concept furnished by Hume's

Principle. It requires securing only so much determinacy as our actual concept of number enjoys. And it is not clear, to say the least, that our concept of number *is* (and hence not clear whether the Neo-Fregean concept of number should be) determinate with respect to objects like b.

To summarize the previous few paragraphs: even helping ourselves to the category ontology, we cannot show that the concept of number given by Hume's Principle is perfectly (horizontally) determinate. However, we can show that it is determinate with respect to any sortal F such that it is determinate whether F and $\mathbb{N}$ have the same criterion of identity. Arguably, our actual concept of number enjoys no more determinacy than this, and so the failure of the category ontology to supply more determinacy does not discredit it.

However, we must consider whether the category ontology does not render the Neo-Fregean concept of number *too* determinate. In particiular, suppose, as seems plausible, that it is determinate whether the concept *real number* shares a criterion of identity with the Neo-Fregean concept of number (I imagine that it does not). Then the category ontology will ensure that it is determinate whether $0_{\mathbb{R}} = 0_{\mathsf{HP}}$. Likewise for the questions whether $0_{\mathsf{HP}} = \emptyset$, $0_{\mathsf{HP}} = 0_{\mathbb{Q}}$, and the like. But it might be thought that, on our intuitive concept of natural number, these identities are *not* determinate. If they are not, then we should be free to develop our mathematical and linguistic practices in different directions. On some such directions, 0_{HP} will be identical to $0_{\mathbb{R}}$, while on others it will not. If the category ontology decides such questions in advance, then it might objectionably curtail the theoretical freedom of mathematicians.

This objection is swiftly answered, however. Hale and Wright can respond that if indeed 0_{HP} and $0_{\mathbb{R}}$ do not share their criterion of identity, then mathematicians had no such conceptual freedom to begin with.

There is a more concerning objection nearby, however. The worry is that the category ontology might determine the *wrong* truth-value for $0_{\mathsf{HP}} = 0_{\mathbb{R}}$. For this identity will only be true if the two numbers share their criterion of identity. But it is hard to see how real numbers could have Hume's Principle as their criterion of identity—Hume's Principle ties its numbers to *counting*, whereas real numbers are plausibly associated with *measuring*. And while one-to-one correspondence may be at the heart of counting collections, it is hard to see what role it could

play in measuring—especially in measuring continuous magnitudes, the bread-and-butter of real numbers. Indeed, to my knowledge, no abstractionist account of real numbers gives equinumerosity a central role.[11] So if we follow this line of response, it will be difficult to avoid the conclusion that $0_{\mathsf{HP}} \neq 0_{\mathbb{R}}$. And this conclusion might seem untenable. For are the real numbers not simply the completion of the rational numbers, themselves the completion of the integers, in turn the completion of the natural numbers, each under the appropriate algebraic operation? And do not completions result in structures containing what they complete?

This objection is not conclusive, however. First of all, mathematical practice does not support the view that these extensions literally contain their base domains, as opposed to an isomorphic copy thereof (Bellomo 2021). Secondly, it is not clear that the bullet biting response is so bad. On the contrary, the Neo-Logicist can turn the table on the objector: the fact that real and natural numbers are associated with different criteria of individuation reflects the fact that they have fundamentally different uses. As such, we should not expect the two kinds of numbers to overlap.

I conclude that the category ontology fares rather well as an answer to the problem of semantic indeterminacy.

Finally let us examine the category ontology as a response to the problem of linguistic understanding. Can the category ontology help secure an understanding of the number concept, where Hume's Principle alone fails?

Note first of all that all of the problems raised in connection with the epistemic problem will recur here. For the category ontology cannot do the work of constituting a speaker's understanding of the number concept if it is not available in some form to the speaker. But let us not rehearse these worries, and grant instead that a speaker has access to the category ontology. Does he thereby grasp the contribution made by $\lambda x.x = Ny : Fy$ to sentences in which it occurs?

Whether he does or not depends on a further question: does understanding a singular term require knowing what kind of thing it refers to? If so, then for any singular term a, our speaker will know the criterion of identity for things of that kind, and will be able to compute the truth-conditions of $a = Nx : Fx$: if the

[11] For examples of such accounts, see Hale (2001b) and Shapiro (2000). See also Wright (2000) for discussion.

criterion of identity for a is given by Hume's Principle, then the sentence is true just in case the concept which a numbers is equinumerous with F, while if its criterion of identity is not given by Hume's Principle, then the sentence will be false.

