2018-07-10

New Directions for Neo-logicism

Thomas-Bolduc, Aaron Robert

UNIVERSITY OF CALGARY

New Directions for Neo-logicism

by

Aaron Thomas-Bolduc

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN PHILOSOPHY

CALGARY, ALBERTA

JULY, 2018

Abstract

In this dissertation I focus on a program in the philosophy of mathematics known as neo-logicism that is a direct descendant of Frege's logicist project. That program seeks to reduce mathematical theories to logic and definitions in order to put those theories on stable epistemic and logical footing. The definitions that are of greatest importance are *abstraction principles*, biconditionals associating identity statements for abstract objects on one side, with equivalence classes on the other. Abstraction principles are important because they provide connections between logic on the one hand, and mathematics and its ontology on the other.

Throughout this work, I advocate that the epistemic goals of neo-logicism be taken into account when we're looking to solve problems that are of central importance to its success. Additionally, each chapter either discusses or advocates for a methodological shift, or sets up and implements a novel methodological position I believe to be broadly beneficial to the neo-logicist project.

Chapter 2 traces thinking about the status of higher-order logic through the mid-twentieth century, setting the stage for issues dealt with in later chapters. Chapter 3 asks neo-logicists to look beyond set theory and consider other foundational theories, or something entirely new when looking for reductions of foundational mathematical theories. Chapter 4 is an extended argument involving non-standard analysis showing that Hume's Principle ought not be considered analytic in *Frege's* sense of the term.

Chapters 5 and 6 move away form the (somewhat) historical work in the first three chapters, and set up new strategies for solving central neo-logicist

problems by integrating formal and epistemic considerations. Chapter 5 introduces the notion of a *canonical equivalence relation* via a discussion of content carving. That notion is a particular way of understanding the relationship between equivalence relations and abstracts. Finally, chapter 6 makes use of canonical equivalence relations to introduce a new direction in the search for solutions to the Bad Company objection.

As whole, the project can be seen as providing, as the title suggests, new directions that ought to be considered by those wishing to vindicate neologicism.

Preface

Chapter 4, "Is Hume's Principle Analytic?", is a slightly revised version of paper I co-authored with Eamon Darnell, that is currently under review. We contributed equally to the paper, though I did majority of the writing. Copyright permission is included as appendix C.

The rest of this dissertation is the original, unpublished, independent work of the author, Aaron Thomas-Bolduc. That work has benefited from discussion with a great number of people, to whom I'm very grateful. The most important of those are mentioned in the acknowledgments.

Acknowledgements

There are a number of people without whose help and support the completion of this dissertation would not have been possible, and great many more whose support has made the process easier and the content better.

I would first like to thank my supervisor, Richard Zach. I could not have asked for a better, more supportive or more helpful supervisor. I would also like to thank the other members of my supervisory committee, Nicole Wyatt and Dave Liebesman. Nicole's continued advice and support has been invaluable, and both Dave and Nicole's comments and suggestions have greatly improved my work. I am also grateful to the other members of my examination committee, Roy Cook and Jeremy Fantl, for their interesting and insightful questions, and subsequent discussion.

A massive 'thank you' also to my friend and collaborator, Eamon Darnell without whom chapter 4 would never have existed, and who has helped improve my work a great deal.

My interests in philosophy of mathematics and Frege were first sparked when I was and undergraduate at the University of Bristol, and I would be remiss not to thank Leon Horsten and Irina Starikova who introduced me to those topics; Richard Pettigrew who supervised my MA thesis; and especially Øystein Linnebo who has continually and enthusiastically discussed my work, invited me to workshops and conferences, and offered his general support over the past decade.

Thank you to Aldo Antonelli who was a great friend and inspiration. He will be greatly missed.

A special thanks to Denise Retzlaff, our Graduate Program Administrator,

who has gone above and beyond to help and support me and the other graduate students. Thanks as well to the other administrators, Merlette Schnell, Sage Murphy, and Tram Nguyen.

The friendship and emotional support, as well as philosophical discussion of Justin Caouette and Alison McConwell, has made this entire process much better than it could have been. It would have been a much different experience without them around. Thanks as well to Ted Shear who is a great friend and interlocutor, and a shout out to the great graduate student community in Calgary, as well.

Of course, support from outside the philosophical community is also very important. I am extremely grateful for the emotional and financial support of my family, especially my parents, Barbara Thomas and Norman Bolduc, my grandmothers, Georgette Sadler and Norma Thomas, and my late grandfather, Bob Thomas, without whom I would never have made it this far.

The last months of this project would have been immensely more difficult without the love, support and understanding of Acacia Morales.

Thanks too, to the Twelfth Avenue Crew who have helped keep me grounded, and have always been supportive.

Finally, the content of this dissertation has benefited from discussions with a great number of people I've not yet mentioned, among them Sean Walsh, Sal Florio, Adam Sennett, Augustín Rayo, Robert May, Stewart Shapiro, and audiences in Calgary, Oslo and Davis.

Table of Contents

Chapter 1

Introduction

In this dissertation I focus on a program in the philosophy of mathematics known as neo-logicism that is a direct descendant of Gottlob Frege's logicist project from around the turn of the last century Frege (1884). The program seeks to reduce mathematical theories to logic and definitions in order to put those theories on stable epistemic and logical footing. When we say that the goal is to *reduce* mathematical theories, or to provide *neo-logicist reductions*, we more or less mean that we are looking for proofs of the canonical axioms of mathematical theories using only logic and definitions. The definitions that are of greatest importance are *abstraction principles*, biconditionals associating identity statements for abstract objects on one side, with equivalence classes on the other. Abstraction principles are important because they provide the epistemological and metaphysical connections between logic on the one hand, and mathematics and its ontology on the other. Abstraction principles are also responsible for most of the mathematical strength of neo-logicist systems.

If neo-logicism is to be successful, it must be ensured that we are warranted in our use of abstraction principles; ensured that we are only availing ourselves of principles that are themselves on firm metaphysical, logical and especially epistemic ground. One obstacle to such assurances, known the *Bad Company* problem, is that there are abstraction principles that are inconsistent, and others that are individually consistent but jointly inconsistent. To ensure the epistemic credentials of the abstraction principles we need for neo-logicist

reductions, we need to find a principled way of excluding bad companions—abstraction principles that lead to inconsistency—otherwise those bad companions would have to be attributed the same foundational status as 'good' abstraction principles.

The recent neo-logicist literature—especially that focused on Bad Company—has a tendency to focus on the logic of abstraction principles, and perhaps the connection of neo-logicist metaphysics or semantics to that formal aspect, while paying scant attention to the epistemological aspects that are central to the program begun by Crispin Wright and Bob Hale in the 1980s (see Wright, 1983; Hale, 1987; Hale and Wright, 2001; MacBride, 2003).

In his forthcoming book, *Thin Objects: An Abstractionist Account*, Øystein Linnebo has this to say about the issue.

> Most worrisome of all, in my opinion, is the extent to which recent work on the bad company problem has become a largely technical undertaking, which has lost touch with the underlying philosophical question of how abstraction might work. Ideally, we would like our philosophical account of abstraction to motivate, or at least inform, our answer to the bad company problem.
>
> In short, it is time to try a new tack.[1] (Linnebo, 2018, p. 55)

This dissertation is, in part, an attempt to rectify that situation. The focus on moving away from purely formal work to explicitly reflect the epistemological components of neo-logicism will perhaps be most obvious in the last couple of chapters, though the importance of epistemology will be an obvious theme throughout.

Each chapter also either discusses or advocates a methodological shift, or sets up and implements a novel methodological position I believe to be broadly beneficial to the neo-logicist project. Additionally, the dissertation as a whole is centered around the Bad Company objection.[2] Because that problem intersects significantly with the epistemic *and* formal aspects of neo-

[1] The tack Linnebo takes is very different from the one I take. See §1.4 for a brief overview.
[2] See §1.3 and chapter 6.

logicism, it provides a useful focus for the discussion. Furthermore, chapter 6 is an extended investigation of one important aspect of Bad Company.

Before going into more detail about the contents of the dissertation chapters in §1.5 of this introduction, some set-up is needed. The next section is a very brief history of logicism and neo-logicism, which is followed by a section laying out the details of the variety of neo-logicism I'm concerned with. That also includes some of the formal details of the background higher-order logic that will be assumed throughout, as well as of abstraction principles (APs).

I then introduce the central objections to neo-logicism that will be of concern throughout. That's followed by brief descriptions of the various directions abstractionist philosophers of mathematics have taken to try to find a consistent, philosophically defensible and mathematically strong system based on APs.[3] In other words, I will sketch the various kinds of attempts to solve the central problems I'll have just described.

1.1 A Brief History

Frege first laid out his logicist program of reducing arithmetic to logic and definition in the Grundlagen der Arithmetik (Frege, 1884). Therein he sets up the project and presents his program informally.

Frege's attempted reduction of arithmetic to logic involves three major components: the concept-script, six "basic laws", and definitions of the basic concepts of arithmetic. Together, these three components allowed Frege to derive the Dedekind-Peano axioms of arithmetic from what he took to be purely logical grounds. The concept-script (in German, *Begriffsschrift*), first presented in his eponymous book of (1879) and expanded on and modified in §§2–52 of the Basic Laws of Arithmetic (*Grundgesetze der Arithmetik*, hereafter *Grundgesetze* (Frege, 1893, 1903, 2013)), is the logical formalism that was designed to represent mathematical reasoning entirely without gaps or ambiguities, and

[3] I use the term 'abstractionist' and cognates to refer to any program in the philosophy of mathematics with APs at it's core, reserving 'neo-logicist' for the particular sorts of systems described in §1.2. Similarly, I use 'neo-Fregean' for any program or doctrine that can be seen as a direct descendant of Frege's thought. Neo-logicists are then a particular sort of neo-Fregean abstractionists.

is usually considered to be the first example of a symbolic quantified deductive system. The basic laws are logical axioms not in need of justification, but are related to certain metaphysical theses. The definitions are explications of the arithmetical vocabulary in the language of the concept-script.

As is well-known, Frege's reduction as presented in the *Grundgesetze* leads directly to Russell's paradox. In modern terms, the inconsistency arises from the combination of Basic Law V (BLV), which says that the extensions of two concepts are identical just in case exactly the same objects fall under them; and the unrestricted second-order comprehension principle that is part of the background logic. Concepts, in Frege's sense, are functions the domain of which are objects, and the range of which are the truth values (also objects).[4] Extensions are abstract objects associated with the collections of objects which fall under specific concepts. Basic Law V then says that there is a unique object associated with every extensionally equivalent class of concepts, but if every formula determines a concept, as Frege thought, then there must be more concepts than objects, by familiar arguments due to Cantor and Russell.[5]

The logicist project was then picked up by Bertrand Russell and A.N. Whitehead, who based their logicist system in the *Principia Mathematica* on the ramified theory of types (Whitehead and Russell, 1927). This allowed Russell and Whitehead to skirt the paradoxes, but required the introduction of two axioms—reducibility and infinity—that couldn't be properly justified from the logicist perspective.[6] Nevertheless, logicism, particularly that of Russell and Whitehead, was taken as an important and serious contender as a foundation for mathematics for roughly two decades following the publication of *Principia* (see Carnap, 1931; Hempel, 1945; Church, 1962).

Chapter 2 includes more historical detail, particularly about the period between 1945 and 1983.

[4]More generally, we use the term 'concept' to refer to something very much like (usually monadic) properties.

[5]See §1.2 for formal definitions. The paradox was pointed out by Russell in a letter to Frege dated 1902, reprinted in van Heijenoort (1967, pp. 124–126).

[6]Frank Ramsey later pointed out that a simple theory of types was all that was needed, obviating the need for the axiom of reducibility (Ramsey, 1926).

1.2 Neo-Logicism

More recently it has been noted that BLV is only needed to derive the identity criterion for cardinal numbers, now known as Hume's Principle or HP, which can then be used, without the need for BLV, to derive the axioms of second-order Peano arithmetic (PA^2, sometimes also known as *analysis*). This result is known as Frege's Theorem, and the system of second-order arithmetic with HP as its sole additional axiom, and with definitions of basic arithmetical concepts, is known as Frege arithmetic (FA).[7]

Frege's Theorem, together with Boolos' (1987a) result that FA is interpretable in PA^2, establishing that that Frege Arithmetic and Peano Arithmetic are mutually interpretable, allowed for the development of what is now know as (Scottish) neo-nogicism, which began with Crispin Wright (1983), and Bob Hale (1987) (but see especially the papers collected in Hale and Wright, 2001).

The common ground between logicists and neo-logicists is the general structure and aims of their foundational projects. In both cases the goal is to reduce centrally important mathematical theories to more basic, uncontroversial logical systems. For Frege, as well as many neo-logicists, the ultimate goal is to show that mathematics (or at least arithmetic) is analytic and/or *a priori*. A bit more generally, the goal is to provide a secure epistemic foundation for mathematics by reducing mathematical theories to, or interpreting mathematical theories in, epistemically secure formal systems consisting of logic and definitions. At this point, the question still remains as to how we are to bridge the ontological gap between the sparse ontology of logic and the rich ontology of mathematics. The answer for Frege and neo-logicists (but not Russell and Whitehead) is to use *abstraction principles*. The insight of neo-logicism is that the general logicist project can be carried out by taking APs to be epistemically privileged, without taking on all of Frege's commitments, e.g. to thinking of numbers as logical objects, or BLV as a law of logic.

The neo-logicist project began in earnest with Wright's (1983) book, *Frege's Conception of Numbers as Objects*. In that book, Wright rejects BLV and instead

[7] The possibility of the derivation of the axioms of PA^2 the was arguably first noted by Charles Parsons (1965). See §2.5 for more discussion of the discovery of Frege's theorem.

takes on HP as the foundational principle for arithmetic. That move is the central feature of what has come to be known as Scottish neo-logicism.[8] The other overarching difference between the neo-logicism of today, and Frege's turn of the (20th) century logicism, has to do with how APs are viewed. For Frege, BLV, from which he derived HP and hence the axioms of arithmetic, was a logical law governing the relationship between concepts and the collections of objects that fall under them. Most neo-logicists, on the other hand, have rejected the inconsistent BLV entirely, along with the view that (some) APs are logical laws.[9] For the neo-logicist project to have any chance of success then, certain consistent APs must be established as being epistemically privileged in some other way.

Second-order Logic

Neo-logicism relies for the most part on *second-order logic*. Syntactically, second-order logic is just like first-order predicate logic with identity, except with quantifiers ($\forall X$ and $\exists X$) that bind variables ranging over n-adic predicates. By convention, second order variables are represented with the upper-case Roman letters F, G, H, R, X, Y and Z. F and G generally represent monadic predicates, H and R dyadic predicates (relations) and $X--Z$ predicates of any arity. The second-order quantifiers are governed by introduction and elimination rules analogous to the first-order case.[10]

Second-order logic also contains an *axiom schema of comprehension* i.e. all axioms of the form

(Comprehension) $$\exists X \forall x (Xx \leftrightarrow \phi(x))$$

where ϕ is a formula containing no free occurrences of X. This means that there's a predicate that corresponds to any formula. If no other restriction's are placed on the formula ϕ we say that we're working with *full* or *unre-*

[8] So called because Hale and Wright both held positions in Scotland at the time.

[9] I briefly discuss a few exceptions in §1.4.

[10] Nowhere in this dissertation do I present any formal proofs in second- (or higher-) order logic, so I won't lay out a particular proof system.

stricted comprehension. In most of what follows I will be assuming that the background logic contains full second-order comprehension.

At the other end of the spectrum, we can restrict ϕ to formulae with no second-order variables. This is *predicative* or *arithmetic* comprehension. We can also make use of amounts of comprehension in between full and predicative. These are differentiated by the complexity of the second-order formulae allowed. We say a second-order formula is Π^1_n if it begins with a block of second-order universal quantifiers and the (blocks of) second-order quantifiers alternate n-times. A formula is Σ^1_n if it begins with a string of second-order existential quantifiers. Finally, a formula is Δ^1_n if it is provably equivalent to both a Π^1_n formula and a Σ^1_n formula. Note that BLV is consistent with no more than Δ^1_1 comprehension (see Horsten and Linnebo, 2016, for an interesting model).

In most of what follows I will be assuming what we call the *standard* (or full) semantics for second-order logic. What that means is that the second-order variables range over the full powerset of the first-order domain.[11] There also also *generalized* or *Henkin* semantics for second-order logic which restrict the available extensions of the predicate variables, allowing second-order logic to be treated as a two-sorted, first-order system.[12]

Second-order logic with standard semantics behaves differently to first-order logic in important ways. Second-order logic is (very) incomplete, the Löwenheim-Skolem theorems fail, and it isn't compact. However, it does allow for the formulation of categorical theories, which means that PA^2 has no non-standard models (unlike first order PA). Note that both first- and second-order logic are decidable in the monadic case, but not polyadic cases. These differences have lead to skepticism about whether second-order logic is logic in the same sense as first-order logic.[13]

Before moving on, we should note that there's nothing stopping us from

[11] The powerset of a set is the set of all subsets of that set, and has cardinality 2^κ, if κ is the cardinality of the first-order domain.

[12] Hale (2015) develops a semantics for second-order logic based on the set of *definable* subsets of the first-order domain that is also of interest to neo-logicists.

[13] See chapter 2 for discussion, and Shapiro (2002) for an extended treatment of second-order logic.

adding third- or higher-order quantifiers in an analogous manner. Indeed Cook (2012); Logan (2015); Cook and Linnebo (2018) and others make use of third-order logic (see also chapters 3 and 6). By convention, third-order variables are typeset in boldface (i.e. $\forall \mathbf{X} \exists X \forall x \mathbf{X} X x$).

Abstraction Principles

At the heart of neo-logicism are *abstraction principles*. APs are sentences of the following (equivalent) forms:

$$\S F = \S G \leftrightarrow F \sim G;$$

$$\partial_E F = \partial_E G \leftrightarrow E(F, G);$$

or

$$Ax.Fx = Ax.Gx \leftrightarrow E_x(F, G)$$

The '§', '∂_E', and '$Ax.$' on the left hand sides represent abstraction operators interpreted as either injective functions from the first- or higher-order domain into the first-order domain, or as variable-binding term-forming operators. The result of appending an abstraction operator to an appropriate (first- or higher-order) formula is an abstraction *term* and denotes an abstract object, or just an 'abstract'. The '$\sim$', '$E(\cdot)$', and 'E_x' (also sometimes $\Phi(\cdot)$) represent equivalence relations on the entities on the RHS of the AP.[14]

Most of the APs we care about are expressible in second-order logic, but not all. Frege's direction principle

$$d(l_1) = d(l_2) \leftrightarrow l_1 // l_2$$

which says that the directions of two lines can be identified if and only if those two lines are parallel, is first-order. Shay Allen Logan's APs for categories

[14] Burgess (2005, §2.6) shows proves that abstraction operators taken as variable-binding term forming operators are inter-definable with abstraction operators taken as functions or type-lowering operators in the presence of full second-order logic.

require third-order logic (Logan, 2015), and the Apogon principles discussed at length in chapter 6 are most naturally formulated in third-order logic.

The two most discussed APs (here and elsewhere) are Frege's BLV and Hume's principle (HP).

(BLV) $$\varepsilon F = \varepsilon G \leftrightarrow F \equiv G$$

(HP) $$\#F = \#G \leftrightarrow F \approx G$$

BLV says that two concepts have the same extension just in case exactly the same objects fall under both. HP says that the concepts F and G have the same number iff they are equinumerous (*gleichzalig*), which is to say there there exists a bijection between the objects falling under F and those falling under G.

The importance of HP to traditional strains of neo-logicism can hardly be overstated. It was suggested by Charles Parsons (1965) that the Dedekind-Peano axioms of (second-order) arithmetic could be deduced from HP in the presence of full second-order logic plus Frege's definitions of natural number, predecessor, and zero, and without recourse to the inconsistent BLV. Parsons' proof sketch was incomplete, as was Wright's attempt in (1983). Nevertheless, such a proof *is* possible, and is central to the hope that neo-logicist reductions of stronger mathematical theories is possible (see Heck, 2011b).

After the publication of Wright's book, it was quickly established that PA^2 interprets FA (Burgess, 1984). This gives us very good reason to think that FA, and hence HP, is consistent, because the consistency of analysis is hardly in doubt. Once Frege's Theorem and the mutual interpretability result were established, there was space to look for other consistent APs that could be used to capture mathematics.

Abstraction principles are not just important as formal tools, however. Scottish neo-logicists require APs to carry significant epistemic weight, too. Since the epistemic grounding of mathematics (or at least arithmetic) is central to neo-logicism, APs must at least have a foundational epistemic status akin

to that attributed to logic.[15] Frege took BLV as, well, a basic law of logic. We now know that that can't be the case, but there are other ways APs could play the epistemic role envisioned for them. I'll have more to say about that in §1.3 (and even more in chapters 4 and 5), but it is worth noting at this point that it is often argued that our understanding of the abstracts identified on the LHS of an AP is grounded in our previous understanding of the concepts and equivalence relation on the RHS (see especially chapter 5).

Platonism

The other key component of the sort of neo-logicism I'll be concerned with here is platonism about mathematics. (That's 'platonism' with a small 'p', not to be confused with Plato's actual views about mathematics, which we might call mathematical Platonism.) Frege certainly believed that numbers are objects—*logical* objects to be precise. The view that mathematics talks about real, mind-independent abstract objects is part of Frege's legacy that has been adopted by neo-logicists (though not just by neo-logicists).

Mathematical platonism is usually glossed as the view that mathematical objects exist non-physically 'outside' space and time, and that their existence doesn't rely on agents, minds or applications—mathematical objects are abstract objects. Many neo-Fregeans also hold that abstract objects are *thin*, a view usually coupled with (or following from) a commitment to meta-ontological minimalism. Thin objects require very little from the world for their existence; meta-ontological minimalism is the view that little or nothing more than coherence is required for existence (Hale, 2011; Rayo, 2013; Linnebo, 2018).

Platonism is assumed throughout this dissertation, often without comment. Certain suggestions made in chapters 3 and 5 are best supported by an understanding of abstracts as being thin, though none of my central claims rely on that assumption.

[15] I will remain largely silent on exactly what that status is, relying on the traditional view that classical first-order predicate logic is unquestionable. I don't think this is obviously the case, though it is a plausible assumption if we're dealing only with classical mathematics.

1.3 Problems

In this section I briefly lay out four objections usually taken to be of central importance to the success or failure of neo-logicism. I spend proportionally more space on Bad Company, as that objection plays a central role throughout this dissertation.

Analyticity

Frege thought that BLV was a law of logic, and thus that by extension, extensions (of which numbers were to be a subclass) were logical objects. Since he derived HP from BLV and definitions, it was, by his own lights, *analytic*.[16] Taking HP as primitive, such a route isn't open to neo-logicists, but an account is needed of the nature of the epistemic privilege possessed by (acceptable) APs.

One of the first proposals, now generally rejected, was that HP is analytic in a sense more standard than Frege's (see, e.g. Wright, 1999; Boolos, 1997).

A different, though compatible, approach to this question is to argue that APs are implicit definitions of the concepts that the objects referred to by abstraction terms fall under, and then arguing that implicit definitions carry enough epistemic weight to ground knowledge about those objects (see Hale and Wright, 2000). For example, HP would be taken as an implicit definition of natural or cardinal number, and the fact that that is so grants us some epistemic access to the nature of numbers. There are various problems with this approach, one of which—the difference between axiom systems and APs as candidates for implicit definitions—has been around almost as long as logicism itself in the form of the Frege-Hilbert controversy (Blanchette, 2014).

Another route is to argue that the the nature of abstraction—what's going on when we use APs—accounts for their epistemic significance. A common route is via content carving or reconceptualization (Hale, 1997; Fine, 2002; Rayo, 2013). Very roughly, the thought is that both sides of an AP express the same content, and so since (or when) we understand (know about, have

[16] Chapter 4 tackles the question of whether HP is analytic in Frege's sense.

epistemic access to. . .) what's going on on the RHS, that understanding can be transmitted to the abstracts identified on the LHS.

I go into some detail about content carving in chapter 5 in a slightly different context, and argue that HP isn't analytic in Frege's sense in chapter 4, but otherwise assume that *some* account of the special epistemic status of APs will be available.

Caesar

In §56 of the *Grundlagen* (Frege, 1980) Frege famously laments that

> . . . we can – to give a crude example – never decide by means of our definition whether Julius Caesar belongs to a number concept, whether this same well-known conqueror of Gaul is a number or not.

Frege's point is that a principle like HP can't tell us whether some object not presented as an abstraction term (i.e. the terms on the LHSs of APs), say Julius Caesar, is identical to an abstract, say the number 7.

More problematic from my perspective is a related issue sometimes known as the C-R problem.[17] The problem is that APs don't tell us whether we can identify abstracts picked out by different APs.

As far as I'm aware, there is not yet any real consensus on how to deal with either objection, though we're not wanting for proposals. These problems play a role in the latter part of chapter 4, and come up, however briefly in various places. Chapters 3 and 5 also contain sketches of broader approaches that may inform how we go about solving them, but I won't go into much detail otherwise.