Now, we are not in a position to answer this question. But we note that an affirmative answer seems to be presupposed by the Neo-Logicist project anyway (note also that many philosophers with no particular affiliation with Neo-Logicism have also argued for an affirmative answer, e.g. Evans 1982). For as we saw in section 1.2, reference to numbers requires a grasp of the criterion of identity for numbers. If this claim can be generalized to other sortals, we will have come much of the way toward defusing the third version of the Caesar Problem. But an in-depth examination of this issue would take us beyond the scope of our argument. So we will be content to record that the success of the category ontology as a response to the problem of linguistic understanding stands or falls—modulo problems about access to the category ontology—with the claim that, whenever a speaker uses a singular term, he is in a position to determine some criterion of identity governing what the term purports to refer to.[12]

3.2 Indeterminism

Indeterminism is an entirely different line of response to the Caesar Problem. On this view, mixed identities are indeed left unsettled by Hume's Principle. But this is not viewed as a problem: these identities can be settled in other ways, and linguistic conventions play a prominent role in such resolutions.[13]

An early defender of this proposal is Linnebo (2018b). Although Linnebo does not spell out a full indeterminist account, he seems implicitly committed to some

[12]Note that if there is to be any hope of success here, definite descriptions will have to be ruled out as singular terms. For it is not plausible to maintain that anyone who uses the description "the thing at the center of the next room" knows the criterion of identity for that thing, whatever it may be.

[13]In earlier drafts of this thesis, this position was called "stipulationism." I have come to think that this is not a good name for it, however. The distinguishing feature of the view is not that mixed identities are a matter of stipulation. It is rather that Hume's Principle and the associated abstractionism can tolerate some indeterminacy, and that this indeterminacy can be resolved in different ways. How precisely the indeterminacy is resolved—whether by explicit stipulation, novel patterns of language use, or metaphysical reflection—is not a core part of the view.

such view. He writes

> What about Caesar-style identity statements such as 'this letter token = that letter type'? In this particular case, it is natural to invoke the uniqueness thesis and this to regard all such mixed identity statements as false. As we have seen, doing so is logically and conceptually unproblematic. Let us therefore write into the assertibility conditions [governing the language under discussion] that no letter type is to be identified with any "old" object that was present already in the domain associated with the [pre-abstraction] language …In other cases, however, it is far more natural to identify some of the "new" abstracts with an appropriate "old" object. Recall plural Law V …:
>
> $$\{xx\} = \{yy\} \leftrightarrow \forall z(z \prec xx \leftrightarrow z \prec yy)$$
>
> Suppose the domain on which we abstract already contains some sets. Then we have good reason to identify a "new" set with an "old" set that has precisely the same elements. Suppose, for example, that the domain of the base language contains the three objects a, b, and $\{a, b\}$. Then we have good reason to identify the "new" set of a and b with the "old" one.
>
> …
>
> The conceptual freedom emphasized throughout this chapter ensures that we need not give a uniform answer to the question of whether or not the "new" objects introduced by some abstraction should be identified with any of the "old" objects that were available prior to this abstraction. Different cases will call for different answers.
>
> Linnebo (2018b: 170–171)

There are two incompatible ways of taking these remarks. One, suggested by the last sentence of the quotation, is that our chosen abstraction principles implicitly settle Caesar questions. Although the answers to these questions are not literally derivable from the abstraction principles, the latter give us enough of a handle on their subject matter—on the nature of the abstracts they introduce—for us

to decide, by general philosophical reflection, a range of Caesar questions. This interpretation is in line with the general orientation of Hale and Wright's response, since it takes Caesar questions to be implicitly resolved by the adoption of the relevant abstraction principles. But it differs in holding that we can answer these questions not by using a general metaphysical principle, such as the Sortal Inclusion Principle or N^D, but through considerations specific to the case at hand. For example, reflecting on the fact that directions are abstract and lines are concrete (if that is indeed a fact) might militate against identifying directions with lines. Likewise, as Linnebo suggests, the extensionality of sets provides a powerful reason to identify the abstract of the plurality of a and b with the already-existing set $\{a, b\}$, since they have the same members.[14]

On this interpretation, there is a matter of fact about whether the number 0, as given by Hume's Principle, is identical to Caesar or is a real number. Another interpretation of Linnebo's remarks holds that mixed identities are not decided by our chosen abstraction principles, even implicitly—that abstraction principles simply do not determine answers to Caesar questions. Instead, Caesar questions represent choice points in our conceptual or theoretical development. We are free to identify or distinguish abstracts and non-abstracts, or abstracts from different abstraction principles. Thus, we may choose to identify each letter type with a particular token of that type, or we may declare that letter types are never identical to letter tokens. In such choices (as in all others), we are guided by considerations of theoretical simplicity, consistency, unity, and so on. But these considerations are not guiding us to an antecedent truth of the matter. Instead, they help us to make a theoretically and practically convenient choice. Until we make such choices, the relevant concepts are not fully deterinate: there is no fact of the matter about whether 0_{HP} is Julius Caesar, or $0_{\mathbb{R}}$.[15]

[14]Linnebo is considering an abstraction principle that allows us to form sets from pluralities: $\{xx\} = \{yy\} \leftrightarrow \forall x(x \prec xx \leftrightarrow x \prec yy)$. $\{\cdot\}$ is the set-of operation, variables xx and yy range over pluralities (that is, their values on a given valuation are one or more objects, not an object that is a plurality), and $x \prec xx$ means that x is one of the xx.

[15]A terminological note. We will refer to these kinds of choices concerning Caesar questions as "stipulations," though this is only for convenience: stipulations by themselves are, to use Jared Warren's phrase, "metasemantically idle." What matters for metasemantics is the use of an expression (understood very broadly, to encompass, for example, connections to perception, action, and the environment—see Greenberg and Harman (2005) for a comprehensive overview). Stipulations are relevant only through their influence on use. For example, if I stipulate that

Linnebo's mention of our "conceptual freedom" leads me to suspect that he would prefer the second interpretation. And in any case, it is clearly the second interpretation that offers the most radical break from traditional solutions to the Caesar Problem. So let us examine it in more detail.

Unfortunately, Linnebo does not elaborate his stipulationism further. He therefore leaves a number of questions open. For example: can we stipulate whatever we want, or are there some constraints on our stipulative activity? What if our stipulations conflict with antecedent facts—if Caesar is not a number, could we somehow stipulate that he is? And how, precisely, does all this stipulation answer the various Caesar Problems canvassed in chapter 2?

To gain some clarity on these questions, let us turn to the work of James Studd. Studd is another defender of stipulationism. According to Studd, "Caesar questions left unanswered by [Hume's Principle] or other abstraction principles may be decided by further stipulations" Studd (2023: 253). Thus, along with Hume's Principle, we may stipulate the truth or falsity of any statement of the form

$$a = Nx : Fx$$

undecided by Hume's Principle. When we do so, we further constrain the meaning of the "number of" operator so as to make our stipulations true (if possible).

One of the attractions of this view is that it relies on the very same traditional conception of meaning determination and apriori knowledge on which the stipulation of Hume's Principle itself depends (see section 1.3.2 above). If Hume's Principle can be the target of stipulation, and if stipulating it leaves some work undone, why not continue to stipulate? The very same metasemantic mechanisms that ensured the (partial) success of our first stipulation will also ensure the success of our later ones.

"zorg" is to mean "distant star" but refuse to call any distant stars zorgs and freely exchange "zorg" and "coffee grounds" in conversation, then, my stipulation notwithstanding, "zorg" means (in my idiolect) "coffee grounds" and not "distant stars." Stipulations can communicate patterns of use, or intentions for use. But they do not ground meanings directly (see Warren 2020b §1.II for elaboration). Note, however, that the objects of stipulation—sentences—have a propositional structure, whereas patterns of use do not. What, then, is the pattern of use that is "summarized" in the stipulation that Caesar is not a number? This is a difficult question. One possible answer is the inferentialist view, which reads such stipulations as summaries of inferential dispositions (Brandom 1994, 2000). We leave the matter open.

Evaluating indeterminism

How does indeterminism fare as a response to the Caesar Problem?

Consider first the epistemic problem. Recall that that problem was to explain our knowledge of certain mixed identities and predications, such as that Caesar and zero are distinct. What explanation does indeterminism offer for this knowledge?