Bad Company

As I mentioned above, the Bad Company objection plays a central role throughout much of this work, and is the primary focus of chapter 6. Very

[17] Named by Cook and Ebert (2005), as the problem arises when we ask whether the real numbers (R) are embedded in or a subclass of the complex numbers (C).

roughly, Bad Company is the problem of how we can establish the epistemic credentials of paradigmatically 'good' APs like HP, without attributing that same status to 'bad' APs like BLV.

So far as I am aware, the Bad Company objection was first raised by Allen Hazen in his review of Wright (1983) (Hazen, 1985), though the thought seems also to have occurred to John Burgess in *his* review of the same book (Burgess, 1984). Hazen's point is that it is at least plausible to think that, if HP is acceptable, so too should the ordinal abstraction principle.[18] He then sketches a straightforward proof that ordinal abstraction leads to the Burali-Forti paradox which asserts that the class of ordinals is bigger than itself. This is an example of the form of Bad Company that asks why we should take HP to be privileged, if there are inconsistent APs of the same form, the more common formulation of which is raised by Micheal Dummett in his book *Frege: Philosophy of Mathematics*:

> [T]he mere fact that, on [Wright's] view, it is unnecessary to define the cardinality operator in terms of classes or value-ranges does not entitle Wright to ignore the problem of the abstraction operator [in BLV]. For Frege's method of introducing the abstraction operator – that is, of introducing value-ranges – was, notoriously, *not* in order. (Dummett, 1991, p. 188, emphasis his)

Here Dummett is pointing out that it is not *prima facie* legitimate to lay down HP as definitional of the concept 'cardinal number' because the analogous case for extensions—BLV—is known to lead to Russell's paradox. This is the core of the Bad Company objection. The problem was then taken up in significant detail by Richard Heck Jr. (2011d). Heck's paper is important for two reasons. First, he argues convincingly, *pace* Wright (1983), that consistency and satisfiability are not sufficient for the acceptability of abstraction principles. Previously it was thought, because the best known examples of abstraction principles were HP and BLV, that any consistent AP could play

[18] Hazen introduces ordinal abstraction as: "$Oxy(Fx; Rxy) = Oxy(Gx; Sxy)$ iff either R and S well-order the Fs and Gs respectively and there is an order-isomorphism between them, or neither relation well-orders the specified field" (p. 253–4).

the same role for the relevant objects as HP does for numbers (whatever that may be). Heck showed that consistency was insufficient by providing his now much employed formula for producing satisfiable, but oft problematic APs (adapting Heck's notation):

$$\partial F = \partial G \leftrightarrow (\forall x(Fx \leftrightarrow Gx) \vee A)$$

where A is any satisfiable second-order sentence. Such principles are problematic: if we allow that they are acceptable because they are satisfiable (and thus consistent), we can essentially force any satisfiable second-order sentence to be true, on the grounds that the class of acceptable abstraction principles should include all and only those principles that can (potentially) be used to ground mathematics logically and epistemically.[19] So, for example, by using Heck's trick we can force the universe to contain strongly accessible-many objects without doing any set theory.[20]

The second major contribution of Heck's article is his suggestion (note 4) of what is now known as stability[21] as an additional necessary condition on acceptable abstraction, setting in motion the search for acceptability criteria for abstraction principles. This was the beginning of a cottage industry constructing model-theoretic acceptability criteria, and counter-examples to those criteria, about which more in §1.4.

The other side of Bad Company was, I believe, first made explicit by Boolos (1990), wherein he introduces the Parity Principle which says that the parity of F is the parity of G just in case the symmetric difference between the Fs and Gs is even. The Parity Principle (PP) is only satisfiable on finite domains, and is thus jointly inconsistent with HP. Boolos makes his point thus: "Hume's principle is inconsistent with the parity principle. Which is the logical truth?" (p. 215).

[19] The formula 'A' is forced to be true if we accept a Heck principle on pain of contradiction via Russell's paradox—the other disjunct is unsatisfiable and inconsistent.

[20] APs like these, where 'A' is a cardinality property, I call *Heck Principles*. They dealt with in great detail in chapter 6.

[21] An AP is stable iff there is a cardinal, κ, such that said AP is satisfiable at any cardinal $\lambda \geq \kappa$. See appendix B for definitions.

This is particularly bad because it means that we can't just pick out individual (consistent) APs for our neo-logicist reductions, rather we have to find a way to pick out a *class* of acceptable APs.[22] This will be important if we want to hold onto FA, but also capture more mathematics.

Good Company

More recently a problem related to Bad Company has been identified and named by Paolo Mancosu: Good Company (Mancosu, 2015, 2016). Where Bad Company asks us to separate the good APs from bad, Good Company asks us to separate the exemplary from the good. Mancosu focuses on the case of the neo-logicist reduction of arithmetic, pointing out that there are ostensibly acceptable APs that can get us PA^2 besides HP.

The problem, then, is that without either a way to identify certain choices as privileged, or a principled reason to identify the abstracts picked out by *all* of the APs fit for a given purpose, we will have no reason to accept the epistemic privilege of one but not the others. But if we can't do that, we also can't claim that the abstracts associated with a particular AP are, say, *the* numbers. In chapter 5, I sketch possible solutions taking both routes, but don't claim to have a worked out solution; the goal is only to demonstrate the usefulness of the notion of a canonical equivalence relation that I develop in that chapter. Nevertheless, Good Company raises its head in various places along the way.

1.4 Solutions

Predicativity

One route toward avoiding Bad Company originates in the recognition that the derivation of Russell's paradox from BLV makes use of the impredicative comprehension principle in the background logic. The thought is that the use of extensions in attempted neo-logicist reductions can be saved by blocking the availability of the requisite comprehension axiom which asserts the

[22] See Cook and Linnebo (2018) for the state of the art.

existence of the Russell concept. In other words, the complexity of formulae allowed to factor into the comprehension schema is restricted in such a way that the formula(e) that would pick out the Russell concept, do not in fact do so. The end result being that there is no instance of BLV involving the Russell concept (because it doesn't exist from the perspective of the restricted background logic).

The first result regarding restricted comprehension was due to Terence Parsons (1987) who gave a model theoretic argument showing that the first-order part of Frege's system with the schematic version of BLV is consistent. Since then, it has been shown that BLV is consistent in second-order systems with at most Δ^1_1-comprehension, which is to say that inconsistencies arise already with Π^1_1 and Σ^1_1 comprehension (Heck, 1996; Ferreira and Wehmeier, 2002; Cruz-Filipe and Ferreira, 2015, fn. 6). Additionally, much of the terrain in between has been investigated, and the various resulting systems compared in terms of interpretability and consistency strength with various systems of second-order arithmetic with restricted comprehension (cf. Walsh, 2012; Burgess, 2005).

What the above results establish is, first, that if the goal is to ground as much mathematics as possible, adopting BLV and restricting comprehension leaves us wanting in that full second-order arithmetic is already out of reach.[23] On the other hand, it is well known that BLV plus predicative comprehension is mutually interpretable with Robinson's **Q**, and thus interprets a fair amount of mathematics (see Burgess, 2005, Chapter 2 for an overview and references). For those like Dummett or Russell who are suspicious of impredicative comprehension regardless of whether BLV is present, this should provide at least some comfort.

Predicativism aside, it looks as if restricting comprehension isn't going to give us a viable route to developing an epistemically secure foundation for mathematics based an APs. This is despite the fact that BLV and other 'inconsistent' APs will no longer lead to inconsistencies. However, this line of research has lead to two new approaches that are more promising, at least

[23]The strongest consistent second-order system that includes BLV is strictly weaker than the weak theory of (first-order) arithmetic known as $I\Delta_0(superexp)$ (Burgess, 2005, p. 226).

from a mathematical perspective.

Sean Walsh has recently established two important results for a (consistent) system containing abstraction principles, predicative comprehension, the Σ^1_1-choice schema, and the global choice principle (Walsh, 2016a,b). The first result is that the joint consistency version of Bad Company goes away, in the sense that, so long as the equivalence relation on the RHS of the AP is provably an equivalence relation in the predicative background theory with global choice, it will be consistent with all other such APs.

The second important result is that Walsh's theory interprets a significant amount of set theory.[24] These are promising results, but it is yet to be seen how the choice axioms can be justified, and to what extent such a system can claim to ground mathematical knowledge. I'll say no more about this other than to mention that Walsh's results lend some support to the sort of system I sketch at the end of chapter 3.

Two other predicative approaches to abstractionism are worth mentioning. Francesca Boccuni and Simon Hewitt have been developing systems that eschew higher-order logic in favour of plural quantification (Boccuni, 2011, 2013; Hewitt, forthcoming). Plural quantification adds to first-order logic quantifiers, variables and names for *pluralities*.[25] Pluralities can be seen as the formal analogue of natural language expressions like "some things", and it's important pluralities not be seen as *reified* individuals like sets. Also important is that adding plural quantification to first-order logic results in a system mutually interpretable with *monadic* second-order logic. That means that HP can't be expressed without adding extra resources, because HP asserts the existence of a binary predicate. Boccuni gets around this by basing her system on a version of BLV, thereby creating what looks like a weak set theory. Hewitt adds a pair-abstraction principle, which allows polyadic second-order logic to be simulated. Though these are interesting developments, it's unclear in what sense such systems can be seen as neo-logicist.

Similar things can be said about the dynamic approach to abstraction developed by Linnebo (2018) and James Studd (2015). Very roughly, dynamic

[24] Specifically, ZF without the powerset axiom.
[25] See Boolos (1984); Linnebo (2017) for details.

abstraction is a method of building up mathematical domains based on abstraction principles in a manner that guarantees consistency—essentially, the paradoxes are 'left out', similar to the way that the liar sentence is left out when constructing Kripke models of the truth predicate. The key to dynamic abstraction is taking a potentialist view of the mathematical universe, and building up an unbounded sequence of models, with each new model extending the previous with the addition of abstracts provided by the available abstraction principles. Self-referential paradoxes like Russell's and Burali-Forti's are then avoided, because there is no way, at any given 'level', to quantify over absolutely everything.

Classical Model Theory

The more prominent approach to Bad Company—and the one I'll be primarily concerned with—assumes Boolos' diagnosis that the inconsistency in Frege's system is due primarily to BLV rather than impredicative comprehension (Boolos, 1993). The idea is to find criteria of acceptability for (classes of) abstraction principles that are necessary and sufficient for the exclusion of APs that are either jointly or singly inconsistent, or otherwise unacceptable, all without excluding APs needed for the development of mathematics. Exactly how much mathematics that should include is contentious—Studd (2015) argues that there should be acceptable APs which allow for the development of a good deal of set theory—but at the very least HP must count as 'good'.[26]

Heck's (2011d) *stability* criterion, and Wright's (1997) conservativeness criterion are among the first.

An AP is conservative, in Wright's special sense, roughly if it implies no restrictions, cardinality- or otherwise, on the domain of a theory to which it is added. So, if your theory is about zebras, but doesn't have anything to say about how many zebras there are, your AP shouldn't demand that there are only 37 zebras; nor that they are solid red. This form of conservativeness has been given a formal presentation which is known as Field-conservativeness

[26]The first half of chapter 6 establishes some baseline acceptability criterion in the context of the mathematical goals of neo-logicism.

(because of a suggestion made by Hartry Field), but the details need not concern us here. An AP is stable roughly if it is satisfiable on (first-order) domains larger than some particular cardinality.[27]

Note that, on pain of excluding HP, the AP in question may require there to be infinitely many objects, just not that there are infinitely many zebras. The stability and conservativeness criteria have been strengthened in various ways, and those strengthenings combined in various way to produce stronger and stronger acceptability criteria, which have been subject to more and more complex counterexamples and countermodels (see Linnebo and Uzquiano, 2009; Linnebo, 2011; Cook, 2012; Cook and Linnebo, 2018).

This dissertation, and especially chapters 5 and 6, bring us closer to both a diagnosis of why attempts such as these have so far failed, and suggest a methodological shift that I believe will bring us closer to a solution to Bad Company. More specifically, I argue that the epistemological concepts central to the neo-logicist project be bought into our analyses and construction of acceptability criteria. I'll say a bit more about that in the next section after I introduce the preceding chapters.

1.5 Outline

Themes relating to methodology, epistemology and Bad Company run through this entire work, but it is nevertheless useful to think of it as being divided into two parts. Chapters 2, 3, and 4 cover central issues in neo-logicism, often from a methodological perspective, and contain a fair amount of historical work. They also set up the positive proposals in chapters 5 and 6. In what follows, I will try to draw out the overarching themes and connections between the chapters, while briefly summarizing them.

Part I

Chapter 2, "The Logic of Meso-logicism", focuses on the status of second-order logic, but from a historical perspective. For neo-logicism to get off

[27] Appendix B is a comprehensive list of proposed model theoretic acceptability criteria.

the ground, not only does second-order logic need to have a foundational epistemic status akin to that often attributed to first-order logic, but it needs to be differentiated from type theory and set theory. The term 'meso-logicism' (suggested to me by Øystein Linnebo when I presented an early version of the chapter in Oslo) refers to the period between the mid-1940s and 1983—after the heyday of logicism, and before the launch of neo-logicism with Crispin Wright's book, *Frege's Conception of Numbers as Objects* (Wright, 1983).

The chapter traces thinking about the differences between higher-order logic (especially second-order logic), type-theory and set theory through that period. I focus only on philosophers writing about neo-Fregean philosophy of mathematics or about logicism, as it is the understanding of these distinctions among those philosophers that is most relevant to the development of neo-logicism. What I then show is that distinctions between higher-order logic, set theory and type theory were slowly sharpened, and that despite Quine's influence (see esp. Quine, 1970), there was space for higher-order logic to be considered as logic proper. Without that space there would hardly have been room for neo-logicism to get off the ground.

The next chapter, "An Entreaty to Neo-logicists" (chapter 3), is a call for neo-logicists to broaden their approach to the formal goal of finding APs that will allow us to capture enough mathematics to be considered foundational. The thought is that we have little reason to think we'll find a way to provide a neo-logicist reduction of set theory directly, and some reason to think that we won't, thus we ought to look in other directions if we're to give neo-logicist foundations for *mathematics* (rather than just arithmetic). I take category theory as my primary example because Shay Allen Logan has recently done some promising work towards a neo-logicist reduction of category theory (Logan, 2015, 2017). Category theory is also our paradigm example of an 'alternative' foundation (Landry and Marquis, 2005; Linnebo and Pettigrew, 2011).

In the latter part of the paper I also suggest that we might be better off working with a general system of APs without the explicit goal of capturing some extant mathematical theory. The possibility is supported by both cases from the history of mathematics, and more recent developments involving

systems of APs that fall outside the confines of the sort of neo-logicism I'm concerned with here (see §1.4, above).

In broader terms, this chapter is important setup for the rest of the dissertation for a couple of reasons. First, not only do I argue for a methodological shift, much as I do in especially the second part of the dissertation, but I set up how I'm thinking about the goals of neo-logicism throughout the dissertation. Second, there are a number of places in the subsequent three chapters where I point out that we need not focus on set theory as *the* foundational theory to be captured by a neo-logicist reduction. This chapter is a way to give a bit more substance to those claims. Furthermore, some of the outlines of possible solutions to various problems for neo-logicism (see §1.3 above) I sketch in the final two chapters support (or are well supported by) the sort of general system of APs I advocate in chapter 3.

The final chapter in what I've been thinking of as Part I of the dissertation—chapter 4, "Is Hume's Principle Analytic?"—tackles another tricky question for neo-logicists, the nature of the special epistemic status thought to be held by APs.[28] In particular we raise (and answer) the question of whether HP is analytic *according to Frege's definition of analyticity*. To answer that question we first pin down what we think is a good explication of Frege's understanding of analyticity—most importantly that an analytic statement can't be consistently denied within the sphere of a 'special science'—and lay out a system of hypernatural numbers that we claim, along with Mancosu (2016), provides a plausible answer to "How many?" questions but disagrees with FA in important ways.

We rely on the strong distinction between set theory and second-order logic discussed in chapter 2 to establish non-standard analysis (NSA) as a special science in Frege's sense, and use that determination to show that, whether we take HP as primitive or as dependent on Frege's definition of number[29], HP can't be considered analytic in Frege's sense. The exception

[28]This chapter is a lightly adapted version of an ipsonymous paper co-authored with Eamon Darnell, currently under review. He and I deserve equal credit for the paper, though I did more than 50% of the actual writing.

[29]Frege's definition of number says that the number belonging to F is the extension of the concept 'equinumerous with *F*'.

is if a very particular sort of solution for Bad Company is adopted. If we were to accept a solution to Bad Company that privileges HP as giving *the* correct identity criterion for cardinal number *necessarily*, our counterargument might be skirted. All of this requires pulling together threads relating to Bad Company, Good Company and questions about which abstracts can be identified. This sets up those problem for more direct treatment in the next two chapters, as well as highlighting some of the connections between the central objections to neo-logicism outlined above.

Part II

Chapter 5 ("Canonical Equivalence Relations") is an attempt to explicitly pick out a notion implicit in much of the literature on neo-logicism and APs. The thought is that APs with the same equivalence relations as all or part of their RHSs pick out (some of the) same abstract objects. For example, the result of restricting BLV to concepts that aren't universe sized (equinumerous with the concept $V = [x : x = x]$), know as New V[30] ostensibly picks out the same sort of abstracts as BLV (would), or any other restriction on BLV does: extensions. Ideally though, we ought not rely on our intuitions or implicit commitments concerning abstract identity, but rather try to ground them in something more robust. I do this by identifying what I call the *canonical equivalence relation(s)* associated with a given AP.

To arrive at a method for identifying canonical equivalence relations, I make make heavy use of both the metaphysical and epistemological aspects of content carving. The overarching idea is that there is no metaphysical or epistemic gap between the equivalence classes based on canonical equivalence relations and the abstract objects that they are associated with, whereas information is lost and a metaphysical gap created by e.g. cardinality restrictions.[31]

[30] This restriction renders the AP consistent. See appendix A for formal definitions.

[31] I'm sure this sounds rather vague and confusing. I'll just say that there's a reason I ended up needing to write a fairly lengthy chapter to establish a notion that was originally meant to be part of a short definition in the final chapter.

After providing what I think is a non-circular method for identifying canonical equivalence relations, I outline how that notion can be used to move us closer to solutions to the C-R problem and Good Company.

Before moving on to my use of canonical equivalence relations in the context of Bad Company in the final substantive chapter, I'd like to mention one more thing. Chapter 5 covers much more conceptual ground than any other chapter, but is also the most speculative. I make a number of suggestions relating to how things might work—content carving for example—without working out the details. I think what I have to say is valuable, but ultimately my hope is that I've adequately described a genuine concept that's genuinely useful, even if my route is perhaps a bit contentious.

Finally, chapter 6, titled "Goldilocks and the Fishes", makes use of the notion of a canonical equivalence relation to tackle a particularly difficult aspect of Bad Company, the problem of fishiness.[32] The chapter begins with a fairly extensive introduction to Bad Company, and particularly model theoretic solutions thereto. I am then in a position to look at the APs identified as fishy by Cook and Linnebo (2018), diagnose the problem, and explain why previous attempts to rule out fishy APs have been unsuccessful.

The basic idea is that APs we've identified as fishy are such that their RHSs tell us *that* the mathematical universe is a certain size without telling us *why* it's that size. In other words fishy APs introduce an explanatory gap between the cardinality of the (abstract, mathematical) domain needed for their satisfaction, and the nature of the abstracts or equivalence classes it's associated with. This is a symptom of the problem identified at the very start of this introduction. In all of our desire to find a formal, model-theoretic solution to Bad Company, we've lost sight of the epistemic goals that are so important to neo-logicism as a whole.

In light of that diagnoses, I suggest that we add an explanatory criterion to our model-theoretic acceptability criteria. I believe that doing so will bring us much closer to a solution to the Bad Company objection as it's faced by Scottish

[32] Wright (1998) and Heck (2011d) both use the word "fishy" to describe certain kinds of APs (the Heck Principles of §1.3), but to my knowledge it is in Cook and Linnebo (2018) that something like a problem of fishiness is identified.

neo-logicists, and thus much closer to justifying the epistemic significance of abstraction principles.

Chapter 2

The Logic of Meso-logicism

> *As the title indicates, this paper concerns, not a contemporary trend . . . but an old question in regard to which developments have now come to a conclusion, or at least a pause. It is not true that opinions now agree. But the cessation of active development means that the matter can be summed up and even that some attempt may be made at adjudication.* –Alonzo Church, "Mathematics and Logic", 1960. (Church, 1962)

During the first half of the twentieth century logicism was a prominent position in the philosophy of mathematics, but the logicism of that time was largely based on that espoused in Russell and Whitehead's hugely influential (and monstrously long) *Principia Mathematica* (Whitehead and Russell, 1927) that took pride of place among logicists of the time. By this I mean first, that it was reductions to the theory of types that were most often taken to be the core of logicism, and second, that Frege's *Grundgesetze* was largely unknown. Even those familiar with Frege's work were more familiar with the *Grundlagen*,[1] and widely held that *Frege's* logicist project was well and truly scuttled by Russell's discovery, in 1902, of the paradox that now bears his name.

Herein I will discuss the developments during the middle of the twentieth century that, in a certain sense at least, lead to the development of the Fregean brand of neo-logicism that was conceived in the mid 1980s by Crispin Wright

[1] See for example, the quotation from Quine's *Philosophy of Logic* near the beginning of §2.2.

(1983), and Bob Hale (cf. Hale, 1987; Hale and Wright, 2001). I will begin with a brief overview of the logicism(s) of Frege, Russell & Whitehead, and Carl Hempel to illustrate the tradition with which the work to be discussed in the remainder of this chapter is concerned. Beginning in §2.2 I will adopt a more thematic approach, looking first at the developments in what was taken to count as logic by those concerned directly with logicism. That will lead naturally to a discussion of Charles Parsons' (1965) sketch of Frege's Theorem (the recovery of Peano's axioms from Hume's Principle—HP) and related issues regarding the development of arithmetic in a Fregean setting (§2.5). The final section (§2.6) briefly discusses the Context Principle and the related *syntactic priority thesis* which lay the semantic and metaontological groundwork for both Frege's logicism and Scottish neo-logicism.

I take it that those components—our understanding of logic, the (re)-discovery of Frege's Theorem, and *neo*-Fregean understandings of semantics and ontology—provide the core of any position that can claim to be neo-logicist in a truly neo-Fregean sense. In this chapter I will be less concerned with semantics and ontology, focusing almost exclusively on the logic of meso-logicism—I have termed advancements in thinking about logicism during this period *meso-logicism* as they fall between the large scale rejection of logicism indicated in the quotation at the beginning of this piece, and the rise of *neo-logicism* in the 1980s.[2,3]

Before diving in, one caveat is in order. A comprehensive survey of the aforementioned material, even during the relatively short period between 1945 and 1985 would fill a rather large book. With this in mind, I have focused on a few influential papers and authors in the hope that they are representative of the shifts in thinking that occurred during this period, thus my aim is to provide a something of a narrative, rather than a comprehensive historical account.

[2]The term meso-logicism was suggested to me by Øystein Linnebo.

[3]An excellent survey of the first two decades of neo-logicism, with a particular focus on semantic and metaphysical issues is provided by Fraser MacBride (2003).

2.1 Beginnings

Frege

Frege's attempted reduction of arithmetic in the Basic Laws of Arithmetic (*Grundgesetze der Arithmetik*, hereafter *Grundgesetze* (Frege, 1893, 1903, 2013)) to logic involves three major components: the concept-script, six "basic laws", and definitions of the basic concepts of arithmetic. Together, three components allowed Frege to derive the (second-order) Dedekind-Peano axioms of arithmetic from what he took to be purely logical grounds. That reduction of arithmetic to logic was meant to show that arithmetic is analytic, *contra* Kant.

The concept-script (in German, *Begriffsschrift*), first presented in his eponymous book of (1879) and expanded on and modified in §§2–52 of the *Grundgesetze*, is the logical formalism that was designed to represent mathematical reasoning entirely without gaps or ambiguities, and is usually considered to be the first example of a symbolic classical quantified deductive system.[4] The basic laws are logical axioms not in need of justification, though they are related to certain metaphysical theses. The definitions are explications of the arithmetical vocabulary—specifically 'natural number', 'predecessor', and 'zero'—in the language of the concept-script.

As is well-known, Frege's reduction as presented in the *Grundgesetze* leads directly to Russell's paradox. In modern terms, the inconsistency arises from the combination of Basic Law V (BLV), which says that the value-ranges (*Wertverläufe*) of two functions, and thus the extensions of two concepts, are identical just in case exactly the same objects fall under them; and the unrestricted principle of substitution that is part of the background logic. Frege's substitution principle allows for the inter-substitution of concept names and formulae, and is, for our purposes, equivalent to the unrestricted comprehension principle of modern second-order logic.[5] Concepts, in Frege's sense,

[4]This is not to impugn the undoubtedly important work of Boole, Schöder, Pierce, Dedekind or Peano. See (Dudman, 1976) and (Boolos, 1994) for overviews and analyses of the development of formal logic in the 19th century.