The standard indeterminist answer is as follows. Someone who has only mastered Hume's Principle does *not* know that Caesar is not a number. To know this (when starting out from Hume's Principle and nothing else), it is both necessary and sufficient to stipulate that "Caesar is a number" is false. Only on the basis of such stipulations can one know that Caesar is not a number, because only such stipulations can settle the truth of "Caesar is a number."

Is this a satisfactory solution to the epistemic Caesar Problem? As usual, the first question to ask is whether the solution is overly demanding. At first sight, this might seem to be the case: did everyone who knows that Caesar is not a number really have to make a stipulation to that effect?

This objection is not quite well-posed, however. As noted in footnote 12, stipulations provide summaries of patterns of language use. It is these patterns, and not the stipulations *per se*, that ground meaning. With this in mind, the indeterminist solution to the epistemic Caesar Problem is really that if one *uses* the expressions "Caesar" and "the number of" in the ways specified by Hume's Principle and "Caesar is not a number," then one is automatically in a position to know the latter sentence.

To muster a convincing solution to the epistemic Caesar Problem, then, the indeterminist must give some account of how one could be in a position to know that Caesar is not a number simply on the basis of one's use of the relevant expressions. There are two ways in which this could go. First, one might reason from an explicit recognition that one is using the expressions in the right way. That is, one might realize that one's use of the expressions "Caesar" and "zero" is such that, thanks to general metasemantic principles, one's utterances of "Caesar is not zero" express a truth. By this point, it should be clear that this is inadequate as an account of our knowledge of Caesar identities. It is far too much to ask of ordinary knowers of that truth that they have a reflective grasp of the metase-

mantic principles underlying the meanings of their expressions. And in light of our discussion of laypersons' knowledge of the category ontology in the previous section, it should be clear that an implicit grasp of such a metasemantic theory will fare no better.

The alternative is that the truth of the metasemantic picture suffices to warrant belief in Caesar identities, without being accessible to the individual believers of such truths. The idea here is that such beliefs are "blind but blameless": blind because not justified by anything in the speaker's psychology, but blameless because constitutive of the meaning of the expressions used by the speaker. The idea that some beliefs or inferences could be blind but blameless has been used to explain our entitlement to believe logical truths or to make logical inferences (see Boghossian 2003; Warren 2020b chapter 6). In that context, the idea is that our entitlement to infer in accordance with *modus ponens*, for example, derives from the fact that that rule of inference is constitutive of the meaning of the conditional. But if our stipulation that Caesar is not a number is likewise constitutive of the meaning of "number," we might also be entitled to accept such sentences despite lacking explicit justification for them.

Obviously, the viability of this position depends on the details of the "blind but blameless" epistemology, which are beyond the scope of this thesis to investigate. We merely flag the view as an interesting option for the indeterminist to explore. One attractive feature of this response, however, is that it predicts the sense of befuddlement one might feel upon being asked to justify one's belief that Caesar is not a number. If our warrant for this belief is indeed blind, we should expect to struggle to articulate it.

What about the problem of semantic indeterminacy? According to this version of the problem, Hume's Principle fails to secure a suitably determinate conception of the natural numbers. In particular, Hume's Principle leaves the identity of numbers unspecified.

The indeterminist solution to this version of the problem is simply to stop worrying and love the bomb. The indeterminist cheerfully accepts that, left unadorned, Hume's Principle leaves us with a rather indeterminate conception of the natural numbers. If we have a determinate conception of the natural numbers, this is thanks to linguistic and conceptual resources that outstrip Hume's Principle:

"If in fact many interesting Caesar questions are answered [i.e. have determinate truth-values], it is because our use of number terms goes far beyond a commitment to [Hume's Principle]" to incorporate, for example, postulates rendering "Caesar = 0" false (Studd 2023: 259).

How convincing is this response to the model-theoretic Caesar Problem? From the perspective of a traditional Neo-Logicist, one for whom Hume's Principle does most of the heavy-lifting, this solution is bound to seem too concessive. For it admits that Hume's Principle is powerless to secure even a modicum of determinacy. While the Neo-Logicist is perhaps free to believe that *some* mixed identities are indeterminate—for example, $0_{\mathsf{HP}} = 0_{\mathbb{R}}$—he cannot allow much more indeterminacy in his concept of number than is present in ours. Widespread indeterminacy would undermine the Neo-Logicist's claim that Hume's Principle is a route to knowlegde of the natural numbers.