[5]In fact, the situation is a bit more complicated than this. However, it is generally accepted that the logic of Frege's *Grundgesetze* is equivalent to modern (full) second-order logic. See (Hale, 2015) for a more detailed discussion.

are functions from objects to truth values (the True and the False, also objects). Extensions are abstract objects representing the collections of objects which fall under specific concepts. Basic Law V then says that there is a unique object associated with every class of coextensional concepts, but if every formula determines a concept, as Frege's substitution principle entails, then there must be more concepts than objects by familiar arguments due to Cantor and Russell.

What is perhaps less well known about Frege's program, even today, are his motivations for pursuing such a reduction of mathematics to pure logic. It was not, as some have thought, that Frege thought that arithmetic *required* a foundation in logic due to some lack of clarity or possibility of inconsistency in the theory of the natural numbers as it was understood at the time; rather Frege thought that by providing such a foundation, he could show that arithmetic needn't rely on Kantian intuition, and thus was analytic *a priori*. It has since been argued that, even had Frege been successful in this philosophical aim, he would have been begging the question against the Kantian given the fairly wide conceptual gap between his and Kant's conceptions of what it means for a statement to be analytic.[6] Whether or not that's actually the case (I think it is, for what its worth), a logicist reduction such as Frege's would provide a promising route to an epistemic foundation for mathematics. It is that particular virtue that has more recently been emphasized by neo-logicists like Hale and Wright.

Apart from an attempted fix rushed to the publisher, Frege never made a serious attempt to rehabilitate his system after Russell's (in)famous letter.[7] Russell (and Whitehead), however did. Russell and Whitehead's solution to the paradoxes was to introduce the ramified theory of types, leading to a predicative system from which the Dedekind-Peano axioms could be recovered with the help of an axiom of infinity. F.P. Ramsey (1926) later pointed

[6] Chapter 4 looks at how Frege understood analyticity.

[7] Frege's attempted solution failed because the modification would have made his system unsatisfiable in models with more than one element. Quine (1955) and Geach (1956) both provide proofs of this, but credit an earlier proof by Leśniewski without citation. The original proof can be found in Sobocinski (1949). Micheal Resnik (1980, pp. 215–19) gives a different proof in more familiar notation.

out that all that was needed was the simple theory of types, and this is where the story I wish to tell actually begins.

Hempel

Carl Hempel's article "On the Nature of Mathematical Truth" (Hempel, 1945) represents something of a culmination of early 20th century logicist thought. Interestingly, Hempel notes that Frege, in addition to Russell and Whitehead, was responsible for the result that formal definitions of **0**, successor, and (natural) number can be given such that "all Peano postulates turn into true statements" (p. 386). He means by this something along the lines of 'fixes the intended interpretation of Peano's axioms'. More significantly, Hempel then sketches the logicist system of *Principia*, ignoring Frege almost entirely. This in itself should be of no great surprise given what I've already said.

Nevertheless, Hempel obviously believed, following Frege, that mathematics (apart from geometry), is analytic *a priori*. Hempel differs from Frege however, in his understanding of analyticity,[8] appearing to hold a more conventional 'truth in virtue of meaning' view. Hempel is also in line with Frege as well as Scottish neo-logicists in taking logicism to be key to epistemically grounding mathematics.

> It is a basic principle of scientific inquiry that no proposition and no theory is to be accepted without adequate grounds. [. . .] But what are the grounds which sanction the acceptance of mathematics? That is the question I propose to discuss in the present paper. (p. 377)

That he thinks that Russell's logicism is the answer to his question is a large part of why his article is so important, especially since he seems to be one of the last major thinkers to hold such a view. Unfortunately, Hempel's views about what counts as an analytic postulate of logic quickly went out

[8]Frege thought roughly that analyticity amounts to reducibility to logic and admissible definitions. See Schirn (2006) and chapter 4 for discussions of Frege's understanding of analyticity.

of favour. As we shall see below (esp. in §2.2), that a certain principle can be defined in purely logical terms would not be enough to establish it as a *logical* principle. Hempel's apparent belief that the concept of class (set) is a logical one did not remain popular for long, though that assumption too, is Fregean in spirit. I say that because not only did Frege take extensions to be logical objects, *Basic Law* V suffices for the definition of a membership relation, too.

2.2 Issues of Logic

As I have already touched on to an some extent, there was a shift in the understanding of the logical basis of logicism from Frege's second-order logic to Russell & Whitehead's type theory, and the type-theoretic understandings of the logical positivists and other early 20th century philosophers of mathematics. By the time Hempel published his influential article in 1945, there was another important change taking place as well.

Set theory(s) based on the work of Cantor, Zermelo, Fraenkel, von Neumann and others were gaining broad acceptance, though there was still much debate about certain axioms and interpretations. In many cases there were not clear distinctions being made between second-order logic, certain type theories, and (first- or second-order) set theory. In other words, set-theory was also often thought of as logic, at least in the sense of "logic" relevant to logicist reductions of mathematics to logic. Quine (1970) makes this point particularly forcefully:

> Pioneers in modern logic viewed set theory as logic; thus Frege, Peano, and various of their followers, notably Whitehead and Russell. Frege, Whitehead, and Russell made a point of reducing mathematics to logic; Frege claimed in 1884 to have proved in this way, contrary to Kant, that the truths of arithmetic are analytic. But the logic capable of encompassing this reduction was logic inclusive of set theory. (pp. 65-6)

Notice here that Quine is talking about the *Grundlagen* (Frege, 1884), and not the later *Grudgesetze* (Frege, 1893, 1903), and is using this passage as a lead

up to his famous argument equating set theory and second-order logic ("Set Theory in Sheep's Clothing") that begins directly thereafter. Essential to the re-emergence of logicism in its Scottish guise however, was the move from thinking in terms of type-theory and instead returning to the Fregean roots of the program in higher-order logic; just as important was the recognition that set theory isn't logic, but nor is higher-order logic set theory. This is not to say that the question of the nature of higher-order logic had been settled, only that space for higher-order logic as logic was reopened.

Implicit in the following discussion is the assumption that Fregean higher-order logic and Russellian type theory are different. From a modern perspective this may seem odd, especially if we are concerned with the simple, rather than ramified theory of types, and if we admit quantifiers of arbitrarily high order as Frege seems to have.[9] We can again look to Quine for a clear example:

> If in response to Russell's paradox Frege had elected to regiment his classes in levels corresponding to those of his attributes, his overall solution would have borne considerable resemblance to that in *Principia*.
>
> Actually, it is not to be wondered that Frege did not think of this course, or, thinking of it, adopt it. It was by having all his classes at ground level that he was able to avoid the use of high-level attributes, and this he liked to do. (Quine, 1955, p. 489)

Quine's diagnosis here is that Russell allows classes of higher types, where Frege allows only 'attributes' (concepts, functions) of higher types, and has classes (extensions, value-ranges) only at the object level. This makes sense

[9] Frege actually has very little to say about this, pointing out that

> It has already been observed in §25 that first-level functions can be used instead of second-level functions in what follows. This will now be shown. As was indicated, this is made possible by the fact that the functions appearing as arguments of second-level functions are represented by their value-ranges, although of course not in such a way that they simply concede their places to them, for that is impossible." Frege (2013, vol. 1, §34, p. 52)

Frege is basically saying that we need not go beyond second-order logic, though third-order entities do exist. He doesn't, to my knowledge, go beyond third-order.

if we are following Frege in maintaining a clear distinction between concepts as (in part) intensional entities, and their extensions which are picked out purely extensionally. This is a distinction easily lost on those of us treating second- (and higher-) order variables as purely extensional, however. I do not wish to pronounce on the viability of a distinction between type theory and higher-order logic, but it appears that such a distinction was a common assumption during the middle part of the twentieth century. A couple of points are, however, in order. First, there are at least a few differences in the metaphysical assumptions of those using higher-order logic, and those using type-theory. For example, type-theorists, following Russell, tended to (and often still do) have concerns about predicativity. Indeed, this is part of the reason Russell originally thought he needed to *ramify* his theory of types. In contrast, Frege and neo-Fregeans are generally happy with impredicativity, at least in most mathematical contexts.

In a similar vein, those working in the Russellian type-theoretic tradition have tended to follow Russell in taking numbers (and other mathematical objects) to be higher-order entities—the number four is the class of all four-membered classes—whereas those following Frege take numbers to be objects—the number eight is an object that can itself be counted, and is the same object picked out by the term 'the number of planets'. For Frege these were a subclass of the extensions, for neo-logicists they are *sui generis* objects. So, whether or not there is a *formal* distinction to be made between type-theory and higher-order logic, the distinction is a useful one for the purposes of discussing approaches to (neo-) logicism.

Hempel and Benacerraf

Although is would be a huge task to track all of the changes in thought concerning what counts as logic through the mid-twentieth century, some more concrete examples of papers that are in any case important in the development of (neo)-logicism will be illuminating. Hempel (1945), as Benacerraf (1960) points out, is obviously thinking in terms of type theory, but nevertheless includes '$\in$' as a basic logical particle—as among the syncategoramata—and

includes the axioms of choice and infinity among the logical axioms. As I pointed out above, this was because Hempel took a much broader view of logicality than has since been mainstream.

Benacerraf's 1960 dissertation "Logicism, Some Considerations" is in a large part an extended discussion of Hempel's article, and Benacerraf also takes a type-theoretic approach as his example of a possible logicist reduction in the first two-thirds of that work. Although he does not say so explicitly, he takes direct aim at Hempel's assertion that the axiom of infinity is a part of logic "because it is statable in purely logical terms" (p. 196, note 1). This is notable not just because of his explicit rejection of a view about logicality that had previously been fairly widespread, but also because of the argument he advances against the logicality of the axiom of infinity. His argument is based on the fact that the truth of the axiom is contingent on the range of the quantifiers. Such arguments were then sharpened and advanced in broader discussions of what counts as logic as we shall see in §§2.3 and 2.4 below.

Earlier, in chapter III of his dissertation, Benacerraf makes another important distinction that sounds odd to the modern ear, but is indicative of discussions of logicism even at that later time.

> The truths of logic (and this at least for our purposes here includes set theory) fall "naturally" into three groups. The first is the propositional calculus. The second includes what is variously referred to as 'quantification theory,' or 'first-order functional calculus (without identity).' The third group includes what we will call set theory, *ignoring for the moment traditional distinctions between type theory and set theory.* (p. 75, my emphasis)

As it can already be inferred from the first parenthetical, Benacerraf is hesitant, at the very least, to actually include set theory as logic proper, but that he *can* include that as an assumption is telling. There are a few other important things to note here. The first is that he is equating type theory and set theory, despite recognising that there is a distinction to be made. This is an assumption he makes frequently, though I think it is a matter of strength and/or ontological commitment rather than some inherent 'logical'

similarity. But again, this is just evidence (to my mind) that such distinctions remained somewhat vague and perhaps not so well understood in the '50s. The second is the notable omission of higher-order logic. This could be due to either the long-standing impression that Frege's program had failed utterly, and so it was the logicism of the *Principia* that was relevant, or (and that's inclusive) that higher-order logic can be treated in the same way as type- and set-theories. The latter suggestion is particularly intriguing in light of Quine's views as articulated a decade later (and mentioned above). The final interesting assumption is that identity is not to be included in first-order quantificational logic. While that isn't particularly important for the current discussion, it is certainly not very Fregean.

Already here we can see that the questions of whether set-theory is equivalent to type-theory and whether either is logic in the sense needed for the success of a logicist philosophy of mathematics (beyond the purely technical aspect at least) were still very much open questions. Benacerraf at least, was highly skeptical that set-theory was logic in the right sort of way. In fact, much of the later part of his dissertation argues against what we might call 'logicism via set theory'. Notably, threads from that work were developed later in his two seminal papers "What Numbers Could Not Be" (1965) and "Mathematical Truth" (1973), though those articles are for the most part not as directly concerned with logicism as his dissertation.[10]

2.3 Two Views on Higher-order Logic

While Hempel, Benacerraf and others' focus on type and set theories stemming from an identification of the logicist project primarily with Russell's work was prevalent at least until the 1970s, Alonzo Church and especially Charles Parsons had things to say more directly about higher-order logic.

In his "Mathematics and Logic" (Church, 1962), Church does not make a clear distinction between type theory and higher-order logic, though he is almost exclusively concerned with Russellian logicisim, having brushed

[10]The former (Benacerraf, 1965) starts of with a parable that directly invokes logicism, but the point is a broader one about under-determination in the foundations of mathematics.

Frege aside because of Russell's paradox (p. 183). That said, Church is quite clear that he regards higher-order logic as logic proper, and that it, as well as simple type theory, is distinct from set theory. To the first point, he states that

> ... $\forall p(p \vee \neg p)$; and in a similar fashion $\forall F \forall y(\forall x Fx \rightarrow Fy)$ may arise by generalization from infinitely many analytic sentences of the appropriate form. [...] Hence they are considered to belong to logic, as not only is natural but has long been the standard terminology.[11] (p. 182)

This makes clear that Church's classification of higher-order logic *as logic* is a result of his understanding of logic as having to do with generalizations of analytic sentences and the relations between those generalizations. That does not however, impugn his general acceptance of higher-order quantification to any great degree (if at all). It appears Church's largest problem with what he calls 'the logistic thesis' is Russell and Whitehead's need to lay down an axiom of infinity, a principle Russell himself famously admitted was probably not part of pure logic.

The last sentence of the above quotation is also very telling. From Church's perspective, it was standard in 1960 to countenance sentences including higher-order quantifiers as purely logical. He does acknowledge that that view is not universal, though he *does* say that denying that the mentioned sentences are part of logic would be an 'extreme position' (p. 183).

Church is less forthright about differentiating between higher-order logic (or type theory), and set theory, but he clearly writes of set theory as an *alternative* to logic, rather than a part of logic. Since he is explicitly dealing with Russellian logicism, in making this distinction Church would not have been able to consistently maintain that reductions of arithmetic to set theory, no matter how weak the set theory, would count as logicist in the relevant sense. Here he is in agreement with Benacerraf (1960, see §2.2 above).

[11] I have replaced Church's symbolism with a more modern variant for readability and notational consistency. Also note that the first formula is a sentence of quantified propositional logic.

Charles Parsons takes a similar line in his influential essay, "Frege's Theory of Number" (1965), though he is less explicit in his arguments for the theses that set theory isn't logic, or that higher-order logic might be. Parsons is concerned mostly with a direct discussion of the viability of Frege's brand of logicism, and rather uniquely at that time, the system of the *Grundgesetze* rather than the earlier *Grundlagen*. In particular he is taking Frege's system as a starting point, and thus accepts higher-order logic as logic "provisionally" (p. 198). This suggests that Parsons' own views on the issue were not solidified at that time, but that counting higher-order logic as logic was at least an open possibility. On the other hand, Parsons does seem to lean towards the opposite conclusion. At the beginning and end of the same paragraph (p. 200), he writes

> As a concession to Frege, I have accepted the claim of at least some higher-order predicate calculi to be purely logical systems. The justification for not assimilating higher-order logic to set theory would have to be an ontological theory of concepts as fundamentally different from objects, because "unsaturated".

and

> Thus it seems that even if Frege's theory of concepts is accepted, higher-order logic is more comparable to set theory than to first-order logic.

The reasons he gives are essentially the same as the Quinean worries about the ontological commitments of higher-order logic (about which I will have more to say in §2.4). So it appears that Parsons was thinking along roughly Quinean lines about the status of higher-order logic, or at least full second-order logic.[12] Whatever Parsons thought about the status of higher-order logic at the time, his 1965 paper takes the foundational status of Frege's

[12]Certain passages in (Parsons, 1965) suggest that second-order logic with predicative comprehension, or with Henkin semantics might be more like first-order logic, as he expresses worries about impredicativity and the need for the full power set of the first-order domain for the standard semantics. The former case is of some interest but has only been investigated in this context much more recently. The latter is of less interest as Henkin semantics essentially

higher-order logic as a working assumption, which undoubtedly helped lead Parsons to his discussion of Frege's Theorem in that paper (see §2.5 below).

> It is denied that set theory is logic. It is also denied that a reduction merely to set theory will suffice for a philosophical foundation of arithmetic or for a refutation of the epistemological theses about arithmetic (e.g., Kant's and Mill's views) against which the reduction is directed.

That could hardly be more clear. It should be noted that Parsons is taking aim here, not just at the idea that set theory might be part of logic, but also at Frege's Basic Law V which introduces what are effectively classes or sets. He follows Benacerraf's reasoning here, the basic idea being that set theory isn't topic neutral in the same way that the predicate calculus is.

What should be fairly clear at this point, is that by the mid-1960s, the distinctions between type theory, set theory, and higher-order logic were still vague, but becoming slowly sharper. It was also widely agreed among those working on logicism that set theory was not logic, and so reductions of other mathematical theories to set theory were not logicist reductions in the strong sense we are here concerned with. More work was still needed on the status of higher-order logic before a neo-Fregean logicist revival would look tenable.

2.4 Boolos on Second-order Logic

As is well known, the late '60s and and early '70s in particular saw Quine arguing for the primacy of first-order predicate logic, and the identification of higher-order logic with set theory, but Quine's was not the only game in town. The mid 70s also saw important developments in the understanding of logicism and the logic behind it. First among these is Boolos's article, "On

leaves us with a two-sorted first-order theory.

Hale (2015, p. 11–12) argues that (early '60s) Parsons takes second-order logic to have similar ontological commitments as set theory, which I more or less agree with, though I don't think that's enough, in this context, to make definitive judgments about Parsons' more general view of higher-order logic. In that paper, Hale also develops an interpretation of second-order logic that may ease the minds of those with Parsons-like concerns.

Second-Order Logic" (1975), in which he argues that second-order logic is certainly *not* set theory, but *can* make many of the same claims to logicality as first-order logic, though he falls short of claiming that second-order logic *is* logic. His arguments are not only important because they have furthered our understanding of second-order logic, but also because of the importance he sees his results as having for the debate about logicism. In the first footnote he describes his motivation thus:

> I am inclined to think that this association [of a truth of second-order logic with each truth of arithmetic] is the heart of the best case that can be made for logicism and that unless second-order logic has *some* claim to be regarded as logic, logicism must be considered to have failed totally. I see the reasons offered in this paper on behalf of this claim as part of a partial vindication of the logicist thesis.[13] (p. 37, fn. 1)

From this it can already be seen that Boolos's paper is a significant departure from the understanding of logicism that we have seen had been prevalent post Russell, which will become all the more obvious with a look at some of Boolos's arguments.

There are two main threads in Boolos's paper relevant to the current discussion: that second-order logic isn't set theory, and that there are reasons to consider second-order logic as *logic*. Without going into much detail, I will rehearse what I take to be Boolos's most important arguments for both of these theses. First, that second-order logic isn't set theory, at least under Quine's

[13] Lest I be accused of cherry picking quotations, I should mention that the following sentence precedes the above quotation.

> [The association of] a truth of second-order logic with each truth of arithmetic. . . can plausibly be regarded as a "reduction" of arithmetic to set theory.

This sentence strongly suggests an association between, or even a Quinean *identification of*, second-order logic and set theory. Two explanations for this occur to me. First, it may be that the influence of Quine on the one hand, or those who identified type-theory and second-order logic on the other may have been ingrained even in Boolos's thinking at the time. The second is that that sentence is a bit of unfortunate sloppiness, though this latter explanation could easily be grounded in the former. In either case, that seeming identification of set-theory and second-order logic runs contrary to the central themes of the rest of the paper.

(1970) assumption—that e.g. $\exists F \forall x Fx$ really means $\exists \alpha \forall x x \in \alpha$—the class of valid sentences changes its extension, and furthermore, there are implications that hold in one syntax and not the other. The two formulae just mentioned are Boolos's example (1975, p. 40)—the first is valid, but the second is not. If, then, we are to take Quine seriously, logicians starting with Frege have been massively in error, which, given even just extrinsic reasons (in the sense of Gödel and Maddy), seems highly implausible.

The second argument, which I take to be similar in kind, amounts to the simple observation that set theory, as we generally think of it, makes quite vast existential assertions, while second-order logic only asserts the existence of one single object, usually taken to be the empty set on standard semantics (p. 48).

The third argument for the thesis that second-order logic isn't set theory may be less familiar than the first two, and is premised on the observation that the truth of many quantified second-order sentences depends on the domain over which the quantifiers are taken to range, while the quantifiers of (first-order) set theory are taken to range over sets, but importantly not a single domain of *all* sets (p. 43ff).[14] To generalise a bit, the point is that second-order sentences can be given different interpretations, whereas sentences of set theory can only be interpreted as ranging over different models of set theory. Note that this is a sharpening of worries expressed by Benacerraf (1960) and Parsons (1965) mentioned above.

Resnik in 1980

To my eye, the previous three arguments succeed in showing that second-order logic is not set theory, especially as the obvious response—that we've been doing second-order logic wrong all this time—is question begging, since we know from Gentzen that second-order logic is consistent. This puts a hypothetical mid-'70s neo-Fregean logicist on good ground, as the view that

[14] This may be one topic where the use of reflection principles in set theory may be particularly useful, at least heuristically, the thought being that any quantifier in first-order set theory can be taken to range over *some* set without loss of generality.

set theory is part of logic was already in serious decline a decade and a half earlier.

Further evidence of the continuation of that decline can be found in the final chapter of Resnik's book *Frege and the Philosophy of Mathematics* (Resnik, 1980). Resnik takes the view that set theory isn't logic as standard.[15]

> [O]n the grounds of complexity and lack of certainty, set theory would not belong to the same epistemological category as first-order logic (if any satisfactory concept of logical truth is to be had). (p. 225)

This is important not just because it is good evidence that the separation of logic and set theory was generally accepted among philosophers writing about Fregean logicism by 1980, but also because of the emphasis on epistemology. Resnik is particularly concerned with logical truth, which makes sense given his focus on Frege himself. The more general concern that whatever we take to count as logic needs to be in "the same epistemological category as first-order logic" goes to the heart of how higher-order logic needs to be viewed—as having the same epistemic import as first-order logic—if Scottish neo-logicism is to be successful. That set theory doesn't have such import would not only mean that set theory isn't logic, but that reductions of other mathematical theories to set theory are no good to the (neo-)logicist.[16]

As to whether second-order logic is logic, what Resnik has to say is less hopeful. He writes,

> Today. . . second- and higher-order logics have been interpreted as carrying a commitment to sets. Thus, it has been argued that second and higher-order logics have no clear title to a place in the domain of true logics. (p. 205)

This quotation doesn't give us much information about what Resnik himself thought about the matter, but it does tell us that Quine's influence was

[15] It's worth noting that Resnik cites Boolos (1975) when raising the issue of higher-order logic on page 205.

[16] Note that the first claim plausibly implies the second, but not the reverse. The epistemic status of set theory need not be tied directly to its logicality.

much stronger than Boolos's in the late '70s.[17] On the other hand, Resnik leaves plenty of room for an interpretation of higher-order logic that eschews any commitment to infinite sets, giving it a claim to status as logic.

So by 1980 set theory generally wasn't considered logic, but nor was higher-order logic. However, the door was closing on the first position and opening on the second.

Before wrapping up I will discuss one further development of possible importance, and certainly of great interest to the development of neo-logicism: Frege's Theorem.

2.5 Frege's Theorem

Perhaps the most important contribution to the development of (neo)-logicism in our period of interest comes from Charles Parsons in 1965. Parsons gives a sketch of what is now known as Frege's theorem, which says that analogues of the Dedekind-Peano axioms can be derived from HP and Frege's definitions of arithmetical concepts without the use of the inconsistent BLV.[18] I say 'perhaps' because Crispin Wright gave a much more detailed sketch of the same proof in his 1983 book (Wright, 1983), but was not aware of Parsons' paper at that time (Wright, 1999, note 1).

Parsons leaves out a lot of detail, and makes a small error (he uses the ancestral instead of the weak ancestral in the definition of natural number), but this is forgivable, since as far as anyone I have read is aware, this is the first time Frege's Theorem had been discussed in print (post Frege). That said, there is a worry here, expressed by Richard Heck in the introduction to (Heck, 2011b), that Parsons' apparent discovery of Frege's Theorem was based on the erroneous assumption that Frege's own proofs were enough to establish

[17] My impression is that this continues to be the case outside of our fairly small circle of neo-logicists, though I also have a sense that tides may change with a new generation of philosophers of mathematics.

[18] It has been suggested by Heck (2011c) and others, that Geach's (1955, p. 293) claim that classes or extensions are not needed to guarantee an infinity of naturals is tantamount to a statement of Frege's Theorem, however such a short passage (1 paragraph) that contains no mention of HP, equinumerosity, etc. can hardly be seen as such as a statement of Frege's Theorem without further evidence or elaboration.

the derivation of the Dedekind-Peano axioms without the use of BLV. Either way, a detailed examination of Parson's assertion may very well have led to an actual proof of Frege's theorem.[19]

Of course, for Frege's Theorem to have any philosophical significance, we must be reasonably certain that HP is consistent when added to second-order logic. Peter Geach (1976, pp. 446-7) hints at a model that might do that important work. Heck has this to say about Geach's construction:

> [Geach's model] verifies HP. The argument can easily be formulated in, say, Zermelo set-theory, so the consistency of HP is thereby assured, if we accept the axioms of Zermelo set-theory. (Heck, 2011c, p. 6)

That was certainly a welcome reassurance, though a consistency proof in a theory closer in strength to arithmetic would have of course been desirable. Once Wright published his book, proofs of the mutual interpretability of Frege Arithmetic (HP plus full second-order logic) and second-order Peano Arithmetic came thick and fast (Burgess, 1984; Hazen, 1985; Boolos, 1987a), but such proofs should then be counted as part of modern neo-logicist project, not meso-logicism.

2.6 An Extra-Logical Ending

Once higher-order logic is considered as logic in the sense needed to provide a part of an epistemically sound foundation for mathematics, Frege's Theorem proved, and HP (and thus Frege Arithmetic) shown to be consistent, the project left for the neo-Fregean neo-logicist is mostly one of metaphysics and epistemology. First, the existence of numbers as abstract objects—platonism about arithmetic—needs to be established. More importantly, since BLV has been jettisoned and with it Frege's derivation of HP from BLV and an explicit definition of (cardinal) number, the stipulation of the truth of HP must be

[19] Heck notes that Wright claims to have given a decent portion of the needed proof in a draft of his BPhil thesis of 1969 (Heck, 2011c, p. 4), but was unable to work out the details of the proof of the successor axiom, the critical step in the proof.

justified. The adoption of principles based on what are now know as Frege's *context principle* and *syntactic priority thesis* go some significant way towards those extra-logical goals, though not the whole way.[20]

The context principle says that it is only in the context of a sentence (or proposition) that that the meaning of a word can be determined. In volume II of the *Grundgesetze* Frege says, "One can ask after reference only where signs are components of propositions expressing thoughts." And of perhaps more obvious relevance, in the *Grundlagen*:[21]

> The self-subsistence that I claim for number must not mean that a number word designates something outside of the context of a proposition, rather I want only therewith to exclude their use as predicate or attribute, whereby their meaning would be somewhat altered. (Frege, 1980, §60)

The context principle gives a starting place for the justification of HP and other similar principles (abstraction principles, at least when they are thought of as partial implicit definitions, as they provide canonical identity conditions for mathematical objects in terms of previously understood concepts. The context principle also works together with the syntactic priority thesis to help establish the existence of abstract mathematical objects.

The syntactic priority thesis is, very roughly, the idea that barring any compelling evidence in specific cases, the syntax of natural language is a good guide to ontology. If number words behave as singular terms, and as Frege told us in the above quotation, not as predicates or attributes, there must be objects that those terms refer to. Obviously there is much more to be said, but again we have a starting point.[22]

[20] It was originally my intention to include a fuller discussion of the the meso-logicist views about these theses in this piece, but it turned out to be too large a project. I hope to say more about the semantics and metaphysics of meso-logicism in the future.

[21] Øystein Linnebo (2018) identifies six passages where Frege mentions the context principle, four from the *Grundlagen* (p. x, §60, §62 and §106) and two from the *Grundgesetze* (I, §29, and II, §97).

[22] I direct the interested reader to the relevant sections of (Linnebo, 2018; MacBride, 2003; Dummett, 1991). See also §1.2.

With an understanding of the logic, mathematics, metaontology and semantics of Frege's logicism in place, the elements of Frege's philosophy of mathematics could be put to use, and his mistakes avoided, making the birth of neo-logicism possible.

Chapter 3

An Entreaty to Neo-logicists

The goal of the neo-logicist program is to establish epistemic, formal, and maybe also metaphysical foundations for mathematics. The general strategy, at least in the Scottish school, is to show that portions of mathematics can be reduced to higher-order logic and abstraction principles. If this can be done, the task left for neo-logicists is that of defending or explaining the foundational epistemic status of abstraction principles.[1] Recently, more attention has been paid to the second goal, though the first is rarely far from our minds.

Before proceeding, we should remember that our paradigm example of a good abstraction principle is Hume's pricniple (HP) which, in the presence of full second-order logic and appropriate definitions, allows us to prove the second-order Dedekind-Peano axioms of arithmetic. This result is known as *Frege's Theorem*; the system consisting of full second-order logic plus HP is known as *Frege Arithmetic* (FA—see §2.5). This is an impressive result by any light, and is central to the hope that we can squeeze more mathematics out of abstraction principles, though that's despite the fact that recent results, due especially to Roy Cook (2017) and Walsh and Ebels-Duggan (2015), suggest that HP is unique beyond just its role in Frege's theorem.[2] Such results

[1] There is also the task of defending the status of higher-order logic as logic, but that can be left aside for now. See (Boolos, 1975), (Shapiro, 2002) and chapter 2 for discussion.

[2] Walsh and Ebels-Duggan show that only systems based on a small class of abstraction principles including HP have certain categoricity properties, and Cook proved a similar result regarding invariance properties.

may indeed be worrisome to a neo-logicist hoping to get beyond arithmetic, but I'll say no more about that here. Rather, what is currently of interest is how neo-logicists might capture enough mathematics for neo-logicism to plausibly be considered foundational. In particular, I want to argue that our focus on reducing some significant portion of *set theory* to higher-order logic and abstraction principles is too narrow, and perhaps even misguided.[3]

Of course, set theory is a familiar and convenient tool for talking about and measuring mathematical strength, as well as looking at the meta-mathematics of neo-logicism more generally. That doesn't mean, however, that we ought to expect a strong mathematical theory based on abstraction principles to look anything like ZFC.[4] To be clear, I am not arguing that we should stop looking for abstraction principles that will give us set theory, only that the neo-logicist project would benefit from taking a broader view of higher mathematics. That said, there are reasons to think that finding an abstractionist route to ZF-like sets is unlikely, as we shall see.

There are two paths to neo-logicist foundations that immediately suggest themselves. The first is a piecemeal approach: Frege's Theorem is the route to arithmetic/analysis; an abstraction principle for Dedekind cuts or Cauchy sequences (say) then gives us the reals; we can get *some* sets from New V or Newer V, or some ordinals from the Size-restricted Ordinal Abstraction Principle (SOAP, see Shapiro and Weir, 1999); and we then go looking for abstraction principles for other kinds of mathematical objects, starting with areas that are currently important, like the complex numbers, groups, or categories. The idea would be to reconstruct each individual area of mathematics using higher-order logic and abstraction principles, thus epistemically grounding areas where we're successful, and deferring, or eschewing altogether, questions about unified foundations. In other words, the idea would be to try to reproduce the success of HP and Frege's Theorem for other parts of mathematics.

There are a couple of problems with this approach beyond just the usual

[3] By 'set theory' we generally mean ZFC or ZFCU.

[4] Well, there's one reason, which is that Basic Law V (BLV) can be used to define a membership relation, but more on that shortly.

concerns with finding and justifying abstraction principles. The first is a problem of identity and a version of the Cæsar problem, and is a concern for philosophers and mathematicians of many bents.[5] The issue is this: there is a lot of apparent embedding and overlap among mathematical domains: the naturals appear to be a subset of the reals, which in turn look to be embedded in the complex field (which is why this was dubbed the C-R problem by Cook and Ebert (2005)), et cetera. Questions about whether identity claims such as $17_{\mathbb{N}} = 17_{\mathbb{R}} = 17_{\mathbb{C}}$ can be made are serious and tricky on the piecemeal approach. This is made worse by having to deal with interpreting interpretability claims between abstractionist systems that could plausibly be thought of as distinct.

Another issue with the piecemeal approach is more an inherent defect than a philosophical or logico-mathematical question that needs to be overcome. If the goal is just to capture specific mathematical theories as best we can, then we're constrained not only by what mathematicians have already discovered, but also the structures of already extant theories. Neo-logicists' efforts at mathematical reduction will be reduced simply to looking for acceptable abstraction principles that capture the identity criteria for the mathematical objects we already know and love.[6] Depending on how the identity problem from the previous paragraph is solved, the Good Company problem may raise its ugly head, which is to say we'll be in the position of choosing between a variety of abstraction principles that, for all intents and purposes, give us the same mathematical objects (Mancosu, 2015, 2016).[7]

The second path to neo-logicist foundations is to look for an abstraction principle or a system of abstraction principles that is powerful enough to

[5]The Cæsar problem is roughly the problem that abstraction principles leave open questions about the distinctness of abstracts from other objects, or objects presented in different ways, such as the question of whether the number two is identical to Julius Cæsar. See §§1.3, 4.6, and 5.5 for discussion of this and related problems.

[6] Determining which abstraction principles are acceptable requires a solutions to the Bad Company objection which asks how we differentiate between 'good' principles like HP, and inconsistent or otherwise 'bad' principles like BLV. This is further compounded by the the fact that there are abstraction principles that look fine alone, but are jointly inconsistent or unsatisfiable. See §1.3 and especially chapter 6.

[7]See §§1.3 and 5.5 for more on the Good Company problem.

capture (in the sense of a neo-logicist reduction) a significant chunk of modern mathematics. In general such a system will include HP—it is our paradigmatic abstraction principle after all—but there is no in-principle reason to include it other than perhaps certain eventual solutions to the Bad Company objection.[8] It is this second path that will be my primary focus in what is to follow, though the piecemeal approach will be a helpful foil. Before that however, I would like to try to ease the minds of those who find the distinction I've just made arbitrary or otherwise dubious.

I take these two paths to be at once extremes of neo-logicist methodology, and at the same time intimately related. In the latter case, taking what I've dubbed the *piecemeal approach* could very well lead to a unified neo-logicist foundation for mathematics, either by only using (intentionally or not) abstraction principles that can work together, say by being strongly or conservatively stable,[9] or by finding a single abstraction principle that allows for a neo-logicist reduction of a mathematical theory strong enough to interpret a large amount of mathematics.

In practice, most neo-logicist or abstractionist programs fall between the two methodologies I've described, but (I contend) are very often aiming at unified foundations. There has long been interest in finding some sort of unified foundation for mathematics in the various senses in which that can be understood, but it is particularly important for neo-logicists precisely because they're trying to do more than merely unify mathematics; they're trying to ground *knowledge* of mathematics. Since that's the goal, we would do well to look at what taking the unified approach entails. By explicitly looking at what a unified abstractionist or neo-logicist methodology might look like, we can see where changes can be made or investigative gaps filled.

Even on a unified approach, it is still not clear exactly how much math-

[8] see footnote 6

[9] An abstraction principle is strongly stable if there is a cardinality at or above which that abstraction principle is satisfiable. Conservative stability puts a further restriction on how many abstracts of the relevant sort there must be. These are proposed model theoretic solutions to the Bad Company objection that guarantee joint satisfiability (among other things). See (Cook and Linnebo, 2018) for definitions and the state of the art. See also appendix A for a list of acceptability criteria, and chapter 6 for further discussion.

ematics we should expect, even from a *system* of abstraction principles. Attempts like those in (Boolos, 1987b) and (Shapiro and Weir, 1999) to try to capture some set theory using size-restricted abstraction principles to avoid the paradoxes, could hardly be seen as successful if we're looking for a unified neo-logicist foundation for mathematics, or even set theory.[10]

It's not my intention to go into any great detail about particular neo-logicist attempts to capture set theory, but rather to point out that despite limited success thus far, a significant amount of work aimed at finding mathematically strong neo-logicist theories has been aimed at capturing set theory. That lack of success shouldn't be seen as evidence of the non-viability of the larger neo-logicist project, however. Given the limited success of neo-logicist reductions based on extensions—the abstractionist analogue of sets governed by abstraction principles like BLV or New V—we should ask whether we can use abstraction principles and higher-order logic to either capture an alternative foundation such as category theory, or something altogether new that's capable of interpreting large swaths of mathematics.

If we take the first route and look for neo-logicist reductions of alternative foundational theories, the obvious options are category theory (see Landry and Marquis, 2005; Linnebo and Pettigrew, 2011) and homotopy type theory (HoTT, Univalent Foundations Program, 2013), though non-standard set theories such as those studied by Peter Aczel (Aczel, 1988; Aczel and Rathjen, 2010) may also be viable options. In any of these cases, an abstraction principle or collection of abstraction principles would need to be found that allows the derivation of the axioms of the theory in question in the presence of higher-order logic and appropriate definitions.[11] Additionally, said abstraction principle(s) would need to be acceptable in some sense that grounds its epistemic significance.

I am unaware of any work that has been done towards neo-logicist reduc-

[10] If we're less restrictive about what sorts of theories we are counting as neo-logicist, then some recent predicative approaches like those of Walsh (2016a); Studd (2015); Linnebo (2018) look to be quite successful *qua* mathematical strength, but we're then left abandoning some traditional neo-logicist principles. See §1.4.

[11] This is a fairly standard description of such projects, but it may be too strong, as we will see.

tions of non-standard set-theories or HoTT, but in the case of categories, some interesting work has recently been done by Shay Allen Logan (2015, 2017).[12] For that reason I will concentrate on the case of category theory, which should be taken as a case study rather than exhaustive of the possibilities.

Logan's work is a promising start to a neo-logicist reduction of category theory; he has found an abstraction principle satisfiable in the set theoretic meta-theory that appears to capture at least *the idea* of category theory in a mathematically useful way. He notes however, that

> What remains is to prove analogs to Frege's Theorem and Boolos's Theorem. That is, given some axiomatization A of the theory of categories, what we need is a mutual-interpretability result ... Unfortunately, there is no standard axiomatization of the theory of categories; so this is a rather difficult task. (Logan, 2017, p. 17)

Perhaps, though, the requirement that a particular set of axioms be reduced to logic and abstraction principles is too strong. Rather, it may be enough that a neo-logicist category theory agree with an understanding of category theory as it is used by mathematicians. Category theory, after all, is unhampered by its lack of a canonical axiom system. Additionally, a Lawverian understanding of categories as belonging to a category of categories, but in a neo-logicist setting (as in Logan, 2015), might go some way towards justifying an abstractionist theory of categories as a neo-logicist reduction in the required sense. There are other possible issues as well, but I will leave those aside for now.[13]

There are two important points to take away from this (admittedly brief) discussion. The first is that category theory appears to be a viable option for a neo-logicist foundation for higher mathematics. The second is the more

[12] Logan (2017, p. 7) notes that "... the neologicist project has to date made no inroads into category theory at all."

[13] For example, Logan requires at least third-order logic, and his category abstraction principle makes use of a size restriction that may not be well motivated from certain perspectives. Neither of these issues would be serious problems for most neo-logicists, however.

general observation that with some ingenuity, the neo-logicist might indeed escape the confines of set-centred thinking.

Even if Logan's or others' attempts at neo-logicist reductions of alternative foundational theories ultimately fail, we are still left with the other possibility I mentioned earlier. The suggestion is that we might use abstraction principles to find some mathematically powerful theory that's distinct from mathematical theories currently considered to be possible foundations for mathematics. This may sound far fetched, and perhaps it is, but I think we have good reason to think that such a discovery isn't impossible. Historically, the discovery and development of significant mathematical theories has often come though work on more modest problems or theories; mathematicians have rarely had grand foundational theories in their sights from the start. Cantor's set theory came out of his work on point-set topology, which itself came out of earlier work on real analysis and algebraic geometry. Marquis (2015, §2) has this to say about the discovery of category theory: "A desire to clarify and abstract their 1942 results led Eilenberg and Mac Lane to devise category theory." More precisely, the discovery of category theory was a result of generalizing work on algebraic topology, particularly work reported in (Eilenberg and Mac Lane, 1942).[14]

My point here is that Cantor didn't set out to find a theory that could interpret most of current (and future) mathematics, nor did Eilenberg and Mac Lane. In fact, in both of these cases, the resulting theories came about through generalizations on developments connected to much more specific mathematical problems.[15] The story of the development of mathematics through generalization is of course a familiar one. We often hear ancient Greek mathematicians praised for developing mathematics via abstraction, which in this sense I take to be a species of generalization.

[14]The first development of category theory can be found in (Eilenberg and Mac Lane, 1945). See (Landry and Marquis, 2005) and (Marquis, 2015) for more details from a philosophical perspective.

[15]This is not to say that no one aims to discover foundational theories. Lawvere (1963) looks have had foundational aspirations in his investigation of the category of categories, and the current debate about Woodin's V=Ultimate L and Martin's maximum concerns how best to characterize the entire set theoretic universe.

Lest we drift too far afield, we can now draw a parallel between the discoveries of set theory and category theory on the one hand, and neo-logicism, or abstractionism more generally, on the other. Frege used one abstraction principle, namely Basic Law V, to (try to) capture arithmetic, and arguably intended to do something similar in the case of real analysis.[16] This is similar to the piecemeal approach described above, though Frege's goals were in certain ways more modest than those of most modern neo-logicists. More specifically, the relevant generalization is in the expansion of the class of abstraction principles available to the neo-logicist. (The fact that Frege took Basic Law V to be a basic logical law is irrelevant.)

This comparison is even more compelling if we look at the unified approach. Instead of using single abstraction principles to capture particular mathematical theories, the unified approach may have us using collections of acceptable abstraction principles to capture large swaths of mathematics. We can even take this a step further. Following the lead of Fine (2002) and his General Theory of Abstraction, we can consider theories that take abstraction principles *in general* as their bases. Such an approach would only be possible if a class of acceptable abstraction principles could be delineated, and the requisite restriction motivated (i.e. a solution to Bad Company found), but such an approach is certainly possible and plausibly fruitful.[17] The thought would then be to take a general theory of abstraction principles and see what we can get, and perhaps also whether moving to yet higher-order settings could get us more interesting mathematics. Crucially I'm *not* proposing we take such an approach having *already* decided that we're trying to capture a bunch of set theory.

Analogies with the history of mathematics are not the only reason we might think that abstraction principles might provide us with a route to a theory that doesn't look very much like ZFC. Walsh and Ebels-Duggan

[16] Volume III of Frege's *Grundgestze der Arithmetik* (Frege, 1893, 1903) was going to cover real analysis (Simons, 1987).

[17] Sean Walsh (2016a) does something like this, though he uses a restricted comprehension scheme and additional choice principles, so it is not clear in what sense the theory he presents is neo-logicist in the strict sense I'm concerned with here (see §1.4). It is quite a mathematically strong theory, though.

(2015) show that, although ZFC is relatively categorical, theories consisting of second-order logic and abstraction principles are not in general. This suggests that theories based on abstraction principles are unlikely to capture ZFC, though more would obviously need to be said.

Perhaps more importantly, Boolos (1997) argues that since HP implies that every concept has a number, so should the concept of *being self-identical*, a universal concept. The fact that we can assign a cardinal number (called anti-zero) to a universal concept contradicts ZFC, which implies that there is no set of all sets (on pain of contradiction), so HP and ZFC can't be true together. In other words, FA is naturally interpreted as confirming the existence of anti-zero, which ZFC denies. Neo-logicists have various responses open to them (see Wright, 1999), however there is no agreed upon solution, and nor is a particular solution obviously entailed by neo-logicist principles. This leaves an unresolved tension between neo-logicism on the one hand, and set theory on the other, a tension that would need to be resolved if neo-logicists insist on trying to capture ZFC.

We might also argue for the possibility of independent neo-logicist foundations from an ontological perspective. Neo-logicists are already committed to mathematical platonism (see §1.2). Indeed, abstraction principles are meant to give us (partial) identity conditions or count as implicit definitions of abstract objects. Then, under the assumption that we are looking for *the* foundation for mathematics, the naturals, reals, sets and categories we get from our specific attempts at neo-logicist reductions are all subsystems or aspects of a more expansive system of (mathematical) abstracts.[18] This is not far removed from the development of the current contenders for mathematical foundations I've been discussing. There is a category of sets; categories can (for the most part) be described in the language of set theory; and HoTT gives us the naturals, the reals, and whole lot more. Furthermore, there is nothing about the neo-logicist brand of platonism that precludes abstracts being embedded in some larger structure. All that's required is that there be

[18]The assumption that we're looking for the foundation rather than *a* foundation is merely to make the following presentation more simple. The rejection of that assumption would not undermine the broader point.

abstract objects, and that those objects obey the identity criteria imposed by the abstraction principles they're associated with.

All of this might sound a bit outrageous, but my goal is merely to show that the neo-logicist project might end up somewhere unexpected, in a way that vindicates the program. There's no particular reason to think that such a place would look like what I've just described, either. That said, there might be some more concrete benefits to accepting the mathematical universe I've just sketched. For one, identity questions about abstracts associated with different abstraction principles would be in-principle answerable.[19] There would also be an obvious approach to solving the Good Company problem: more or fewer of the very same objects are being picked out by similar principles, though possibly in slightly different ways.[20]

It's worth reiterating at this point that all of this—any kind of alternative neo-logicist foundation of mathematics—will require answers to some tricky epistemological questions if there is to be any hope of success. Recall that the ultimate goal is not just formal or metaphysical foundations, but also epistemological foundations. However they see the ultimate neo-logicist theory, the neo-logicist must be able to explain how abstraction principles give us epistemic access to the truths of mathematics. A large part of that will come down to solving the Bad Company problem, but conceptually more important is establishing what exactly it is that makes abstraction principles epistemically special. In the early days of neo-logicism, Crispin Wright held that abstraction principles—in particular HP—are analytic (Wright, 1999). Analyticity in Wright's sense seems less likely an answer these days (see e.g. Boolos, 1997), but there are other contenders—other formulations of analyticity and conceptual truth, accounts of content carving (Linnebo, 2018, Part II), or of privileged definitions—that might fill the role.[21] But such question need not be answered for formal work to continue. The exception is the case where solutions to Bad Company rule out consistent, satisfiable abstraction principles that were of use to the formal program. None of that ought to

[19] Whether such questions would be answerable in practice is less clear.

[20] I pick some of this up again in §5.5.

[21] I take ou this question in the next chapter.

dampen the resolve of the determined neo-logicist, though.

I'll end by reiterating the point I most hope those working in and around the neo-logicist program take away from this chapter. Whatever the approach, we must keep in mind our ultimate goal, whatever that may be, and consider paths less well trodden, be they other foundational contenders, or new paths that must be cut through the mathematical jungle.

Chapter 4

Is Hume's Principle Analytic?

4.1 Introduction

Is Hume's Principle analytic?[1] Several authors have discussed this question according to the "classical account" of analyticity (Wright, 1999; Boolos, 1997). Yet, few seem to have devoted special attention to addressing whether or not the Principle can be considered analytic according to *Frege's* account of *analyticity*. Crispin Wright describes the classical account of *analyticity* as holding (minimally) that, "the analytical truths...[are] those which follow from logic and definitions" (Wright, 1999, p. 8). For Frege, a statement is analytic, roughly, if it is provable from *general logical laws* and *admissible definitions*. The latter charactarisation is *similar* to the former, but the ways in which it differs suggest that the latter cannot be adopted if Hume's Principle is to be deemed analytic. Below, we will explain why we take this to be the case. We begin with a brief elucidation of Hume's Principle (HP) and then sketch Frege's definition of analyticity. Next, we argue that, if HP is taken as a definition (as some neo-logicists take it to be), its admissibility depends on a proposition which is not, itself, analytic (according to Frege's conception). Specifically, we show that the proposition belongs to the "sphere of a special science" (namely, standard analysis) because it can be denied, without con-

[1] This is a lightly adapted version of a paper of the same name co-authored with Eamon Darnell, that is under review at the time of this writing.

tradiction, within another (non-standard analysis). We conclude on this basis that HP fails to satisfy Frege's definition of analyticity.

We then go on to argue that even if we follow the neo-logicist line taking HP as primitive—not dependent on an explicit definition of "cardinal number"—there is a very narrow set of conditions that would have to be met for HP to be considered analytic in Frege's sense. These conditions have to do with three of the most pressing concerns for contemporary neo-logicists: Bad Company, Good Company, and Caesar (see §1.3). We conclude with some ways our results might be expanded or generalised.

4.2 Hume's Principle

We will follow Wright (1999, p. 6) in formulating Hume's Principle as follows. For any (appropriate[2]) concepts F and G,

(HP) The Number of Fs is the same as the Number of Gs if and only if there is a one-to-one correspondence between the Fs and the Gs.

Frege provides an example which is helpful for getting a better grasp on how HP is to be understood. He writes,

> If a waiter wishes to be certain of laying exactly as many knives on a table as plates, he has no need to count either of them; all he has to do is to lay immediately to the right of every plate a knife, taking care that every knife on the table lies immediately to the right of a plate. Plates and knives are thus correlated one to one, and that by the identical spatial relationship. (Frege, 1980, §70)

The purpose of this example is to motivate HP by showing how one can determine that there are exactly as many Fs as Gs without counting each F and each G, or appealing to the concept of number. The determination can be made in virtue of a certain relation that each F bears to a certain G and *vice versa* (in this case for each knife k, k is *immediately to the right of* a plate p,

[2]See footnote 3 below.

and for each plate p, p is *immediately to the left of* a knife k). If such a relation obtains, then there is a one-to-one correspondence between the Fs and the Gs and so, the Number of Fs (whatever that is) and the Number of Gs (whatever that is) are the same Number (i.e. there are exactly as many of the one as there are of the other).

In light of Frege's example, we will interpret HP as follows. First, we will take there to be a one-to-one correspondence between the Fs and the Gs just in case there is a *bijection* between F and G (hereafter, we will use the expressions, 'F is *equinumerous* with G' and '$F \approx G$', as synonymous with, "there is a one-to-one correspondence between the Fs and the Gs"). Given these conventions, the formal version of HP with which we are operating is (the universal closure of) the following.