From a more neutral perspective, however—one not antecedently committed to Hume's Principle doing all the work—indeterminism is an attractive response to the model-theoretic Caesar Problem. This is because it is a very flexible solution: however determinate or indeterminate our concept of number may be, it will always be possible to adjust the abstractive concept of number to coincide in determinacy with the ordinary one. Indeed, it seems like this strategy is bound to succeed: for however determinate our concept of number is, some fact about our mathematical or linguistic practice makes it so. And what is to prevent the indeterminist from modeling this aspect of our practice by some stipulation that achieves the same effect?

Finally, let us consider how indeterminism fares as a response to the problem of linguistic. Recall that on this interpretation, by failing to provide truth conditions for mixed identities, Hume's Principle fails to ground an understanding of the "number of" operator, and ultimately of the concept of number. The distinctive challenge posed by this version of the Caesar Problem is that it does not tolerate any unexplicated sentences containing number terms: all such sentences must be assigned a truth-condition from within the language.

It is not clear that indeterminism provides a satisfactory answer to this version of the Caesar Problem. For while specific gaps in Hume's Principle's assignment of truth-conditions may be filled by stipulations, as long as some such gaps remain,

the problem will go unanswered. Thus, unless we have provided a comprehensive package of stipulations, we will be left not with an indeterminate concept of number, but with no concept of number at all.

3.3 Summary

Let us review our two solutions. We argued that Hale and Wright's category ontology provides a convincing response to the semantic indeterminacy problem. However, its viability as a response to the epistemic problem and the problem of linguistic understanding is vitiated by the demands it makes upon would-be knowers of arithmetic. Indeed, one of the main threads of our argument has been that, in addressing these problems, there is a serious risk of *over-intellectualization*: of requiring highly sophisticated conceptual resources to secure knowledge that is in fact available to ordinary thinkers as well as children. While these concerns do not categorically rule out the category ontology approach to the Caesar Problem—after all, it is possible for a defender of this approach to retreat to the weaker claim that the solution merely provides a "proof of concept" for the knowability of such truths—they need to be addressed if the full ambitions of the Neo-Logicist program are to be sustained.

The situation is less clear cut with the indeterminist response, perhaps because that position is itself less fully developped than Hale and Wright's. It is a pleasingly flexible and ecumenical idea. But to provide a convincing response to all three Caesar Problems, it must flesh out and defend a number of theoretical assumptions. Chief amongst these are the connection between stipulations, conceived of as propositionally structured entities (such as sentences), and patterns of language use, which are manifestly different kinds of "entities." For it is the latter that do the real metasemantic work (or so we argued). The task is therefore to explain what pattern of use could result in an assignment of meanings to "number," "zero," and the like, that would render mixed identities determinate. Another important challenge, this time concerning the epistemic problem specifically, is to articulate the connection between such meaning-determining patterns of use and knowledge of mixed identities. We hinted at one way such an explanation could go—our acceptance of mixed identities could be blind but blameless—but work

clearly remains to be done.

In closing we note that the two solutions are—unfortunately—incompatible. For as MacBride (2006) notes, the indeterminist position presupposes that mixed identities are semantically indeterminate. If they are determinate, then there is a fact of the matter about whether Caesar is zero. And stipulations are powerless to settle that fact one way or another, or to yield knowledge of it. Now, there is a powerful argument that such identities *are* indeterminate, at least if Hume's Principle is left unsupplemented—that was the burden of section 2.3. But as we saw, the strength of the category ontology lies in its response to the problem of indeterminacy. If the category ontology is correct, there is in fact rather little indeterminacy to the concept of number. Thus, for the indeterminist solution to get off the ground, the category ontology must fail.

Conclusion

The aim of this thesis has been two-fold. First, to articulate in detail various versions of the Caesar Problem, explaining clearly the different threats they pose to Neo-Logicism. We identified three versions of the Caesar Problem: the epistemic problem, the problem of semantic indeterminacy, and the problem of linguistic undersanding. If Neo-Logicism is to live up to its ambition to explain the nature of our grasp of numbers and the grounds of our knowledge of arithmetic, all three problems must be answered. We argued, however, that the third problem is the most stringent: failure to answer the first two would mean that Neo-Logicism falls short of its goals; failure to answer the third would prevent from Neo-Logicism from even getting off the ground.

Our second goal has been to critically survey extant approaches to the Caesar Problem. The point has not been to argue for one solution at the expense of another, but instead to highlight the strengths and weaknesses of each and to identify specific issues which each must address if it is to be viable. A recurring cause of concern has been over-intellectualization: so long as a solution to the Caesar Problem is to explain *our* grasp of number, it cannot impute conceptual resources which most laypersons and mathematicians lack.

Another theme has been the proper aim of Neo-Logicism. At several points the Neo-Logicist has been able to dodge, or lessen, the blow of an objection by retreating to a weaker claim: instead of explaining our grasp of numbers, Hume's Principle can explain how it could be *possible* for *something* (a pure rational agent?) to have a concept of number. In closing, we wish to highlight just how costly this reaction is. We have been talking somewhat loosely of "a concept of number," "the concept *number*," "our concept of number," and so on. It is common ground in the debate,

however, that concepts are constituents of propositions.[16] If two propositions contain different concepts, then they are different propositions. Thus, if the target of inquiry is our knowledge of arithmetic truths, it is essential that the concepts that figure in such truths be our concepts. It will not be possible to provide a "different route" to the same truths if the route goes via a different concept of number than the one we actually possess. In particular, the following is not a tenable position: Hume's Principle secures a concept of number less determinate than ours, but the indeterminacy only affects non-arithmetical (i.e. mixed) sentences; thus, Hume's Principle might suffice to explain our knowledge of arithmetical truths, even if it leaves some of our further knowledge involving numbers unexplained. If what I am pressing is right, then an concept of number less determinate than ours is simply not our concept of number; and so truths containing that concept will not be the truths of arithmetic that we know. Thus, the worry is that if the Neo-Logicist lowers his sights even by an inch, he will in fact lose sight of his target entirely. Developing this point, however, is a task for another paper.

[16]Note that this is not the Fregean sense of "concept" as an unsaturated entity.

Works Cited

Bellomo, Anna (2021). "Domain Extension and Ideal Elements in Mathematics?" In: *Philosophia Mathematica* 29.3, 366–391.

Boghossian, Paul A. (2003). "Blind Reasoning". In: *Supplement to the Proceedings of the Aristotelian Society* 77.1, 225–248.

Boghossian, Paul Artin (1996). "Analyticity Reconsidered". In: *Noûs* 30.3, 360–391.

Boolos, George (1997). "Is Hume's Principle Analytic?" In: *Language, thought, and logic: essays in honour of Michael Dummett*. Ed. by Richard G. Heck. Oxford University Press.

Brandom, Robert (1994). *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, Mass.: Harvard University Press.

— (2000). *Articulating Reasons: An Introduction to Inferentialism*. Cambridge, Mass.: Harvard University Press.

Burge, Tyler (2003). "Logic and Analyticity". In: *Grazer Philosophische Studien* 66.1, 199–249.

— (2005). "Introduction". In: *Truth, Thought, Reason: Essays on Frege*. New York: Oxford University Press.

— (2020). "Entitlement: The Basis for Empirical Epistemic Warrant". In: ed. by Peter J. Graham and Nikolaj J. L. L. Pedersen. Oxford, UK: Oxford University Press.

Button, Tim and Sean Walsh (2016). "Structure and Categoricity: Determinacy of Reference and Truth Value in the Philosophy of Mathematics". In: *Philosophia Mathematica* 24.3, 283–307.

— (2018). *Philosophy and Model Theory*. Ed. by Sean Walsh and Wilfrid Hodges. Oxford, UK: Oxford University Press.

Cook, Roy T. and Philip A. Ebert (2005). "Abstraction and Identity". In: *Dialectica* 59.2, 121–139.

Davidson, Donald (1967). "Truth and Meaning". In: *Synthese* 17.1, 304–323.

Dedekind, Richard (1963). "The nature and meaning of numbers". In: *Essays on the theory of numbers*. Trans. by Wooster Woodruff Beman. New York: Dover.

Dummett, Michael (1991). *Frege: Philosophy of Mathematics*. Harvard University Press. Cambridge, MA: Cambridge University Press.