HP: $\#F = \#G \leftrightarrow F \approx G$

where '#' is a function from Fregean concepts to objects, and '≈' is a second-order formula asserting the bijectability of the Fs and Gs.

Second, as the waiter example suggests, we will take the Number of Fs to be the same as the Number of Gs iff there are *exactly as many* Fs as there are Gs. Accordingly, and as is implied by Wright in (Wright, 1999, p. 12), we will take the referent of 'the Number of Fs' to be that which (correctly) answers the question: How many Fs are there? (and likewise for G).[3]

What does 'less-than' Mean?

There is one other important feature of HP, and of Frege's account of cardinality more generally, that may seem obvious and natural now, but is in the background of much of what follows (and the foreground in §4.4): in asserting HP Frege is, like Cantor, asserting that one-to-one correspondence is the

[3]Wright introduces this question as part of his response to the objection that HP is not analytic on the grounds that not every concept has a number (e.g. *is self-identical*). In short, his point is that a restriction is needed such that substitutions for 'F' are restricted to those concepts such that the question, "How many Fs are there? makes sense—or at least has a determinate answer..." (Wright, 1999, p. 12), e.g. *count nouns* and expressions for *sortal concepts*. We don't necessarily endorse Wright's view on anti-zero, however—see chapter 3.

correct criterion for cardinal identity for both finite and infinite collections.[4] Another way of putting this is that if we say that there are fewer Fs than Gs, i.e. the number of Fs is less than ($<$) the number of Gs, an injective function from F into G could not be surjective—there would be Gs "left over". This is now the standard way to think about the less-than relation, at least among those of us familiar with 20th century mathematical logic. There is, however, another common intuition about the meaning of the less-than relation having to do with the part-whole relation.[5] This is perhaps best expressed using set/subset discourse but that does not mean that it is only applicable in formal set-theoretic settings. The principle is roughly this: if the Fs are a proper subset of the Gs, then the number of Fs is less than the number of Gs.

Take a bowl of fruit as a toy example. There are some mangoes and some figs in our fruit bowl. Without having to count either all of the pieces of fruit or the just the figs, we know that the number of figs is less than the number of pieces of fruit because the figs are a proper subset of the fruit. Likewise, if we were to eat all of the figs we would know that the number of mangoes is equal to the number of pieces of fruit because the mangoes are *not* a *proper* subset of the fruit. Notice in this case, as with the case of the waiter, above, nothing was ever *counted*, but answers to "number of" questions were compared with respect to the less-than and equal-to relations.[6]

We will have more to say about the two intuitions about the less-than relation after developing our central argument against the Fregean analyticity of HP, but two things are worth keeping in mind for what follows. First, that the two conceptions of the meaning of 'less-than' agree entirely for finite cases. Second, that the one-to-one correspondence (Fregean/Cantorian) understanding entails the subset/ part-whole understanding in the infinite case *but the converse does not hold* (Mancosu, 2015, p. 384).

[4]The ideas in the subsection owe much to the reading of Mancosu (2009, 2015, 2016). We encourage readers interested in what follows to look at those works.

[5]For the history of the use of these different conceptions among mathematicians dating back to the medieval period see (Mancosu, 2016, chapter 3).

[6]We realize that more work will have to be done to compare the cardinalities of disjoint collections, but as we will see, that is possible. See also the references in footnote 4.

4.3 Frege's Account of *Analyticity*

With the above interpretation of HP in place, we will now turn to Frege's account of *analyticity*. To begin, consider Frege's contrast between *analytic* and *synthetic* truths. He writes:

> The problem becomes, in fact, that of finding the proof of the proposition, and of following it up right back to the primitive truths. If, in carrying out this process, we come only on general logical laws and on definitions, then the truth is an analytic one, bearing in mind that we must take account also of all propositions upon which the admissibility of any of the definitions depends. If however, it is impossible to give the proof without making use of truths which are not of a general logical nature, but belong to the sphere of some special science, then the proposition is a synthetic one. (Frege, 1980, §3)

Here, Frege claims that a statement ϕ is *analytic* just in case there is a proof of ϕ and that proof relies only on general logical laws and admissible definitions. A definition is *admissible*, in this context, just in case the propositions upon which that definition depends are, themselves, analytic. *General logical laws*, as opposed to truths that belong to the sphere of some special science, apply to any subject matter whatsoever.

Matthias Schirn (2006, pp. 199–200) provides a useful (and we would argue correct) interpretation that brings the positive parts of Frege's definition of analyticity together nicely. It can be summarised as follows. For any statement Q, Q expresses an analytic truth just in case:

(1_F) Q expresses a primitive truth or,

(2_F) Q expresses a general logical law or,

(3_F) Q expresses an admissible definition or,

(4_F) There is a proof of Q such that that proof begins with primitive truths and each of its steps appeals (only) to general logical laws or (admissible) definitions.

But of particular relevance to the arguments in the following two sections is that, for Frege, the following principle holds.
If it is not possible to prove a statement ϕ, without making use of truths that belong to the sphere of some special science, then ϕ is not analytic.

In order to understand this claim, it is useful to consider Frege's explanation as to why the truths of geometry are *synthetic* and not *analytic*. He states,

> For purposes of conceptual thought we can always assume the contrary of some one or other of the geometrical axioms, without involving ourselves in any self-contradictions when we proceed to our deductions, despite the conflict between our assumptions and our intuition. The fact that this is possible shows that the axioms of geometry are independent of one another and of the primitive laws of logic, and consequently are synthetic. (Frege, 1980, §14)

It is worth specifically highlighting two of the points that Frege makes here. First, a (true) statement can fail to be analytic even if its denial, or the consequences of its denial, are not intuitable. Frege does not think that one can intuit any non-Euclidean space.[7] Yet, he does think that, for the purposes of conceptual thought, one can consistently assume that there are such spaces. Second, if a particular statement (say, the Parallel Postulate) can be denied within the sphere of a special science (like a non-Euclidean geometry) without contradiction, then that statement is *synthetic*.

Accordingly, we think that it is safe to say that Frege's definition of *analyticity* entails the following:

(A) If it is not the case that $\neg\phi$ entails a contradiction within the sphere of some special science (like Euclidean or non-Euclidean geometry), then it is not the case that ϕ is analytic.

[7] Frege makes this point explicitly stating, "To study such conceptions is not useless by any means; but it is to leave the ground of intuition entirely behind. If we do make use of intuition even here, as an aid, it is still the same old intuition of Euclidean space, the only one whose structures we can intuit." (Frege, 1980, §14)

4.4 Is HP *Analytic*?

Under the assumption that HP is a definition,[8] does HP satisfy Frege's account of *analyticity*? Before answering this question, it will be worth giving a brief account of what it takes for a statement to satisfy (3_F). Frege identifies two kinds of admissible definitions. The first consists in introducing an expression, not previously in use, to express the sense of a complex expression. Frege writes,

> We construct a sense out of its constituents and introduce an entirely new sign to express this sense. This may be called a 'constructive definition' ['*aufbauende Definition*'], but we prefer to call it a 'definition' *tout court*. (Frege, 1997b, p. 316)

For the sake of clarity, we will refer to any definition of this type as a "constructive definition" (despite Frege's preference). Giving a constructive definition involves introducing a new expression (sign) and stipulating that it expresses the same sense (*Sinn*) as the sense already known to be expressed by a particular complex expression.[9] Constructive definitions are supposed to merely serve the pragmatic role of abbreviating long or complex expressions.[10] The second kind of admissible definition that Frege identifies consists in providing a logical analysis of an expression that already has an established use. Frege states,

> We have a simple sign with a long established use. We believe that we can give a logical analysis of its sense, obtaining a complex expression which in our opinion has the same sense. We can only allow something as a constituent of a complex expression if it has a sense we recognize... That it agrees with the sense of the long established simple sign is not a matter for arbitrary stipulation, but

[8] Note that this is a neo-logicist idea, rather than Frege's own. Frege feels he needs to *derive* (each direction of) HP for Caesary reasons. See §4.6 below and §1.3 of the introduction.

[9] Frege gives a series of rules which govern this type of definition in (Frege, 1997a, §33).

[10] Frege does not think that such definitions are of logical importance but recognizes that they are of importance for thinking (Frege, 1997b, p. 315).

> can only be recognized by an immediate insight. (Frege, 1997b, p. 316)

We will call definitions of this sort, "definitions of analysis."[11] Ideally, definitions of analysis serve to clarify or sharpen one's understanding of the sense expressed by a simple expression with an established use (Frege, 1997b, p. 317). This is accomplished by showing that the sense of the simple expression is the same as the sense of a complex expression, each constituent of which expresses a sense that is recognized and clear.

In Frege (1997a), (Frege, 1903) and elsewhere, Frege lays out certain rules by which a *logical* constructive definition or definition of analysis must abide. He states,

> [T]he laws of logic presuppose concepts with sharp boundaries... Accordingly all conditional definitions, and any procedure of piecemeal definition, must be rejected. Every symbol must be completely defined at a stroke so that, as we say, it acquires a *Bedeutung*. (Frege, 1903, §65)

Both the laws of logic themselves plus the fact that logical laws apply universally mean that logical definitions must satisfy two requirements. First, any concept, function, name, etc. must be defined such that the exact extension of that concept, range of values for that function, referent of that name, etc. is determinate (no vagueness or ambiguity is permitted). Second, a definition must be determinate, in the way just mentioned, for everything in the domain of logic, not merely for a proper subset of that domain. We will take a definition to be *logical* just in case it is correct and satisfies these two requirements.

For Frege, a definition is admissible (i.e. *analytic*[12]) if it is either a logical constructive definition or a logical definition of analysis and, only if the

[11] Frege calls them, "analytic definitions", but suggests that it might be better to refrain from using the term 'definition' to describe this sort of logical analysis altogether (Frege, 1997b, p. 315).

[12] Henceforth by "analytic" we mean "analytic in Frege's sense" unless it is expressly noted.

propositions upon which it depends are analytic. HP depends, at least, upon the following: For any (appropriate) F,

(N) The Number of Fs is the extension of the concept *equinumerous with* F.

We will also express (N) as, $N(F) = Eq(F)$ (Where '$N(F)$' means "The Number of Fs" and '$Eq(F)$' means "is the extension of the concept *equinumerous with* F).

We will take the admissibility of HP, as a definition about when the Number of two concepts is the same, to depend upon (N) for two reasons. First, Frege understands 'the Number of Fs' in terms of (N) (Frege, 1980, §68). Second, HP holds if (N) does. If the admissibility of HP depends upon (N), then, according to Frege's account of *analyticity*, HP is analytic only if (N) is analytic.

It is well known that Frege took HP to depend on (N) in the sense that he finds it necessary to derive HP from Basic Law V (BLV—more about this in §4.5) and (N). The former is (meant to be) a *basic logical law* and (N), or its expression in the concept-script, to be an admissible definition. But we are in a different position than pre-1902 Frege, so it's worth investigating whether (N) *is* an admissible definition.

Is (N) *Analytic*?

It looks as though (N) fails to satisfy Frege's conditions for *analyticity*. If, within the sphere of some special science, not-(N) is true, then (N) fails to satisfy Frege's conditions for analyticity. Not-(N) is true within the sphere of some special science Σ, just in case there is at least one F such that $N(F) \neq Eq(F)$ is true in Σ. There is at least one F such that $N(F) \neq Eq(F)$ is true in Σ if, in Σ, one can, consistently, give a value for $N(F)$ (i.e. correctly answer the question: How many Fs are there?) such that that value $\neq Eq(F)$.

Within the sphere of non-standard analysis (NSA) one can, consistently, make $N(F) = num(F)$, where $num(F)$ = the *numerosity* of F, and show that

there is at least one F such that $num(F) \neq Eq(F)$.[13] To demonstrate this, we will compare the set of natural numbers including 0 ($\{0, 1, 2, 3, ...\}$), $\mathbb{N}_0$, and the set of natural numbers excluding 0 ($\{1, 2, 3, 4, ...\}$), $\mathbb{N}_1$,[14] and show that where $N(\mathbb{N}_0) = num(\mathbb{N}_0)$ and $N(\mathbb{N}_1) = num(\mathbb{N}_1)$, either $N(\mathbb{N}_0) \neq Eq(\mathbb{N}_0)$ or $N(\mathbb{N}_1) \neq Eq(\mathbb{N}_1)$.

The *numerosity* of a set F, is (roughly) the hypernatural number that answers the question: How many Fs are there? To define the *numerosity* of F, then, requires two things. First, a construction of the hypernatural numbers. Second, a means of mapping F to a particular hypernatural number according to the size of F (i.e. how many Fs there are). We will present each in turn.

Hypernatural Numbers[15]

The hypernatural numbers can be constructed by mapping $\mathbb{N}_0$ to its hypernatural extension: ${}^*\mathbb{N}_0$. This can be done by, first, defining a free or *non-principle* ultrafilter $\mathcal{U}$,[16] on $\mathbb{N}_0$. We will follow Wenmackers and Horsten (2013, p. 44) in defining $\mathcal{U}$ such that:

(U1) $\mathcal{U} \subset \mathcal{P}(\mathbb{N}_0)$

(U2) $\emptyset \notin \mathcal{U}$

(U3) $\forall F, G \in \mathcal{U}(F \cap G \in \mathcal{U})$

(U4) $\forall F \subset \mathbb{N}_0(F \notin \mathcal{U} \rightarrow \mathbb{N}_0 \backslash F \in \mathcal{U})$

(U5) $\forall F \subset \mathbb{N}_0(F \text{ is finite} \rightarrow \mathbb{N}_0 \backslash F \in \mathcal{U})$

(U1) makes $\mathcal{U}$ a proper subset of the power set of $\mathbb{N}_0$. (U2) states that the empty set is not an element of $\mathcal{U}$. (U3) holds that for any pair of sets in $\mathcal{U}$,

[13] We are assuming that the sphere of NSA constitutes the sphere of some special science. This is plausible as it is a coherent theory about a particular domain, whose reliance on sets precludes it from belonging to pure logic.

[14] Strictly, using sets here (as opposed to *concepts*) is a departure from Fregean terminology, however, doing so will make the presentation simpler.

[15] For those familiar with the construction of the hypernatural numbers, or who wish to take our word on the matter, can feel free to skip this section, and similarly for the following section where we define numerosities.

[16] If there is no finite set in an ultrafilter, it is *non-principle*.

the intersection of that pair of sets is also in $\mathcal{U}$. (U4) states that for any proper subset of $\mathbb{N}_0$, F, if F is not in $\mathcal{U}$, then the set of all elements in $\mathbb{N}_0$ that are not contained in F is a member of $\mathcal{U}$. Lastly, according to (U5), for any proper subset of $\mathbb{N}_0$, F, if F is finite, then the set of all elements in $\mathbb{N}_0$ that are not contained in F is a member of $\mathcal{U}$. Together, (U1)–(U5) make $\mathcal{U}$ the set of all infinite subsets of $\mathbb{N}_0$.[17]

Using $\mathcal{U}$, $\mathbb{N}_0$ can be mapped to ${}^*\mathbb{N}_0$ as follows. For all infinite sequences of natural numbers, $\langle s_n \rangle$ and $\langle r_n \rangle$:

(M1) $\langle s_n \rangle \approx_{\mathcal{U}} \langle r_n \rangle \leftrightarrow \{n \mid s_n = r_n\} \in \mathcal{U}$

(M2) $[\langle s_n \rangle]_{\mathcal{U}} = \{\langle r_n \rangle \mid \langle s_n \rangle \approx_{\mathcal{U}} \langle r_n \rangle\}$

(M3) $\forall n \in \mathbb{N}_0 : n = [\langle n, n, n, n, n, \ldots \rangle]_{\mathcal{U}}$

(M1) says, roughly, that a pair of infinite sequences of natural numbers are $\mathcal{U}$-equivalent just in case the set of numbers that label the places where the terms in each sequence are equal is in $\mathcal{U}$.[18] (M2) defines the $\mathcal{U}$-*equivalence class* of an infinite sequence of natural numbers $\langle s_n \rangle$ as the set of infinite sequences of natural numbers that are $\mathcal{U}$-equivalent with $\langle s_n \rangle$. (M3) states that for any natural number n, n is equal to the $\mathcal{U}$-equivalence class of infinite sequences of natural numbers that has, as a member, the constant sequence, $\langle n, n, n, n, n, \ldots \rangle$. The set of hypernatural numbers ${}^*\mathbb{N}_0$, is the set of $\mathcal{U}$-equivalence classes of members of the set of all infinite sequences of natural numbers. (M3) serves to embed $\mathbb{N}_0$ in ${}^*\mathbb{N}_0$ (Wenmackers and Horsten, 2013, pp. 44–45).

Numerosity

The hypernatural number that constitutes a measure of the size of a set F is the *numerosity* of F. Numerosity has been discussed by a number of authors[19]

[17] For a more general discussion of ultrafilters, see Komjáth and Totik (2008).

[18] To illustrate with a simple example, if $\langle s_n \rangle = \langle 0, 3, 4 \rangle$ and $\langle r_n \rangle = \langle 1, 3, 4 \rangle$, then $\langle s_n \rangle \approx_{\mathcal{U}} \langle r_n \rangle$ iff $\{2, 3\} \in \mathcal{U}$. Keep in mind this example is meant merely as an illustration. $\mathcal{U}$ does not contain any finite sets.

[19] See, especially, Benci and Di Nasso (2003) and Mancosu (2009).

but Wenmackers and Horsten (2013) provide a particularly clear definition of the notion. For this reason, we will closely follow their procedure below (with some minor variance from their original notation).

Wenmackers and Horsten define *numerosity* in three steps. First, they define a function C, that gives the *characteristic bit string* of a set of natural numbers (Wenmackers and Horsten, 2013, p. 47). The characteristic bit string of a subset F of $\mathbb{N}_0$ is constructed from the following function:

$$\begin{aligned} \chi_F \colon \mathbb{N}_0 &\to \{0,1\} \\ n &\mapsto \begin{cases} 0 \text{ if } n \in \mathbb{N}_0 \backslash F \\ 1 \text{ if } n \in F \end{cases} \end{aligned}$$

χ_F takes natural numbers as arguments and gives the value 0 if the given natural number is not in F and gives the value 1 if the number *is* in F. The function C is now defined as follows:

$$\begin{aligned} C : \mathcal{P}(\mathbb{N}_0) &\to \{0,1\}^{\mathbb{N}_0} \\ F &\mapsto \langle \chi_F(0), \chi_F(2), \chi_F(3), \ldots, \chi_F(n), \ldots \rangle \end{aligned}$$

C maps F to a sequence of 0s and 1s. In particular, the sequence of 0s and 1s that results from applying χ_F to each number in the linearly ordered sequence of natural numbers ($\langle 0, 1, 2, 3, 4, \ldots \rangle$). To illustrate, if F is $\{0, 2, 3\}$, $C(F)$ is $\langle 1, 0, 1, 1, 0, 0, \ldots \rangle$.

The second step in defining numerosity is to define *partial sums of characteristic bit strings* of F: $sum\text{-}C(F)$. Wenmackers and Horsten (2013, pp. 47–48) define this as follows,

$$\begin{aligned} sum\text{-}C \colon \mathcal{P}(\mathbb{N}_0) &\to \mathbb{N}_0^{\mathbb{N}_0} \\ F &\mapsto \langle S_n \rangle \end{aligned}$$

where,

$$S_n = \chi_F(0) + \ldots + \chi_F(n)$$

This function maps the sequence given by $C(F)$ to a new sequence where the value of the term at the n-th place in the new sequence consists of the sum of

all of the terms in places $\leq n$ in the sequence $C(F)$, for all places n. To illustrate, again suppose that F is $\{0, 2, 3\}$. Accordingly, $C(F) = \langle 1, 0, 1, 1, 0, 0, ... \rangle$ and so, $sum\text{-}C(F) = \langle 1, 1, 2, 3, 3, 3, 3, 3, ... \rangle$.

The final step in defining the *numerosity* of a set F, is to give a means of interpreting $sum\text{-}C(F)$ as one hypernatural number. This is done with the following function (Wenmackers and Horsten, 2013, p. 48):

$$num\colon \mathcal{P}(\mathbb{N}_0) \to {}^*\mathbb{N}_0$$
$$F \mapsto [sum\text{-}C(F)]_{\mathcal{U}}$$

The value of $sum\text{-}C(F)$ is an infinite sequence of natural numbers. The $\mathcal{U}$-equivalence class of an infinite sequence of natural numbers is a hypernatural number. Accordingly, the $\mathcal{U}$-equivalence class of the value of $sum\text{-}C(F)$ is a single hypernatural number.

The *numerosity* of a set F is the hypernatural number given by $num(F)$: the $\mathcal{U}$-equivalence class of the partial sums of characteristic bit strings of F (i.e. the value of $[sum\text{-}C(F)]_{\mathcal{U}}$). When the sizes of finite sets are given in terms of their *numerosities*, (N) holds (as does HP). To illustrate, assume that F is $\{0, 2, 3\}$. Now measure the size of F in terms of its numerosity. As before,

$$sum\text{-}C(F) = \langle 1, 1, 2, 3, 3, 3, 3, ... \rangle$$

and so,

$$num(F) = [\langle 1, 1, 2, 3, 3, 3, 3, ... \rangle]_{\mathcal{U}}$$

$[\langle 1, 1, 2, 3, 3, 3, 3, ... \rangle]_{\mathcal{U}}$ denotes the set of all sequences $\mathcal{U}$-equivalent with $\langle 1, 1, 2, 3, 3, 3, 3, ... \rangle$ and since $\{4, 5, 6, 7, ...\} \in \mathcal{U}$,

$$\langle 1, 1, 2, 3, 3, 3, 3, ... \rangle \approx_{\mathcal{U}} \langle 3, 3, 3, 3, 3, 3, ... \rangle$$

Hence, $[\langle 1, 1, 2, 3, 3, 3, 3, ... \rangle]_{\mathcal{U}}$ has $\langle 3, 3, 3, 3, 3, 3, ... \rangle$ as a member and so (by (M3)),

$$[\langle 1, 1, 2, 3, 3, 3, 3, ... \rangle]_{\mathcal{U}} = 3$$

The value of $Eq(F)$ is 3,[20] as is the value of $num(F)$ and so, (N) holds with respect to F. This result generalises to all finite sets of natural numbers.

[20] It follows from the manner in which Frege defines cardinal numbers (Frege, 1980, §§ 77–86) that, for any set of natural numbers F, the value of $Eq(F)$ is equal to the standard cardinality of F.

If num is applied to infinite sets, it gives a value in ${}^*\mathbb{N}_0 \backslash \mathbb{N}_0$. To see this, make $F = \mathbb{N}_0$. Thus, $C(F) = \langle 1, 1, 1, 1, ... \rangle$ and so, $num(F) = [\langle 1, 2, 3, 4, ... \rangle]_{\mathcal{U}}$. There is no place at which $\langle 1, 2, 3, 4, ... \rangle$ begins (infinitely) repeating some finite number n. Hence there is no n such that $\langle 1, 2, 3, 4, ... \rangle$ is in the $\mathcal{U}$-equivalence class containing $\langle n, n, n, n, ... \rangle$. Thus, $[\langle 1, 2, 3, 4, ... \rangle]_{\mathcal{U}}$ must be larger than any finite number and so, $[\langle 1, 2, 3, 4, ... \rangle]_{\mathcal{U}} \in {}^*\mathbb{N}_0 \backslash \mathbb{N}_0$. We will follow the lead of Wenmackers and Horsten (2013, p. 48) in following the lead of Benci and Di Nasso (2003, p. 52) and call this number, α (i.e. $num(\mathbb{N}_0) = \alpha$).[21]

(N) is False in NSA

When the *sizes* of $\mathbb{N}_0$ and $\mathbb{N}_1$ are given by their (respective) *numerosities*, (N) is false of either $\mathbb{N}_0$ or $\mathbb{N}_1$. By stipulation,

$$num(\mathbb{N}_0) = \alpha$$

Now consider $\mathbb{N}_1$. $C(\mathbb{N}_1) = \langle 0, 1, 1, 1, 1, 1, ... \rangle$ and so, $sum\text{-}C(\mathbb{N}_1) = \langle 0, 1, 2, 3, 4, 5, ... \rangle$. Subtracting one hypernatural number from another is done by taking sequences from the relevant $\mathcal{U}$-equivalence classes and then subtracting (in the standard way) their corresponding entries one by one.[22] Accordingly,

$$\begin{aligned}
\alpha - 1 &= [\langle 1, 2, 3, 4, ... \rangle]_{\mathcal{U}} - [\langle 1, 1, 1, 1, ... \rangle]_{\mathcal{U}} \\
&= [\langle 1, 2, 3, 4, ... \rangle - \langle 1, 1, 1, 1, ... \rangle]_{\mathcal{U}} \\
&= [\langle (1-1), (2-1), (3-1), (4-1), ... \rangle]_{\mathcal{U}} \\
&= [\langle 0, 1, 2, 3, ... \rangle]_{\mathcal{U}}
\end{aligned}$$

Hence, the $\mathcal{U}$-equivalence class of $\langle 0, 1, 2, 3, ... \rangle$ is the hypernatural number $\alpha - 1$. Since $num(\mathbb{N}_1)$ is the $\mathcal{U}$-equivalence class of $\langle 0, 1, 2, 3, ... \rangle$,

$$num(\mathbb{N}_1) = (\alpha - 1)$$

[21] Note: Wenmackers and Horsten do not consider the set of natural numbers *including* 0 and so, strictly, do not call $num(\mathbb{N}_0) = \alpha$. Rather they stipulate that $num(\mathbb{N}_1) = \alpha$. Following their stipulation, $num(\mathbb{N}_0)$ should be $\alpha + 1$. We've chosen to overlook this detail to keep things simple.