Evans, Gareth (1982). *The Varieties of Reference*. Ed. by John Henry McDowell. Oxford: Oxford University Press.

Fine, Kit (2002). *The Limits of Abstraction*. Ed. by Matthias Schirn. New York: Oxford University Press.

Frege, F. L. Gottlob (1892). "Über Sinn und Bedeutung". In: *Translations From the Philosophical Writings of Gottlob Frege*. Oxford, UK: Blackwell.

— (1950). *The Foundations of Arithmetic: A Logico-Mathematical Enquiry Into the Concept of Number*. Ed. by J. L. Austin. New York, NY, USA: Northwestern University Press.

— (1956). "The Thought: A Logical Inquiry". In: *Mind* 65.259, 289–311.

— (2013). *Basic Laws of Arithmetic*. Ed. by Philip A. Ebert and Marcus Rossberg. Oxford University Press UK.

Gentzen, Gerhard (1969). "Investigations into Logical Deduction". In: *The Collected Papers of Gerhard Gentzen*. Ed. by M. E. Szabo. Amsterdam: North Holland Publishing Company.

Greenberg, Mark and Gilbert Harman (2005). "Conceptual Role Semantics". In: *The Oxford Handbook of Philosophy of Language*. Ed. by Ernie Lepore and Barry C. Smith. Oxford University Press.

Hale, Bob (2001a). "*Grundlagen* §64". In: *The Reason's Proper Study: Essays Towards a Neo-Fregean Philosophy of Mathematics*. Ed. by Bob Hale and Crispin Wright. Clarendon Press.

— (2001b). "Reals by Abstraction". In: *The Reason's Proper Study: Essays Towards a Neo-Fregean Philosophy of Mathematics*. Ed. by Bob Hale and Crispin Wright. Clarendon Press.

Hale, Bob and Crispin Wright (2001a). "Implicit Definition and the A Priori". In: *The Reason's Proper Study: Essays Towards a Neo-Fregean Philosophy of Mathematics*. Ed. by Bob Hale and Crispin Wright. Clarendon Press.

— (2001b). *The Reason's Proper Study: Essays Towards a Neo-Fregean Philosophy of Mathematics*. Oxford: Clarendon Press.

— (2001c). "To Bury Caesar …" In: *The Reason's Proper Study: Essays Towards a Neo-Fregean Philosophy of Mathematics*. Ed. by Bob Hale and Crispin Wright. Clarendon Press.

— (2009). "Focus Restored: Comments on John Macfarlane". In: *Synthese* 170.3, 457–482.

Heck, Richard Kimberly (2011a). "Cardinality, Counting, and Equinumerosity". In: *Frege's Theorem*. Oxford, UK.: Oxford University Press. Originally published under the name Richard G. Heck.

— (2011b). "Finitude and Hume's Principle". In: *Frege's Theorem*. Oxford, UK.: Oxford University Press. Originally published under the name Richard G. Heck.

— (2011c). *Frege's Theorem*. Oxford, UK.: Oxford University Press. Originally published under the name Richard G. Heck.

— (2011d). "Frege's Theorem: An Introduction". In: *Frege's Theorem*. Oxford, UK.: Oxford University Press. Originally published under the name Richard G. Heck.

— (2011e). "Julius Caesar and Basic Law V". In: *Frege's Theorem*. Oxford, UK.: Oxford University Press. Originally published under the name Richard G. Heck.

— (2011f). "The Development of Arithmetic in Frege's *Grundgesetze der Arithmetik*". In: *Frege's Theorem*. Oxford, UK.: Oxford University Press. Originally published under the name Richard G. Heck.

Linnebo, Oystein (2018a). "Dummett on Indefinite Extensibility". In: *Philosophical Issues* 28.1, 196–220.

— (2018b). *Thin Objects: An Abstractionist Account*. Oxford: Oxford University Press.

MacBride, Fraser (2006). "The Julius Caesar Objection : More Problematic Than Ever". In: *Identity and modality*. Ed. by Fraser MacBride. Oxford University Press, 174–202.

MacFarlane, John (2009). "Double Vision: Two Questions About the Neo-Fregean Program". In: *Synthese* 170.3, 443–456.