[22] See Wenmackers and Horsten (2013, p. 50) for a brief explanation of how addition on ${}^*\mathbb{N}_0$ is defined.

Since $(\alpha - 1) < \alpha$,

$$num(\mathbb{N}_1) < num(\mathbb{N}_0)$$

Therefore, when the sizes of $\mathbb{N}_0$ and $\mathbb{N}_1$ are compared in terms of their respective *numerosities*, $N(\mathbb{N}_1) < N(\mathbb{N}_0)$. With respect to the infinite number $\aleph_0$,[23] Frege states that it applies to the concept F (i.e. $Eq(F) = \aleph_0$) just in case, "there exists a relation which correlates one to one the objects falling under the concept F with the finite numbers" (Frege, 1980, § 84). Since the objects falling under $\mathbb{N}_0$ *are* the finite numbers,

$$Eq(\mathbb{N}_0) = \aleph_0$$

Let f be the function from $\mathbb{N}_1$ to $\mathbb{N}_0$:

$$\begin{aligned} f : \mathbb{N}_1 &\to \mathbb{N}_0 \\ n &\mapsto (n - 1) \end{aligned}$$

Accordingly, f correlates one-to-one the objects falling under $\mathbb{N}_1$ with the finite numbers. Thus,

$$Eq(\mathbb{N}_1) = \aleph_0$$

Hence, $Eq(\mathbb{N}_1) = Eq(\mathbb{N}_0)$ and so, either $N(\mathbb{N}_0) \neq Eq(\mathbb{N}_0)$ or $N(\mathbb{N}_1) \neq Eq(\mathbb{N}_1)$. In either case, it follows that there is at least one F such that $N(F) \neq Eq(F)$ is true.

HP is not Analytic in Frege's Sense

It follows from the above that (N) is not analytic. There is a sphere of some special science (NSA) such that not-(N) does not lead to contradiction. It is consistent within NSA to make $N(F) = num(F)$. Furthermore, it seems that $num(F)$ does correctly answer the question "How many Fs are there?" As was shown above, with respect to any finite F, $num(F) = Eq(F)$. Hence, with respect to any finite F, if $Eq(F)$ correctly indicates how many Fs there are, so must $num(F)$. Other than presupposing the analyticity of (N), we see little reason to suppose that $num(F)$ should fail to also correctly indicate how many

[23] Frege uses ∞_1 rather than $\aleph_0$.

Fs there are with respect to any infinite F. Hence, by (A), (N) is not analytic. Under Frege's account of *analyticity*, HP is analytic only if (N) is analytic. Therefore, HP is not analytic according to Frege's account of *analyticity*.

As mentioned above, and argued in much greater detail by Mancosu (2016), the question of whether numerosities or Fregean cardinalities should be used to measure infinite cardinalities comes down to whether we wish to privilege the intuition that subsets are always strictly smaller that their (proper) supersets (if $F \subset G$ then $F < G$), or the intuition that cardinality is completely captured by bijectability/ one-to-one correspondence as with HP. Both understandings coincide for finite numbers, but diverge in the case of infinite cardinals. Thus it appears that we're in an even better position than Frege was in the case of geometry, because although Frege found non-Euclidian spaces to be unintuitable, the intuition that proper parts are strictly smaller than their wholes is a common one. Indeed, Mancosu (2016), traces a venerable history of mathematicians relying on that intuition both before and after Frege and Cantor 'decided' on one-to-one correspondence.

4.5 An Easier Route

Beginning again with the assumption that the analyticity of HP relies on (N) being an admissible definition, and is thus itself analytic, we can take a much shorter route to the conclusion that HP isn't analytic in Frege's sense. The principle (N) is straightforwardly *in*admissible. Here's why. It says that numbers are a particular class of *extensions*, which are (according to Frege) *logical objects* governed by BLV (1903), which says that two concepts have the same extension just in case exactly the same objects fall under both concepts, i.e. the two concepts are coextensional.[24] But in his (in)famous letter to Frege in 1902 (see van Heijenoort, 1967, pp. 124–126), Russell shows that BLV is inconsistent. Despite its suggestive moniker then, BLV is *not* a basic logical law. Even if we were to find a way to characterise extensions with consistent,

[24]See §1.2 or appendix A for a formal definition.

analytic axioms, Frege's derivation of HP from (N) relies heavily on BLV.[25] So, HP, if it relies on (N), is not analytic (in Frege's sense).

Our appeal to numerosities is not, however, a superfluous exercise of our mathematical muscles. Recall from §1.2 that Scottish neo-logicists want to take HP as primitive and then argue that it is analytic, or has some equally important epistemic status that will allow us to ground our epistemology of arithmetic.[26] If we are concerned with the analyticity of HP from a neo-logicist perspective, then we need not concern ourselves with (N). In such a case, HP would have to either qualify as a basic logical law, or as an admissible definition.

In the first case, one would be hard pressed to find anyone willing to endorse the claim that HP is a basic logical law, so we won't go into any great detail here. It is first worth noting though, that if one *were* to claim that HP is a basic logical law, and assuming that basic logical laws are analytic (which is the point), the neo-logicist reduction of arithmetic to logic falls out immediately.[27] But HP almost certainly isn't a basic logical law. The obvious arguments against HP as a basic law are ontological in character. For one, if we accept HP as a basic logical law, then we are already committed to there being infinitely many objects. This is a much larger ontological commitment then first-order logic (1 thing) or second-order logic (1 thing).[28]

There's more to be said here, but to keep laying into a straw man seems unfair. So we're now left with the possibility that HP is an admissible definition. It is here that our earlier development of numerosities will come in

[25]This is not to say that it would be impossible to find a consistent theory of extensions, the objects of which could be used in the formulation of (N), and HP derived therefrom. However, our current best theory of extensional entities is Zermelo-Fraenkel set theory. If we then take extensions to be governed by such a system, we would have to show that the axioms of ZF are analytic. And if we can do *that* we can declare victory for logicism without having to worry about (N) or HP, other than to perhaps pick out which sets to call the natural numbers.

[26]It strikes us that Fregean analyticity as we have represented it here differs enough from the standard Kantian or neo-Quinean accounts of analyticity that it may provide such a status even if we may not consider it to be a species of analyticity proper. Discussion of this possibility would take us too far afield, however.

[27]That is it falls out immediately from a proof of Frege's theorem, which though non-trivial, is by now well known (Boolos, 1996; Heck, 2011b, see also §§1.2 & 2.5).

[28]Quine (1970) was wrong about the vast ontological commitments of second-order logic. See Boolos (1975) for the canonical refutation of Quine on this count.

handy (again).

HP isn't an Admissible Definition

First, as we know, both HP and BLV are abstraction principles (APs), as are uncountably many other sentences, including a numerosity AP (Mancosu, 2016, §9, of which more below). As we have seen, we can't simply argue that APs are *all* analytic or otherwise epistemically privileged, because, for one thing, BLV is inconsistent, while HP is taken to be the paradigm case of a 'good' AP.[29] This is the problem of Bad Company.[30]

What all of this has to do with NSA and analyticity is the following. Establishing HP as an admissible definition requires a solution to Bad Company unless we have some reason to think that HP is privileged even among APs.[31] The reason we would need a solution to Bad Company is that we would presumably need to give strict grounds for thinking that HP successfully cashes out phrases like 'the number of Fs is the same as the number of Gs' while at the same time denying that BLV (for example) *doesn't* successfully cash out phrases like "the Fs and the Gs are coextensional". This isn't the usual way of framing Bad Company, but it is effectively the same problem.

Since we are dealing with a single sentence, what we would need to do to show that HP isn't an admissible definition and thus not analytic, is to show that there is some special science where HP fails, but is itself coherent. NSA looks like a good candidate. Indeed, Heck thinks that the case of numerosities closes the door on HP *qua* conceptual truth. In discussing Mancosu (2009) Heck writes the following.

> Mancosu's announced goal in his paper is "to establish the simple point that comparing sizes of infinite sets of natural numbers is a legitimate conceptual possiblility" (Mancosu, 2009, p. 642). I think

[29] It's straightforward to construct a model of HP (see e.g. Boolos, 1998, Chapter 9). Additionally, HP plus full axiomatic second-order logic, known as Frege arithmetic (FA) is equiconsistent with PA^2.

[30] For more detail see §1.3 and chapter 6.

[31] See chapter 3 for a brief discussion of reasons we might think HP is privileged. Accepting that HP is unique would likely spell the end of neo-logicism beyond arithmetic, however.

> it is clear that he succeeds. But if it is conceptually possible that infinite cardinals do not obey HP, then it is conceptually possible that HP is false, which means HP is not a conceptual truth, so HP is not implicit in ordinary mathematical thought. (Heck, 2011b, pp. 265–6)

This is in fact, very broadly, the argument we're in the midst of giving and it turns out that there are a few more holes to plug.

4.6 A Final Worry (or Three)

The 800 pound pink gorilla in the foyer happens to be called Caesar, and is concerned with the question of whether the Numbers of HP and the numerosities of NSA are commensurable in the first place. It might be the case that we're equivocating when we say that we can use both Numbers and numerosities to answer the question "how many are there?" To put it another way, invoking NSA as a 'special science' in which HP fails, is to say that we are talking about the *very same* cardinal numbers in both cases. Although there are good reasons to think that they *are* the same cardinal numbers, it's a metaphysical assumption that can be consistently denied.

Identifying Cardinals

In §56 of the *Grundlagen* (Frege, 1980) Frege famously laments that "... we can – to give a crude example – never decide by means of our definition whether Julius Caesar belongs to a number concept, whether this same well-known conqueror of Gaul is a number or not." This passage has since given a name to the so-called Julius Caesar objection, or Caesar problem. The core of the issue is that HP gives us no way to determine whether an object not identified by an expression of the form $\#\Phi$ is a number or not. Frege gets around this problem by introducing an explicit definition of 'the number of': (N). As we've already shown that that strategy fails, we have to look elsewhere if we want to figure out whether numerosities and numbers can be identified.

Since APs are (partial) identity criteria (Fine, 2002, ch. 1), an obvious place to start would be to see whether there is a suitable equivalence relation that will allow us to construct an AP for numerosities. Numbers and numerosities would then be on equal conceptual footing, and we could appeal to the literature on the identification of abstracts.[32] Alas this strategy is unlikely to bear fruit. Mancosu (2016, §9) points out that any AP that fully satisfies the Part-Whole principle will be (massively) inflationary.[33] Because of Cantor's theorem and related arguments, we know that inflationary APs (like BLV) are unsatisfiable in classical, static settings. So we're left with less direct arguments.

To our minds the most compelling evidence that Numbers and numerosities ought to be identified is that they agree for all finite cases. There are practical as well as theoretical reasons to identify the finite cardinals, the finite ordinals, the real whole numbers, etc. This is a thorny issue for structural realists, as well as others who take a piecemeal approach to the foundations of mathematics—is the natural number structure embedded in the real number structure, the real number structure in the complex number structure, or $2_{\mathbb{N}} \neq 2_{\mathbb{R}} \neq 2_{\mathbb{C}}$?[34] Denying the identity of numbers presented in different ways would wreak havoc on ordinary mathematics, and there don't look to be reasons to uphold such distinctness claims (beyond perhaps some currently unpopular metaphysical theses). It would be much easier to say that the various properties of numbers converge on the naturals, or slowly diverge as they become more complex.

That leaves the possibility, though, that numerosities and Numbers are much like classical cardinals and ordinals, agreeing in finite cases, but diverging for infinite cases. No-one to our knowledge holds that we can't have both infinite cardinals and infinite ordinals, and still maintains the identity of the

[32] See (Mancosu, 2015, §9) for an overview as well as a discussion of some issues related to NSA and Caesar. See §5.5 for a sufficient condition for abstract identity based on the concepts developed in that chapter.

[33] An AP, Γ, is inflationary if it entails there be more Γ-abstracts than there were objects in the original domain. BLV is inflationary; HP is inflationary on finite but not infinite domains. See chapter 6 for more discussion.

[34] See (Cook and Ebert, 2005), who call this the 'C-R problem, for more discussion in the context of neo-logicism.

finite cardinals and ordinals.

Where this breaks down is with the less-than relation. To hold that numerosities and Numbers are fundamentally different we would have to give up the motivation for considering numerosities in the first place, namely that the less-than relation should (or at least could) be defined according to the part-whole principle *rather than* bijectibility. So once again it looks like we have a dilemma. We can insist on a univocal less-than relation, or give up on the identity of finite "natural numbers" that have been defined in different ways.

If we take hold of the first horn—that there is a univocal less-than relation—we should conclude that HP isn't an admissible definition, and thus not analytic because it is inconsistent within NSA. If we grasp the second horn, the possibility that HP is an admissible definition is still open, in which case more work needs to be done.

Companions

Given our arguments thus far, an obvious strategy presents itself: find another "special science" where HP fails. While we admit that such a strategy may eventually be successful, a problem is immediately apparent. Since we will have already given up on identifying Fregean numbers and numerosities, we would hard pressed to find a domain that meets the requisite criteria, but won't allow us to make a similar move. We could just keep claiming that the objects of the domain under consideration and Fregean Numbers are incomensurable.

Two other issues also arise if we are trying to establish HP as an admissible definition. The first, already briefly introduced, is Bad Company. The second has been recently dubbed Good Company by Mancosu (2015). Both of these problems are closely related to the criterion of universal applicability that is behind the search for special sciences in which HP fails. Bad Company asks us to weed out APs that will lead to inconsistency, while Good Company asks us to choose between principles that will do the same work as one another. Indeed, the motivation for worrying about Good Company is essentially the

same as our concern about the comensurability of numerosities and Fregean numbers. If we have multiple ways to construct or define a concept like 'cardinal number', in this case different APs, how ought we decide between them?[35]

Bad Company presents a slightly (but only slightly) different problem. If we have a principle that we're claiming to be universally applicable, analytic, or an admissible definition, why should we think that it is fundamentally different than other, inconsistent principles of the same form?

These are serious issues for neo-logicists, and there is a great deal of literature proposing and rejecting possible solutions to Bad Company (see Linnebo, 2011; Cook, 2012; Cook and Linnebo, 2018, and chapter 6), much of which will be applicable to Good Company. For our purposes however, the issues are somewhat narrower in scope.

Since we are only concerned with the analyticity of *HP*, we need not worry about delineating the class (or classes) of acceptable APs. Instead we can look at justifying an assumption underlying much of the literature on Bad Company that is explicitly challenged by Good Company: HP is special. We would need to show that there is something conceptually, and/or logically different about HP that puts it above other similar principles, and also conceptually above other understandings of cardinality such as that provided by NSA. In particular we would need to show that it is broadly applicable in a way that other options are not. We have already argued in §4.2 that our 'natural' understanding of 'number of' won't be enough. That was the point of looking at NSA.

The other option requires solutions to Good Company and Bad Company, issues we won't take a stand on here other than to note that they are both open questions, and the state of the debate on Bad Company suggests that a new route needs to be taken if we are to hope for a solution. (See chapter 6 for one suggestion for a new route.)

All of this is to say that even if we assume the incomensurability of Fregean numbers and numerosities, the best hope for establishing the analyticity of

[35] §5.5 of the next chapter suggests a way out of Good Company.

HP is to find a very specific kind of solution to problems that have proven extremely contentious and difficult.

4.7 Concluding remarks

There are some important take-ways from our analysis here. First and foremost, it is exceedingly unlikely that HP is analytic on Frege's understanding of analyticity. This in itself is interesting for a couple of reasons. First, it puts an important bound on how much of *Frege's* logicist project can be reconstructed without BLV, at least with respect to Frege's chosen method for showing the purported analyticity of arithmetic. In another way though, that HP isn't analytic vindicates Frege's desire to ground HP on more fundamental principles and definitions. It furthermore highlights a more insidious aspect of the Caesar problem which arguably was Frege's impetus for that decision: there are apparent abstracta that are much more difficult to differentiate from HP's numbers than the "well-known Conqueror of Gaul".[36]

More generally we have highlighted just how closely entwined the Caesar problem and the Good and Bad Company problems are. This may turn out to shed light on the importance of resolving all of these issues if the epistemic supremacy of HP is to be established as is required by Scottish neo-logicism.

Finally, we contend that the analysis we have herein provided will be useful in showing that HP isn't analytic in senses other than Frege's, but that's a project for another day.

[36] Again see (Mancosu, 2015, §7), where he also uses the word 'insidious'. We liked it enough to use it and add this footnote. This problem is picked up again in §5.5.

Chapter 5

Canonical Equivalence Relations

5.1 Introduction

Generally when we introduce abstraction principles (APs), we do so by introducing them as variants of one of the following three forms (suppressing quantifiers).

$$\S F = \S G \leftrightarrow F \sim G;$$

$$\partial_E F = \partial_E G \leftrightarrow E(F, G);$$

or

$$Ax.Fx = Ax.Gx \leftrightarrow E_x(F, G)$$

For consistency with other chapters I will adopt the first convention, but all of these are perfectly good representations of APs in their most general form. I will argue however, that when we represent or conceptualize APs that we're working with in a coarse-grained a manner as this, we are ignoring information about the equivalence relations that is vital for answering questions about identity and acceptability. In particular, I will be advocating for what I call a *fine-grained approach* to the analysis of abstraction principles.

The general strategy, which I will present in detail in §5.2, is to associate abstracts with their *canonical equivalence relations* which need not be represented by the entire formula on the right hand side of an AP. This will allow us to give a sufficient condition for the identity of abstracts produced by different APs—abstracts ought to be identified if they are associated with the same equivalence classes of the same canonical equivalence relations—discussed in more detail in §5.5.

This approach will also provide a strategy for ruling out what have been called *fishy* APs (Cook and Linnebo, 2018, and Chapter 6), as well as some abstraction principles that are at the core of the Good Company objection in §5.5 (Mancosu, 2015, 2016). In other words, if the strategy I develop below is adopted, it will bring us closer to solutions to both the Good Company and Bad Company objections to neo-logicism.[1]

5.2 The Approach

I will start, as is now customary, with the two most well know APs. The first is Frege's *Basic Law V*:

(BLV) $$\varepsilon F = \varepsilon G \leftrightarrow F \equiv G$$

The second is *Hume's Principle*:

(HP) $$\#F = \#G \leftrightarrow F \approx G$$

Basic Law V, though inconsistent, wears it's heart on its sleeve, which is to say the equivalence relation at play is just coextensionality. The usual presentation of HP, on the other hand, uses '$\approx$' to abbreviate a second order sentence asserting the existence of a bijection. However, although that abbreviation obscures what's going on on the RHS of HP, it won't cause us any trouble (so long as we remember what it's abbreviating). This is in part because the abstracts HP produces are equinumerosity-abstracts (although we

[1] In chapter 6 I use the notion of a canonical equivalence relation to first provide a more precise definition of fishiness, and use that to show that likely solutions to Bad Company will require more attention to the connection between equivalence relations and abstracts; in particular an explanatory connection.

usually call them 'numbers')—it only produces abstracts that are associated with the whole formula on the right hand side of the AP.[2] More precisely, the abbreviation in HP doesn't obscure any information that will be relevant in determining its canonical equivalence relation, as we will see.

Let's now look at another well-known AP known as *New V*:

(NewV) $$\varepsilon F = \varepsilon G \leftrightarrow (F \equiv G \vee (F \approx V \wedge G \approx V))$$

where 'V' is the universal concept ($V = [x : x = x]$). The reason I've used the same abstraction operator as I did for BLV is that New V behaves exactly the same as BLV for concepts that are less than universe-sized and sends the rest to a dummy abstract, rendering it satisfiable. (I'll save further justification of the identification for §5.5.) We could also represent New V like this (for example):

(NewV) $$\varepsilon F = \varepsilon G \leftrightarrow F \sim_N G$$

Here we've obscured information that was more readily available in the first version, namely the size restriction on concepts to which coextensionality is being applied.[3] There's nothing particularly *wrong* about presenting New V in the second way, but it obscures the fact that the equivalence classes associated with the abstracts produced by New V are coextensionality equivalence classes.[4] Before I make this more precise, here two more abstraction principles that will be helpful in what follows.[5]

(BCA) $$\%F = \%G \leftrightarrow (F \approx G \wedge \neg F \approx \neg G)$$

If we call the associated abstracts *conumbers* then BCA says that the conumber of F is the same as the conumber of G iff the Fs and Gs are equinumerous *and* their complements are equinumerous.

[2] If you're thinking that this is always the case, bear with me for a moment.

[3] Note that we've not restricted the applicability of the principle as a whole.

[4] In fact New V also produces what we generally call a dummy abstract. I'll say more about this, especially in 5.3.

[5] See Walsh and Ebels-Duggan (2015, §5), Mancosu (2016, chapter 4), and Cook (2017) for discussion of BCA and similar APs.

Finally, an abstraction principle I'll call Hume's V:

$$\begin{aligned} \Diamond F = \Diamond G \leftrightarrow [&(F \approx G \wedge (F < \omega \vee G < \omega)) \\ &\vee (F \equiv G \wedge (\omega < F < |V| \wedge \omega < G < |V|)) \\ &\vee (F \approx V \wedge G \approx V)] \end{aligned} \tag{HV}$$

This behaves like HP for finite concepts, and like New V for infinite concepts (take '$X < \omega$' to be a second-order sentence asserting that X is Dedekind finite). What's interesting about HV is that it produces both numbers and extensions.[6]

We can then classify these APs into four groups. There are those like BLV and HP that divide every available concept into equivalence classes based on one equivalence relation without restriction—call these (Case-1) abstraction principles. There are those like BCA that combine equivalence relations, but without restricting which concepts are available—(Case-2). There are those like New V that do the same, but restrict the available concepts to those of a certain size (Case-3). Finally, there are those that combine both of the latter (Case-4). Each of these kinds of APs will pose different problems and thus have to be dealt with a bit differently in what follows. As we shall see however, it is (Case-3) that will cause most of our trouble. As I have already said, (Case-1) is unproblematic, and (Case-4) ought to be easy to tidy up if we have dealt with the others (see §5.4). (Case-2) will also be easily taken care of, as we will see in §5.2.

Canonical Equivalence Relations

The general strategy for the approach I'm advocating is to identify the *canonical equivalence relation* associated with each unique sort of abstract, and then to use that categorization to motivate solutions to tricky problems having to do with analyses of APs. A rough gloss on the idea of canonical equivalence relations is the following. APs provide identity criteria for abstracts by associating abstracts with equivalence classes of concepts. So, the 'shortest'

[6] Like New V, HV also produces a dummy abstract. I don't find dummy abstracts to be problematic, but I'll suggest some ways to think about them in §5.3.

AP that would provide an identity criterion for a given sort of abstract has *only* the canonical equivalence relation for that sort of abstract on its right hand side. For example, the canonical equivalence relation for both BLV and New V is coextensionality, and the canonical equivalence relations for HV are equinumerosity *and* coextensionality, since HV ostensibly picks out two distinct kinds of abstracts. Note that a single AP can be associated with more than one canonical equivalence relation. This makes sense since we expect each unique sort of abstract to be canonically associated with exactly one equivalence relation, so an abstraction principle like HV that ostensibly picks out two kinds of abstracts ought to have two canonical equivalence relations associated with it.[7]

As it stands, this gloss won't get us very far without a lot of unpacking, and care needs to be taken to ensure that we don't end up begging the question. As we will see however, once the concept is made precise, an ability to identify canonical equivalence relations will allow us to analyze and categorize APs in interesting and fruitful ways. But first a couple of caveats. A certain amount of bootstrapping will be necessary to get this analysis off the ground, but I will aim to assuage concerns about circularity when they might arise. Second, I will not provide a mechanical or purely formal 'test'—there will be a need for some interpretation. This is a departure from much of the recent literature on APs which is often very technical, but is consistent with e.g. Wright's general approach around the turn of the (21st) century.[8] Such an approach is in the interest of the neo-logicist project. Allowing ourselves the use of epistemically loaded concepts like understanding and explanation in our analyses of APs will not only help solve problems like Bad Company, but help ensure that the APs we deem acceptable are capable of carrying the epistemic weight required of them.

Perhaps the largest problem we will face in trying to clarify the notion of a 'canonical equivalence relation' can already be seen by the presentation in

[7] We'll see later that it's actually a bit more complicated that this, but this is a good starting point.

[8] See, for example, the first appendix of Wright (1999). See also the quotation at the beginning of the introduction.

the introduction. By definition the right hand side of any AP must consist solely of an equivalence relation, which is why it's acceptable to present the right hand side of New V as either $F \equiv G \vee (F \approx V \wedge G \approx V)$ or $F \sim_N G$.[9] This is to say, adding the size restriction to BLV (giving us New V) results in a new equivalence relation, though one very similar to BLV. The problem then is picking out the relevant (i.e. *canonical*) 'internal' equivalence relation.

I will present one way of doing so (I expect that there are others), but we should first note that such identifications are implicit in much of the literature. We almost always call the abstracts associated with both New V and (those that would be associated with) BLV 'extensions'.[10] And lest we think that this move is licensed by the fact that there aren't actually any abstracts associated with BLV because it is inconsistent, notice that in settings where BLV is rendered consistent by restricting the second-order comprehension scheme, the resulting abstracts are still called extensions (see e.g. Walsh, 2016b). Additionally, the abstracts associated with Cook's Finite Hume's Principle (FHP, the restriction of HP to finite concepts) are referred to as 'numbers', and in both of these cases the same symbols are used for the (respective) abstraction operators (Cook, 2012).[11] This latter is important because coding the identification of abstracts in the syntax suggests that, rather than being a convenient *façon de parler*, the move is made intentionally, or at least the identification of the resulting abstracts is acceptable to those proposing restricted abstraction principles.

Of course you might think that since FHP is meant to *replace* HP (likewise for New V and BLV), plus the related fact that it would be pointless to include both the restricted and unrestricted version of a principle in the same system, FHP and New V are *refinements, corrections*, or just *new definitions*. Indeed, that's an extremely plausible analysis (though as we shall see in §5.5 the situation is complicated by the Good Company problem), but the use of the same terms and symbols for abstracts associated with the same (canonical)

[9] Jonathan Payne (2013) argues that, if neo-logicists adopt a negative free logic, partial equivalence relations suffice.

[10] The notable exception here is George Boolos (1987b) who introduced New V in the hopes of capturing some set theory, but dubbed the New V-abstracts 'subtensions'.

[11] See a appendix A for a list of the formal definitions of APs.

equivalence relations suggests at least an implicit identification. In any case, it should be obvious why we might identify the abstracts associated with, say, HP and FHP, and why it wouldn't be unreasonable to do so. For example, if we're trying to capture arithmetic we should be talking about the same objects—numbers—whether we adopt HP, FHP or something else.[12]

Given all of this talk of identifying abstracts, one route to picking out canonical equivalence relations presents itself. If you were antecedently committed to identifying the abstracts associated with certain (distinct) APs, then you could just go look for whatever the APs associated with the identical abstracts had in common (ensuring that what was picked out was an equivalence relation). You might be in such a position, for example, if you were already committed to particular solutions to Good Company, or the Caesar problem. It would ideally also be the case that the same equivalence relations would need to come out as canonical on that account as will on mine, but that seems quite likely: it would be a strange theory indeed that allowed abstracts from different APs to be identified, that also ruled the abstracts from HP and FHP distinct.[13]

I don't find the possibility of some neo-logicist being in the position just described implausible, but *I* am not in such a position.[14] Furthermore, if (something like) the notion of 'canonical equivalence relation' I develop herein is adopted, a sufficient condition of inter-abstract identity will be a consequence, and a route to a solution to Good Company will present itself.

Before moving on, it is worth noting again that (Case-1) is unproblematic. At least for present purposes we should take it that there are APs whose right hand sides are just their canonical equivalence relations. We can also assume that we can often pick out such cases by inspection. That said, it's likely that there will be less tractable cases, but the methods for identifying canonical equivalence relations developed below are easily adaptable to instances of

[12]Mancosu (2016, chapter 4) identifies other APs that could potentially play the role of HP in a neo-logicist reduction of arithmetic.

[13]That is, as a matter of metaphysics; if we're taking APs as a (partial) guide to the metaphysics of abstract objects, as we should if we're good neo-logicists, we should expect HP and FHP to both be picking out (the actual) natural numbers.

[14]At least, I'm not currently able to adequately defend my positions on those matters.

(Case-1).

(Case-2)

As I see it, there are two ways to look at APs belonging to (Case-2). On the one hand, we might think we're working with an entirely new equivalence relation, and hence a new variety of abstract. On the other hand, we might remember our Boolean algebra and notice that APs like BCA are merely restricting us to the intersection of two equivalence classes.

Call the equivalence relation expressed by (the universal closure of) $\neg F \approx \neg G$ *binumerosity*, and the associated abstracts *binumbers*. Since we are looking for the canonical equivalence relation associated with BCA, we ought to ask whether conumbers are actually numbers or binumbers. The answer is, of course, both. It may be convenient to give a name (conumbers) the bi-cardinality abstracts, but really we are considering a subclass of numbers that is also a subclass of binumbers.[15]

To look at this another way, if we are given the class of numbers, and the class of binumbers (over the same domain) and took their intersection, we would end up with the same class of abstract objects as we would had we used BCA to pick out the conumbers. The canonical equivalence relation*s* for BCA are both equinumerosity *and* binumerosity.

This analysis gives us the right answer to questions about the natures and identities of the relevant abstracts.[16] Some care does need to be taken though. It's natural to think that conumbers are both numbers and binumbers, but other cases may seem less plausible. For example, we might wonder about

[15]Cook (in private communication) pointed out to me that in most models we might consider, the intersection of the equinumerosity abstracts and binumerosity abstracts will be empty. However, the broader point still holds—nothing hear or in what follows relies on the details of the example of BCA. Any AP whose RHS consists of the conjunction of two equivalence relations that produce disjoint classes of equivalence classes with a non-empty intersection would do just as well (or better) as illustration.

[16]This justification for the foregoing analysis may seem circular, and perhaps it is. The circularity is harmless however. The appeal to intuitive identity claims can be seen as just an appeal to a desirable consequence of the fine-grained approach as a whole. I ask those who are still skeptical to withhold judgment until the full analysis has been carried out, or at least grant that the above is a plausible analysis of the underlying nature of (Case-2).

the intersection of coextensionality and equinumerosity (suitably restricted). Obviously all coextensional concepts have the same number, so it looks like we're forced into Frege's position that numbers are just a subclass of extensions. The solution would be to argue that, in fact, BCA abstracts are *sui generis*, in which case BCA would belong to (Case-1). I don't think this route would significantly affect the rest of my analysis, however.[17]

There will be a bit more to say once some more work is done, but I will first consider how we might identify the canonical equivalence relations for APs belonging to (Case-3), and how those identifications might be justified. What follows relies on some specific assumptions about neo-Fregean metaphysics and epistemology that are not part of the core of neo-logicism that we can assume to get a project like this off the ground (see 1.2). Nevertheless, I think it's a plausible route to understanding and identifying canonical equivalence relations. I also think there there are other routes to the same destination, e.g. via previous commitments to solutions to Good Company or Caesar as mentioned, or by making use of an analysis of APs as analytic or implicit definitions (but see chapter 4). But that's a project for another day. Even if you find the assumptions in the next section problematic, it will be possible to adopt them temporally for the purpose of picking out canonical equivalence relations, as what is important is the association of canonical equivalence relations with APs more generally.[18]

5.3 (Case-3)

Content (re-)Carving

There is a long and interesting thread in the neo-logicist literature that will help with the current analysis. In his *Grundlagen der Arithmetik* (Frege, 1884),

[17]The first option coheres well with the general theory of abstracts sketched at the end of chapter 3. For example, the thought might be that the universe of abstracta is carved up into sorts of abstracts based on the equivalence classes of concepts they're associated with. We would then have a case where that 'structure' contains overlapping substructures.

[18]Whatever stand you take on this issues, the following analysis should be of interest for methodological reasons relating to the use of epistemic concepts in the analysis of APs, which will be particularly apparent in the next chapter.

Frege famously says this:

> The judgment 'line a is parallel to line b', in symbols,
>
> $$a//b,$$
>
> can be understood as an identity. If we do this, we obtain the concept of direction and say: 'the direction of line a is identical to the direction of line b'. We therefore replace the symbol // by the more general =, while we redistribute the particular content of the former expression to a and b. We carve up the content in a different way than the original, and thereby produce a new concept. (Frege, 1980, §64)

This has been picked up by a number of authors and developed in various ways, most often with the aim of explaining the epistemic or metaphysical import of abstraction principles, be that as contextual definitions, implicit definitions, or something else.[19]

The general idea is that the content represented on the LHS of an AP is simply a recarving of the content on the RHS, or if you prefer (as Fine, 2002, does), a reconceptualization. How this is to be cashed out is contentious, but luckily most of the contentious details can be ignored for present purposes.

There are three important things to keep in mind. The first is that although the recarving of content presupposes a sameness of content, there are a narrow range of possibilities available if 'sameness' is to be cashed out in a useful way. If we're too strict, all of the cases of genuine content carving will be trivial and the cases we care about will be excluded. If we're not strict enough, formulae that have little or nothing to do with each other will count as providing recarvings of the same content, rendering the the concept of recarving unfit for (any) purpose.[20]

The second thing we ought to keep in mind is that there are are both metaphysical and epistemological elements required for successful content carving. On the metaphysical side, Linnebo glosses content recarving as "a

[19] See Hale (1997); Fine (2002); Rayo (2013); Linnebo (2018).

[20] Linnebo (2018, chapter 4) discusses this problem at some length.

matter of two statements' imposing the same requirement on reality"[21] , the thought being that no more is required for one side of an AP to be true, than is required for the other side to be true.

Similarly, Augustín Rayo (2013, p. 23) interprets Frege to be saying

> For the direction of line *a* to equal the direction of line *b just is* for *a* and *b* to be parallel.

Rayo has a particular interpretation of 'just is' in mind in the just cited *The Construction of Logical Space*, upon which he bases his entire system. He does describe accepting *just is*-statements as closing a metaphysical gap (pp. vii-viii). That is to say that accepting Rayo's paraphrase of Frege amounts to saying that there is no metaphysical gap between two lines being parallel, and their having identical directions. This jibes well with how I interpret content carving below. Before looking at the metaphysical side of content carving in more detail, however, it will be useful to have a look at an important epistemological feature.

For content carving to be at all useful to neo-logicists, there needs to be what Linnebo calls a *sufficiency condition*, which is to say that epistemic properties like knowledge, or justification must be allowed to carry over from the (representation of) one piece of content to its recarving (Linnebo, 2018, chapter 4). If that weren't the case, knowledge of concepts like one-to-one correspondence or parallelism would be useless in grounding our knowledge of numbers or directions via APs, which is part and parcel of the explanatory usefulness of APs thought of as tools for recarving content. Indeed, it has often been argued that it is precisely the fact that we *can* have such a transfer of epistemic properties across APs that grounds our knowledge of abstracts, and by extension, arithmetic. For this to work, both the epistemic and metaphysical conditions must be met, which is to say that there is no metaphysical gap between the contents of the two sides of an AP, and that epistemic properties can carry through.[22] It also requires that we are capable

[21] in an unpublished draft

[22] I think that the lack of metaphysical gap gives rise to to the epistemic closeness, but nothing here relies on that.

of grasping the content of the RHS of an AP in a manner sufficient to then actually understand the content of the LHS.[23]

What I will now argue is that how we grasp the content of the RHS of an AP, as well as the nature of that content affect whether relevant epistemic properties can actually carry across the biconditional. More precisely, I will argue that how we look at that content, and how it is recarved, will give us insight into which equivalence relations are canonical.

Of Dessert Spoons and Dinner Guests

To see how the role of APs as carvers of content can help us identify the canonical equivalence relations associated with those APs, let's again take the example of HP, our most successful AP. The idea is that we are to have an antecedent understanding of the content captured by the RHS of the AP which will allow us to 'see' that the content on the LHS is the same, just represented in a different way. The RHS of HP says that the objects falling under two concepts can be correlated one-to-one, which HP then tells us is just the same as the concepts having the same number (or the number of objects falling under them being the same).

Take the following adaptation of Frege's well known example of a table settings (Frege, 1884, §70; see also §4.2). If Hera is hosting Thanksgiving for her family and asks Hero to set the table, he might set a plate in front of each chair, a fork to the left and a knife to the right of each plate, and so on.[24] So long as Hera has put out one chair for each guest (and one for herself), Hero will have set the right number of places, and if Hera asks Hero whether he has put out the correct number of dessert spoons, Hero needs only to check that there is a small spoon above every plate. He need not count spoons, or indeed place settings. Hero and Hera are able to go back and forth between questions about numbers of spoons and statements about on-to-one correspondences between

[23]No particular person need actually fully understand the content of either side of an AP, but such an understanding must be in principle (humanly) possible. If important information about the equivalence relation on the RHS of an AP is simply unavailable, that AP will provide us with no better understanding of its associated abstracts than was already possible.

[24]Hero and Hera are siblings with PhDs in philosophy of mathematics. See 6.3 for more about them.

guests and spoons seamlessly, and without loss of relevant information. In other words, Hera was able to recarve the content of Hero's statement about a one-to-one correspondence between dessert spoons and chairs (via plates, which is possible because equinumerosity is an equivalence relation), into a judgment about whether there are the same number of spoons as guests. In other words again, Hero and Hera were warranted in moving from the content of statements about correspondences between the concepts spoons (on Hera's table), plates (on Hera's table), chairs (at Hera's table) and guests (invited by or identical to Hera), to judgments about the equality of the numbers belonging to those same concepts.[25]

Zooming back out, and remembering that HP is an instance of (Case-1), and so wears its heart (canonical equivalence relation) on its sleeve, we can now consider how such content recarving might play out in instances of (Case-3).

Ignoring for the moment that BLV is inconsistent, we might apply an analysis based on content carving and say that understanding what it means for two concepts to be coextensional will allow us to recarve that content and instead refer to extensions—the objects associated with coextensional classes of concepts. Now consider New V. We are asked to use our understanding of what it is for two concepts to be coextensional *or* for both concepts to be equinumerous with the universe. Putting aside the difficult issue of whether we can understand what it means to be equinumerous with the universe, we can ask whether a faithful recarving of the content of the RHS of New V could give us extensions for small concepts while lumping the rest together.

At first it looks like that might be fine; we have recarved the content into (identity claims about) extensions and have carved *out* the problematic concepts—those that are too big. At the risk of stretching the metaphor too thin, we should then ask whether *carving out* is, or can be, a type of *re*carving. The language of the metaphor suggests not, but our metaphysical and epis-

[25] As we saw in chapter 4, had Hera instead been cooking for the guests at Hilbert's hotel, asking about the number of spoons would have been ambiguous, though the answer provided by HP would have been the correct one in this context. Luckily Hera didn't have to entertain such a large (and unhappy) crowd.

temic desiderata regarding content carving will give us a more satisfying answer. First we should recall a feature of (Case-3) APs like New V: the choice of size restrictions on concepts is constrained only by the threat of contradiction (or unsatisfiability). For example, there is an AP I'll call *c-V* that restricts the applicability of coextensionality to concepts with cardinality no more than that of the continuum.[26] With that in mind we can see whether such APs recarve content in a manner that meets the above mentioned metaphysical and epistemological criteria.

The example of c-V presents an immediate problem—it could easily be argued that at worst we can't understand the RHS without first deciding on the status of the continuum hypothesis (CH), and at best, we can't evaluate the success of the content carving without deciding CH. Issues like this may provide us with additional evidence of the undesirability of including size restrictions in APs, but that need not concern us right now.

Leaving thorny questions about the nature of the mathematical universe aside, we should still be wondering whether epistemic properties can carry through the AP from right to left. We are immediately faced with a dilemma. If we conclude that the requisite epistemic carry through is impossible, we should throw out such APs, since we are committed to content carving as part of the explanation for their importance. APs can't do the epistemic work we've asked of them; case closed. If we find that such APs do meet a proper sufficiency condition, we will notice that all of the useful information to be gained by recarving comes from coextensionality in the case of APs like New V, or equinumerosity in the case of FHP, which is to say their respective canonical equivalence relations.This at once suggests a method for picking out canonical equivalence relations: they are the equivalence relations that allow fine-grained information to be passed from right to left when APs are understood as content recarvers.[27]

We can make a similar move from the metaphysical point of view. Content recarving requires that there be no metaphysical space—no gap—between the

[26] I've called this AP 'c-V' because 'c' is often used to denote the cardinality of the continuum.

[27] I say "fine-grained" information because *something* is captured by acknowledging the existence of a dummy abstract.

content of the two sides of an AP. I admit that I have only a tenuous grasp on the idea of metaphysical space in this sense, but I hope that what follows will be clear without my having to articulate a theory of metaphysical gaps in any great detail.[28]

As I hope the example of Hera's thanksgiving demonstrated, (an informal version of) HP allows us to move back and forth between (representations of) content involving one-to-one correspondences, and content involving numerical identity. Likewise, New V allows us to move back and forth between content about the coextensionality of small concepts, and the extensions of small concepts. But what about universe-sized concepts? In that case there is a mismatch between the right and left hand sides of New V. There are no metaphysical fences between the classes of concepts of different cardinalities; indeed, we saw that we could easily use c-V instead of New V if we're just trying to pick out some extensions. If we were to do so, we might end up picking out fewer extensions, and sending more cases to a dummy abstract, but the nature of the content recarving would be essentially the same in both cases. That is to say, in both cases content about coextensionality is being recarved into content about extensions and a single dummy abstract.

Notice that the LHS of both Vs refer to the same sort of abstracts, and that content carving really ought to go both ways. The necessity of the bidirectionality of content recarving is even more obvious if we adopt either Rayo's *just is* operator or something like Linnebo's full account of sufficiency conditions as (perhaps partial) explications of the behaviour of APs.[29] These observations certainly make it look as if adding size restrictions of the RHS of APs introduces a metaphysical gap that isn't present in (Case-1) or (Case-2).

Since the LHSs of New V and c-V turn out to be the same,[30] they must not capture the cardinality restrictions. How could they?—the LHSs of APs merely state that particular terms pick out the same objects. This will cause

[28] I find the account in (Rayo, 2013) attractive, but a detailed discussion of that view would take us too far afield.

[29] Linnebo's account would have to be modified to be suitable here, as it rules out many of the cases we care about as content recarvers.

[30] This is at least a reasonable assumption. In any case, we have no reason to think that the abstraction terms of either AP are obscuring information about cardinality restrictions.

problems if we want content carving to work in both directions. And we can make a similar observation in the epistemic case. If we understand the LHS of New V, it isn't obvious that that understanding would transfer to the RHS (unless we insist that understanding extensions necessarily involves understanding e.g. Cantor's theorem).

At this point it will be worth pausing to deal with a possible objection. Since we agree that the *important* direction for content carving is right to left, perhaps we could relax our requirement that content carving go both ways, thereby greatly weakening the preceding argument. I find this move unattractive, but admit that it isn't necessarily an unreasonable move to make, especially if our goal was to rule out size restricted APs. That said, it is natural to cash out 'sameness of content' in a way that carries a commitment to bidirectional content recarving.

If we think of content as behaving less like a roast that can only be carved once, and more like clay that, in careful hands, can be reformed endlessly, we can begin to get a handle on bidirectional content recarving.[31] The thought is that if we start with some region of metaphysical space that's carved up into equivalence classes, and then recarve that content into abstract objects, we ought to then be able to change our minds and return to where we started without any loss of content. If there really is *no* metaphysical gap between the content on two sides of an AP, and we take metaphysical content carving talk seriously, there ought to be no reason that the content in question shouldn't permit recarving in any number of ways—all of the same 'material' is available.[32]

Another reason we might think that content carving is loss-less in both directions is that there might be cases where we have a previous understanding or even knowledge of some abstract objects, but aren't sure exactly what

[31] In fact it might have been better if we had spoken content reformation all along. But the language of metaphysical content carving is too well established to be easily dislodged. Fine's use of 'reconceptualization' arguably captures what's going on better than content carving, but I find the term clumsy and less conducive to analogical reasoning. Furthermore, Linnebo uses reconceptualization explicitly for an asymmetric relation in a similar context (Linnebo, 2018, chapter 2)

[32] For what it's worth, I think that carving a roast duck introduces a metaphysical gap.

equivalence classes they are associated with. We might be in such a position if we want to integrate abstracta presented in other ways into an abstractionist system.

In any case, what we notice in the mathematical cases we're concerned with here is that the metaphysical and epistemic space between the abstracts referred to on the LHSs of APs and coextensionality (or equinumerosity) remains static across the introduction or variation of size restrictions on the RHSs. Our understanding of one-to-one correspondences carries over to numbers just as easily whether we adopt HP or FHP.[33] Likewise, the metaphysical gap between an ontology carved into extensions and an ontology carved up into coextensionality equivalence classes remains equally nonexistent whether we wish to use New V or c-V—*that* content remains the same.[34] Any pair of coextensional concepts with continuum-many or fewer objects falling under them is associated with the same abstract object, in the same way.

The takeaway here is that the canonical equivalence relation associated with a given AP is the equivalence relation that allows for successful lossless, bidirectional content carving regardless of whatever else might be going on on the RHS of that AP. Note that this accords well with the analysis of (Case-2)—both conjuncts will contribute equally to the content carving.

Dummy Abstracts

I've mentioned the presence of dummy abstracts in cases of size-restricted APs a number of times, but have had little to say about them thus far. The presence of dummy abstracts can be a useful heuristic for picking out canonical equivalence relations, though it's not clear exactly how to think about them from a metaphysical perspective.

When dealing with APs like New V, we're often told that they produce such-and-such abstracts for concepts of certain sizes, and send everything else to a dummy abstract. We can then ask why there is such a thing as a

[33]This point is developed in §5.5, below.

[34]Chapter 6 deals explicitly with the connection between canonical equivalence relations and the cardinalities of the classes of abstracts picked out by APs.

dummy abstract. The answer is that we are essentially throwing a bunch of concepts into a class we might call 'too big'. We do that so that we can avoid restricting our (second-order) domain or comprehension scheme.[35] Which concepts are too big is entirely determined by the relevant size restrictions, which if I'm allowed a bit more bootstrapping, are *not* (part of) the canonical equivalence relations.

Looking a bit more closely at dummy abstracts, we might think that their very presence counts against the notion of a canonical equivalence relation. After all, FHP purportedly picks out something that HP doesn't—its dummy abstract. Surely that tells us that there is a something fundamentally different on the RHS of those APs. Whether we accept this line of reasoning comes down to how we interpret size restrictions. Formulae like '$F \approx V \wedge G \approx V$' are naturally read as the conjunction of two formulae, each free in one second-order variable. Interpreted thus, size restrictions needn't be seen as equivalence relations, because, *prima facie*, they're not relations at all. In that case, if we want dummy abstracts to be associated with equivalence classes, it will have to be equivalence classes associated with the entire RHS of the AP in question. We might also think that dummy abstracts aren't associated with equivalence classes at all, or at least that any such association they do have is independent of the relevant AP.

One the other hand, we could interpret size restrictions like the one in the version of New V I've been working with as expressing the relation that holds between two concepts just in case both concepts are equinumerous with V. This is a bit of an odd relation, but it's obviously an equivalence—it's equinumerosity for just one size of concept. In this case, we can interpret the size restriction as a second canonical equivalence relation. In that case, presumably any size restriction based on the same cardinality would be associated with the very same dummy abstract.

Consider an AP that's like FHP, but instead has the same size restriction as BLV, namely a restriction to concepts that are (strictly) less than the size of the universe. Our new AP, call it SHP for Small Hume's Principle, and New

[35] There are other reasons we might do this too. See Heck (2011b).

V send exactly the same pairs of concepts to a dummy abstract. We can then ask ourselves whether those dummy abstracts are distinct. It certainly isn't obvious that they are.

In any of these cases, if we know that an AP produces a dummy abstract, we can usually also see *why* it does so. In other words, we will often be able to see what equivalence relation(s) is(are) associated with the abstracts we introduced the AP to pick out, and what's producing the dummy abstract. The former is(are) the canonical equivalence relation(s). Of course this won't always be possible, but it will often be a good starting place, especially when we're dealing only with a small number of APs.

5.4 Pragmatics and Recap

We now have the tools to identify the canonical equivalence relation(s) for any abstraction principle we come across, but before laying out those tools, I'll discuss some more general considerations and heuristics that will make the identification and use of canonical equivalence relations more tractable for practical purposes. This will also allow use of the concept without the need for a particular interpretation of content recarving.[36] After discussing those considerations and briefly laying out my analysis covering all four cases in this section, I will sketch approaches to Good Company and abstract-identity questions made possible by our new found ability to identify canonical equivalence relations.[37]

A Few Pragmatic Considerations

My analyses so far have required a certain level of familiarity with APs and some accompanying intuitions, which were then bolstered by more concrete (or at least more detailed) conceptual and technical considerations. I'm now

[36] As I mentioned above, I think there are other routes to the identification of canonical equivalence relations, that will depend on other aspects of your interpretation of neo-logicism.

[37] I originally developed the idea of canonical equivalence relations when working on the problem of fishiness, and the concept will be uniquely useful in solving that problem, but further discussion will have to wait until I've set that problem up in the next chapter.

going to suggest that for practical purposes we can rely more heavily on our understanding and intuition, at least in the majority of interesting cases.

In most simple cases it will be obvious to those of us steeped in abstraction what the canonical equivalence relation(s) associated with an AP is, but the following heuristic may help in other cases.