McGee, Vann (1997). "How We Learn Mathematical Language". In: *Philosophical Review* 106.1, 35–68.

Murzi, Julien and Brett Topey (2021). "Categoricity by Convention". In: *Philosophical Studies* 178.10, 3391–3420.

Parsons, Charles (1982). "Objects and Logic". In: *The Monist* 65.4, 491–516.

— (2008). *Mathematical Thought and its Objects*. New York: Cambridge University Press.

Peano, Giuseppe (1967). "The principles of arithmetic, presented by a new method". In: *From Frege to Gödel*. Ed. by Jean van Heijenoort. Cambridge, MA: Harvard University Press.

Pedersen, Nikolaj J. L. L. (2009). "Solving the Caesar Problem Without Categorical Sortals". In: *Erkenntnis* 71.2, 141–155.

Perry, John, ed. (1975). *Personal Identity*. Berkeley: University of California Press.

Putnam, Hilary (1977). "Realism and Reason". In: *Proceedings and Addresses of the American Philosophical Association* 50.6, 483–498.

— (1980). "Models and Reality". In: *Journal of Symbolic Logic* 45.3, 464–482.

Quine, Willard V. O. (1951). "Two Dogmas of Empiricism". In: *Philosophical Review* 60.1, 20–43.

— (1954). "Carnap and Logical Truth". In: *Synthese* 12.4, 350–74.

Quine, Willard Van Orman (1948). "On What There Is". In: *Review of Metaphysics* 2.5, 21–38.

Ramsey, Frank P. (1927). "Facts and Propositions". In: *Aristotelian Society Supplementary Volume* 7.1, 153–170.

Schwitzgebel, Eric (2013). "A Dispositional Approach to the Attitudes". In: *New Essays on Belief: Constitution, Content and Structure*. Ed. by Nikolaj Nottelmann. Palgrave, 75–99.

Shapiro, Stewart (1991). *Foundations Without Foundationalism: A Case for Second-Order Logic*. New York: Oxford University Press.

— (2000). "Frege Meets Dedekind: A Neologicist Treatment of Real Analysis". In: *Notre Dame Journal of Formal Logic* 41.4, 335–364.

Studd, J. P. (2023). “The Caesar Problem – a Piecemeal Solution”. In: *Philosophia Mathematica* 31.2, 236–267.

Tarski, Alfred (1956). “The Concept of Truth in Formalized Languages”. In: *Logic, semantics, metamathematics*. Ed. by Alfred Tarski. Clarendon Press, 152–278.

Walsh, Sean (2014). “Logicism, Interpretability, and Knowledge of Arithmetic”. In: *Review of Symbolic Logic* 7.1, 84–119.

Walsh, Sean and Sean Ebels-Duggan (2015). “Relative Categoricity and Abstraction Principles”. In: *Review of Symbolic Logic* 8.3, 572–606.

Warren, Jared (2020a). “Infinite Reasoning”. In: *Philosophy and Phenomenological Research* 103, 385–407.

— (2020b). *Shadows of Syntax: Revitalizing Logical and Mathematical Conventionalism*. New York, USA: Oxford University Press.

Westerstahl, Dag, Alexandru Baltag, and Johan van Benthem (2021). “Compositionality in Context”. In: *Outstanding Contributions to Logic: Samson Abramsky*. Ed. by A. Palmigiano and M. Zadrzadeh. Springer.

Wright, Crispin (1983). *Frege’s Conception of Numbers as Objects*. Aberdeen: Aberdeen University Press.

— (2000). “Neo-Fregean Foundations for Real Analysis: Some Reflections on Frege’s Constraint”. In: *Notre Dame Journal of Formal Logic* 41.4, 317–334.

— (2001a). “Is Hume’s Principle Analytic?” In: *The Reason’s Proper Study: Essays Towards a Neo-Fregean Philosophy of Mathematics*. Ed. by Bob Hale and Crispin Wright. Clarendon Press.

— (2001b). “On the Harmless Impredicativity of $N^{=}$”. In: *The Reason’s Proper Study: Essays Towards a Neo-Fregean Philosophy of Mathematics*. Ed. by Bob Hale and Crispin Wright. Clarendon Press.

— (2004). “Warrant for nothing (and foundations for free?)” In: *Aristotelian Society Supplementary Volume* 78.1, 167–212.