For many APs we already have an idea of what sort of abstracts we expect it to pick out, so we can often begin by asking ourselves what equivalence relation would be associated with that sort of abstract were we to disregard other details included on the RHSs of APs for other reasons—to guarantee satisfiability, for example. In many cases this will be obvious, or if it isn't we are given an informative name for the abstracts (number, singleton abstract, extension) and can work back from there. In more difficult cases it will be an advantage to first make sure our platonist hats are firmly placed on our heads, as that will make it easier to see the connections between equivalence classes and other sorts of abstract objects.

This strategy, along with the heuristic involving dummy abstracts discussed at the end of the last section, ought to give us all we need to make use of the concept of a canonical equivalence relations in the manner sketched in the next section. It is important to keep in mind however, that absent an independent justification for identifying certain abstracts associated with different APs, these are merely heuristics designed to let us make use of the notion of a canonical equivalence relation once we've already agreed that such a notion is available. The 'official' method of picking out canonical equivalence relations is analyzing the metaphysical and epistemic gap(s) between the two sides of an AP.

Before looking at some possible uses of canonical equivalence relations, I'll recap our methods for picking out canonical equivalence relations.

Cases 1–3 again

In the first three cases, we can look at what equivalence relation allows for seamless transfer of understanding between the content of the two sides of the AP, or ensures no metaphysical gap arises between the content on each

side. That's the canonical equivalence relation. In (Case-1) we should usually be able to identify the canonical equivalence relation pretty much at a glance. In (Case-2) we need only ask ourselves whether there is an overlap in the classes of abstracts being picked out. If there is, *both* equivalence relations are canonical. If not, we should treat that AP as an instance of (Case-4). In (Case-3) we need to keep in mind that size-(or other) restrictions may need to be carved away for the canonical equivalence relation to be identified.

If we come across an AP we've not seen before, it may not be obvious which case it falls under, especially if it's in one of the forms at the very beginning of this chapter, i.e. the equivalence relation is represented by a single operator. In such cases we ought to behave as if the AP is (Case-3) or (Case-4), and depending on the situation, apply either the 'official method' or the heuristics mentioned just now. (If the equivalence relation is represented by a single operator, we may not be able to use heuristics involving dummy abstracts unless we have a way to unpack the RHS.)

(Case-4)

For the most part (Case-4) can be dealt with in a similar manner to (Case-2) or (Case-3), but there is one major difference: APs in (Case-4) produces (at least) two varieties of abstract that don't necessarily overlap.[38] Luckily this admits of a straightforward solution. If there is no overlap between the classes of abstracts picked out by the AP, we can treat the single (Case-4) AP as two APs which could in principle belong to any of the four cases. If the classes of abstracts overlap, the overlap can generally be seen as analogous to (Case-2). Here, as in (Case-2) we must of, course, not be tricked by the syntax of the LHS into thinking that the AP must be associated with only one type of abstract, or that the kinds of abstracts it produces are unique (which is kind of the point of all of this).

[38] For a worked out example of (Case-4) see §6.5.

5.5 Using Canonical Equivalence Relations

In this section I'll sketch a couple of ways the notion of a canonical equivalence relation might be useful in advancing the neo-logicist project. However, one caveat ought to be mentioned first. For the details of the following to be worked out, especially given the formal nature of much of the recent neo-logicist literature, a formal precisification of the definition of a canonical equivalence relation will likely be needed. This appears to be possible, though I don't yet have a worked out solution.

Identity

I argued in §5.2 that we generally act as if the abstracts picked out by BLV, New V, and c-V are all extensions, the abstracts picked out by HP and FHP are all numbers, and so on. This makes sense, but now we have a way of articulating better *why* it makes sense. Using canonical equivalence relations, we can easily state a sufficient condition of the identity of abstracts.

Abstracts a_1 and a_2 are identical ($a_1 = a_2$) if they are associated with the same equivalence class of the same canonical equivalence relation.

Note that the reason this condition fails to be necessary is exactly analogous to the Caesar/C-R objections. The only difference is that instead of considering the objects picked out by individual APs, we're concerned with the abstracts picked out by any APs with the same canonical equivalence relations. Shifting our focus in that way doesn't give us more information about whether abstracts are identical to Julius Caesar, or whether abstracts associated with different canonical equivalence relations can be identified.

Good Company

The above sufficiency condition for abstract identity will also be helpful in solving Good Company. Recall from §1.3 that the Good Company problem, first identified by Mancosu (2015) as a generalization of an argument due to Heck (2011a), points to the fact that many of the abstracts we care about (mathematically), can be picked out by more than one AP. The problem then is that

many of the alternatives are *prima facie* acceptable—hence 'Good'. Mancosu (2016, §§4.4–4.6) picks out FHP, as well as APs that that are meant to capture how some 19th century mathematicians (including Peano) understood (cardinal) number.[39] All of these agree with HP with respect to the finite numbers, and allow derivations of the axioms of PA^2. So which one do we choose?

Since we can identify the canonical equivalence relations of all of these APs as either just equinumerosity, or equinumerosity and binumerosity, we can apply the above identity condition, and conclude that, perhaps apart from dummy abstracts, all of the good companions are picking out numbers (at least in finite cases). We're in a similar situation as before, but making this identity claim explicit we are able to be more precise about what is required of a solution. There are two lines of thought that I can see.

Mancosu (2016, §4.6) builds on a line of reasoning due to Heck (2011b) which roughly says that if we are trying to capture our intuitive understanding of arithmetic, there doesn't seem to be an in-principle way to distinguish between a host of Good APs (for reasons similar to those discussed in chapter 4). It's difficult, if not impossible, to pin down a common, pre-theoretic understanding of infinite numbers. We can now go some way towards fixing this problem, though I think we'll ultimately have to throw out a reliance on 'intuitive understanding' to successfully ground the epistemic significance of APs.

The partial answer we can give is that when we're concerned only with the natural numbers, our understanding proceeds via our understanding of equinumerosity—the canonical equivalence relation associated with numbers—not via the entire RHSs of the APs. From this perspective, it truly doesn't matter which AP we choose. But we will still presumably want to say something about the infinite. Unfortunately, there doesn't seem to be a principled way to decide how to pick a principle. There's another way to look at this, though.

If we are serious about abstract objects being metaphysically thin, as well as about (acceptable) APs being a reliable way of acquiring knowledge or

[39] In fact Mancosu calls FHP 'PP' for the Peano Abstraction Principle. I've adopted Cook's nomenclature because it makes it more obvious what's going on.

understanding of those abstract objects, it is reasonable to commit ourselves to some sort of maximality principle for APs.[40] By this I mean a principle that, given a canonical equivalence relation (or some canonical equivalence relations), we ought to adopt the least restrictive APs associated with that canonical equivalence relation. In other words we ought to choose the AP that picks out the most abstracts with respect to a given canonical equivalence relation without leading to paradox. Such a principle would tell us to pick HP (or BCA) in the cases Mancosu is worried about, New V if we're wondering about extensions, and so on. We're now a good way down the road to a solution to Good Company.

5.6 Concluding Remarks

I've argued that we can pick out the *canonical equivalence relations* associated with APs, and sketched how the identification of those relations might help us move forward on difficult problems faced by neo-logicists. Much of what I've done here relies on integrating the epistemic goals of neo-logicists into our reasoning about the APs. I've also had to thread some pretty thin conceptual needles to get us to where we are. That said, I've tried to at least give the impression that the notion I'm aiming at isn't reliant on our ability to thread those particular needles. I think canonical equivalence relations are a general feature of APs understood as neo-logicists understand them.

In the next chapter, I use the notion of a canonical equivalence relation to shed light on an aspect of Bad Company known as the problem of fishiness. That allows us again to integrate epistemic concerns into the analysis of that complex and difficult problem, as well as to suggest similar methodologies be employed in investigations of Bad Company more generally. It also demonstrates the usefulness of the notion better than the brief sketches I've given here could.

[40] A commitment to thin objects is generally accompanied by a commitment to ontological maximalism or metaontological minimalism, fairly common views among neo-Fregeans. See Rayo (2013); Linnebo (2018) and §1.2 of the introduction.

Chapter 6

Goldilocks and the Fishes

Neo-logicists, and especially those working in the tradition of Bob Hale and Crispin Wright ("Scottish Neologicists") (Wright, 1983; Hale and Wright, 2001) are looking to ground mathematics and mathematical knowledge on abstractions principles (APs):

$$\S F = \S G \leftrightarrow (F \sim G)$$

The two most well-known examples are

BLV $\varepsilon F = \varepsilon G \leftrightarrow (F \equiv G)$

HP $\#F = \#G \leftrightarrow (F \approx G)$

where '$\equiv$' denotes coextensionality, and '$\approx$' denotes equinumerosity, or equivalently, bijectability.

HP is consistent (if arithmetic is), and is at the core of the neo-logicist project because the (second-order) Dedekind-Peano axioms can be derived from just HP and the background second-order logic (plus some appropriate explicit definitions). This result is known as *Frege's Theorem*.[1] The problem is now this: If HP is to have the special epistemic status needed to ground our knowledge of arithmetic, why does the inconsistent BLV not have this

[1] See §2.5 for a historical discussion of Frege's Theorem.

same status? This is the *Bad Company* objection. The problem is made much worse by the availability of APs that are individually satisfiable, but jointly unsatisfiable or inconsistent. This latter version of Bad Company implies that we'll have to do more than just ban inconsistency if we're to delineate a class of acceptable APs

The project has been to find a way of delineating a 'natural class of acceptable abstractions'. It is one major thread in this program that will be my focus here. I will begin by setting up a problem with one of the most popular proposals—strong stability—and analyzing that problem as a tool for investigating what we might mean by 'natural' and 'acceptable'.[2] Before doing so however, I will outline some minimal, necessary conditions on a class of acceptable abstraction principles, as well as on solutions to Bad Company.

6.1 Restrictions on a Solution

The sorts of solutions to Bad Company that I'm concerned with here are what have been called *model-theoretic acceptability criteria* as they are couched in terms of the models of APs constructed in a strong metatheoretical background theory, usually ZFC. This means that such solutions will not be available to what Kit Fine (2002, p. 10) calls uncompromising abstractionists—those neo-logicists who eschew the use of any mathematics not already reduced to logic and APs, or who at least require that the ladder of traditional mathematics can be kicked away when we're done. Although it's an interesting question whether Bad Company is soluble from such a position, such concerns are tangential to our current question. That said, even 'compromising' neo-logicists should be careful to avoid acceptability criteria that are in some way *ad hoc*, question begging, or otherwise unnatural *from their own philosophical perspective*. What I mean by this will become clear as we get into more detail below.

If the neo-logicist program is to be successful, some account has to be given of the special epistemic status of APs. Frege held (in)famously that BLV

[2] See §6.3 for a definition and discussion of strong stability. There is also a list of definitions in appendix B.

was, true to its name, a basic logical law. Since HP was derived from BLV in pure logic, it was, by his own lights, analytic. Thus because numbers were extensions, they counted as *logical objects*. This line won't work for neo-logicists (who take HP as primitive), but it has been argued, most famously by Crispin Wright, that HP is analytic.[3] That position has largely gone out of favour, I think for good reason, but some other special kind of conceptual truthiness, quite possibly bolstered by an account of reconceptualization or content carving, must be at play.[4] The point here is that any formal restrictions on the class of acceptable APs ought not conflict with the epistemic foundations on which we've chosen to ground our acceptance of APs.

To that very weak requirement we should add independent philosophical arguments for the use of any acceptability criteria we wish to employ. That some criterion "gets it right" isn't enough. One way to go is to argue that the criteria are, in some philosophically robust sense, *natural* and not *ad hoc*. The bulk of this paper will look at a concept that's been dubbed 'fishiness'. The discussion of fishiness will bring us closer to understanding *why* certain classes of APs should be excluded, beyond their being for some reason or another 'problematic' or 'unnatural'. I will also sketch two possible ways to rule out fishy APs, one based on the notion of a canonical equivalence relation developed in chapter 5, the other on a proposal in (Cook and Linnebo, 2018). Before that I'll discuss some minimal acceptability criteria.

6.2 Minimal Acceptability Critieria.

There are a few basic criteria that should be met by acceptable APs, and thus permitted by any stronger criterion we consider. The justifications for these minimal criteria are partially grounded in the background assumption that we are concerned primarily or exclusively with mathematics (as opposed to other domains that might make use of APs). Abstraction principles can be

[3]Frege's understanding of analyticity differs from more recent conceptions like Wright's and Quine's, and also from Kant's, against whom Frege was ostensibly arguing. See (Schirn, 2006) for an analysis, and chapter 4 where we argue that neither Frege, nor modern neo-logicists should take HP to be analytic in Frege's sense.

[4]See chapter 5 for a discussion of content carving in a similar context.

used to provide accounts of other varieties of abstracta like letter types and propositions, but we don't look to have a *prima facie* reason to think that the acceptability criteria for abstraction principles in other domains need to match those needed for a (neo-)logicist account of mathematics (although it would be pleasing if they did). I will say more about this when discussing specific cases. For now the point is only that I will assume that we care only about logical or mathematical applications of abstraction principles.

Consistency

It should be obvious from the example of BLV, or just general considerations regarding our (classical) mathematical goals, that APs from which contradictions can be derived need weeding out. However, since we are working with fairly strong systems using full second- (or higher-) order logic, it isn't always easy to determine whether an AP is consistent in this sense. For example, we have good reason to think that HP is consistent, because Frege Arithmetic (FA) (the name for the system of second-order logic with HP as its sole non-logical axiom), is mutually interpretable with second-order Peano arithmetic, which we have very good reason to think is consistent (see Boolos, 1995, 1996; Burgess, 2005), not to mention the fact that HP has a nice model—the natural numbers. The effect of this is that we want our more complex/restrictive proposals to allow us to weed out APs that are inconsistent for well understood reasons. One way to start weeding out inconsistent APs is to solve what Roy Cook[5] has dubbed the *Goldilocks problem*. The Goldilocks problem is the problem that we need to restrict ourselves to APs that neither produce too many abstracts, like BLV, and are thus unsatisfiable, nor too few abstracts and thus run into other problems with respect to the mathematical domains we care about.

That said, in practice most of the model theoretic acceptability criteria that I'll consider are designed in part to guarantee satisfiability on (first-order) domains of certain sizes. Such criteria will at least rule out APs that

[5] in private communication

are inconsistent for the usual reasons, and criteria such as *irenicity* aim to do the same for collections of APs.[6]

Infinite Satisfiability

Related to consistency and satisfiability is the requirement that APs be satisfiable on infinite domains. The reason for this is again that our goal in picking out a class of acceptable APs is to use those principles to recover as much mathematics as possible. Even if we're focused only on arithmetic, we'll need a countably infinite domain, though this is unproblematic given the almost universal desire among neo-logicists to ensure that HP—which guarantees the existence of countably infinite cardinal numbers—is acceptable.

Cook (2017, p. 5, fn. 11) sketches the obvious argument:

> - Neo-logicism is a non-starter if the paradigm 'good' instance Hume's Principle is not acceptable.
> - Hume's Principle has models whose first-order domains are of size κ if and only if κ is infinite.
> - All acceptable abstraction principles must be compatible – that is, co-satisfiable – with one another.
>
> It follows that any acceptable abstraction principle must have an infinite model.

Demanding satisfiability on an infinite domain causes a dialectical problem, but no serious logical or mathematical concerns. The dialectical problem concerns what Weir (2003) calls the Embarrassment of Riches problem which is often motivated by nuisance type principles which are only satisfiable on finite domains, so are jointly inconsistent with HP.[7] Sean Walsh and Sean Ebels-Duggan (Walsh and Ebels-Duggan, 2015) give a generalized nuisance

[6] An AP is *irenic* iff it is conservative, and jointly satisfiable with every other conservative principle. A list of definitions of acceptability criteria is provided appendix B.

[7] The Embarrassment of Riches problem is the version of the Bad Company objection that is specifically to do with APs which are individually, but not jointly satisfiable.

principle. Unpacking their definitions we have:

$$\ddagger F = \ddagger G \leftrightarrow (\{x : Fx\} \backslash \{y : Gy\}) \cup (\{y : Gy\} \backslash \{x : Fx\}) < \omega$$

which says that the symmetric difference of the sets of objects falling under concepts F and G is Dedekind finite.[8] This forces the domain to be finite, and is thus disallowed by requiring infinite satisfiability. That means nuisances won't be available to motivate the Embarrassment of Riches version of Bad Company. However, there are plenty of other problematic cases. For example, it is well known that there are APs that are only satisfiable at limit ordinals, and others that are only satisfiable at successor ordinals.

Given all of that, one might wonder why we would include infinite satisfiability as a separate criterion at all, especially given that most of the proposed model theoretic solutions to Bad Company entail this anyway. My reasoning is that this criterion is motivated by the desire to either capture or interpret a large amount of extant mathematics, or to independently develop a mathematically strong system (see chapter 3 and Fine, 2002), either of which will require an infinite domain. I take this motivation to be independent of the project of delineating a class of true principles. That said, it may be the case that if we are looking to employ APs in non-mathematical domains, we'll need APs that aren't satisfiable on infinite domains.

The more general idea here is that independent mathematical or philosophical justifications for acceptability criteria for APs will strengthen the neo-logicists' case, and provide a more robust ground for the epistemic goals of the project.

Purely Logical

The final basic requirement on APs, again directly related to the goal of recovering mathematics is that they be *purely logical*.[9] This simply requires that the

[8] Ebels-Duggan (2015) show that under the assumption of the *pairing axiom* these principles are inconsistent in infinite models. He glosses *pairing* as "if the universe is Dedekind infinite, then there is an injection from pairs of objects into the universe" (p. 265).

[9] It's an interesting question as to how much mathematics we want to recover. If we're following Frege, geometry is synthetic, and thus not amenable to the same sort of treatment as

equivalence relation on the RHS of an AP be expressible in pure higher-order logic. This criterion is often taken to be just a simplifying assumption, and it does often make things easier. However, as with infinite satisfiability, pure logically can be justified independently.

A gloss on one such justification goes as follows. Neo-logicists assume a special epistemic status for higher-order logic that is on par with that often attributed to first-order logic. This still leaves us the task of establishing the epistemic credentials of APs.

Given that we accept the special epistemic status of logic, plus the neo-logicist commitment to the continuity of mathematics and logic, it wouldn't seem natural to allow non-logical concepts in the APs we are using to justify or construct mathematical theories. The exception to this is the inclusion of abstraction operators. If we are working with a system of APs, there is no in-principle reason not to allow the various APs or the abstracts picked out by them to interact explicitly,[10] because they are presumably all acceptable. In other words, there is no reason to disallow abstraction terms from appearing on the RHSs of other APs.

It's worth reiterating that if we're looking to ground knowledge of abstracta in other domains, we may very well need to employ non-logical concepts in equivalence relations. For example, an AP for letter types will presumably have to reference letters; APs for propositions, sentences; and so on. Allowing such things in the case of mathematics runs counter to the core of neo-logicist thinking, however.

There's another problem with this criterion that will have been noticed by anyone familiar with the literature on logicality more generally. Even accepting second-order logic as logic proper, the question of which operators count as 'logical ' in the relevant sense can hardly be seen as settled. Frege

the rest of mathematics. Then there's the debate about foundations – set theory vs. category theory vs. HoTT. . . . Most neo-Fregeans are in the set-theory camp, but that's not the only way to go. See for example Leach-Krouse (2017) on structural abstraction, and Logan (2015, 2017) on abstractionist routes to categories, and chapter 3 on why we should consider such programs.

[10]For example, it is allowable to include abstraction terms from one AP on the RHS of another (provided they are both acceptable).

himself doesn't go much beyond the heuristic of being *generally applicable*. There's much to be said here, especially since we're assuming that higher-order logic *is* logic. For the present discussion I'll make do with the practical, if somewhat unsatisfying answer that 'logical' tracks some standard set of first- and higher-order operators that we are all used to. That said, one proposal due to Cook (2017) is worth mentioning here, as it will come up again in later discussion. Cook suggests that some particular understanding of *invariance* will at least entail logicality, and may exhaust it. This is in line with the more general proposal that the logical operators are those that are permutation invariant (see Sher, 1991). This is important because invariance and related notions have been put forward by Cook (2017), Antonelli (2010), Fine (2002) and Walsh and Ebels-Duggan (2015) as alternative ways of approaching Bad Company.

6.3 Cardinality and Stability

It has often been assumed that *the* sources, or at least some very important sources of unacceptability for APs have directly to do with cardinality considerations. It is not hard to see why. Frege's notorious BLV leads to paradox because it requires there to be as many objects in the first order domain as there are functions, contradicting Cantor's Theorem. Similarly, Hazen's ordinal abstraction principle (Hazen, 1985) leads directly to the Burali-Forti paradox, which is to say that Hazen-style ordinal abstraction allows you to prove that the class of ordinals is bigger than itself. Such principles are *inflationary*, which essentially means that they force the first-order domain to be bigger than it already is, which is impossible in a static setting.[11] This is something we'll come back to below.

On the other side of things, as we have already seen, there are consistent APs that are jointly unsatisfiable for cardinality reasons.

These sorts of cases have naturally lead to the proposal of acceptability criteria that focus on cardinality. An important example, suggested by Heck

[11] Dynamic settings like those developed by Studd (2015) and Linnebo (2018) allow the first-order domain to grow in a sense that makes BLV satisfiable. See §1.4.

(2011d) and recently defended at length by Cook (2012), is *strong stability*.[12]

Definition 1 (Strong Stability)**.** An AP Σ is *strongly stable* iff there is a κ such that it is λ-satisfiable iff $\lambda \geq \kappa$. Σ is λ-*satisfiable* iff Σ is satisfiable on a domain of size λ.

The basic thought motivating strong stability is that adding more things to the domain shouldn't affect the existence of abstracts, or even more simply, once we have some stuff, adding more stuff shouldn't block the existence of the stuff we already had.

Note that neither BLV, nor nuisance principles are strongly stable.

It turns out, however, that strong stability doesn't do all of the work we want it to. It is plausible to think, given most neo-logicist understandings of the role of APs, that acceptable APs should pick out or 'provide' the abstracts needed for their own satisfaction.[13] The canonical example is again HP. HP guarantees there be (at least) countably infinite cardinal numbers by *providing* the next finite cardinal given its predecessor. The easiest way to see this is by starting with Frege's definition of zero: $\textsc{zero} = \#\{x : x \neq x\}$.[14] The cardinal number ONE can then be defined as "identical to ZERO", TWO as "identical to ZERO OR ONE", and so on. It can then easily be seen that that process won't end, which is to say that it gives us an ω-sequence (because it is only picking out successors).

Note that this method of generating the finite cardinals relies not only on some second-order comprehension (see fn.14), but also on the impredicativity of HP—the concepts on the LHS fall within the range of the quantifiers on the RHS.

Heck (2011d) gives a simple demonstration of why strong stability is not enough. Call instances of the following schema *Heck principles*.

$$\partial F = \partial G \leftrightarrow ((F \equiv G) \vee A_\kappa)$$

[12]Cook has since backed away from his claim that strong stability is sufficient for acceptability, though he still maintains that it is *necessary* (see Cook, 2017, p. 5).

[13]This is also originally a suggestion of Heck's.

[14]Because we are working in second-order logic with unrestricted comprehension, we are guaranteed that set abstraction terms like this define available concepts.

where 'A_κ' is a purely logical second-order sentence that is true just in case the universe is at least size κ.[15] Heck principles are *strongly stable* by design, but don't tell us anything at all about the infinitely many abstract objects it says that there are. It's strange to say the there are strongly inaccessibly many objects of necessity (because the acceptable APs, which the proposal says are the strongly stable ones, are true). In other words, Heck principles allow us to assert that the universe is some size κ where κ is strongly inaccessible (say), but they don't give us *any* information about the abstract objects that must exist for the universe to be that size.

6.4 Fishiness

Heck has the following to say about the Heck principle that asserts that the universe is at least the size of an inaccessible.

> But there is obviously something fishy about this principle. What's fishy about it, it seems to me, is that it *requires* there to be inaccessibly many objects, but it does not *provide* them. (p. 232, Heck's emphasis)

And he's right, there *is* something fishy going on here, not, I think, unlike Russell's "theft over honest toil".[16] Often the first reaction people have on seeing this problem is to suggest ruling out 'disjunctive' principles, or better, to rule out Heck principles on the grounds of their being unnatural, or *ad hoc*. Throwing out 'disjunctive' principles is obviously a non-starter. There isn't a principled way of picking out which APs fall under that heading. Think of the case where the only primitive connectives are negation and disjunction; we'd almost certainly end up throwing the orphans out with the asbestos. At first blush, it may look more promising to impose a complexity constraint. The problem with that sort of solution is that the complexity of second-order formulae doesn't look to track anything that we're actually interested in from

[15] Also of interest are cases where the second disjunct is independent of the meta-theory, say GCH, though those cases aren't directly relevant here.

[16] Cook and Linnebo (2018, p. 63) make a similar observation.

a neo-logicist perspective, nor do we have good reason to think it might. Basic Law V is not very complex at all. In fact it's Π^1_1 (it has only universal quantifiers), and only concerns unary concepts. HP, on the other hand, requires existence claims about binary concepts. Furthermore, many of the APs that we care about (for example, the cardinality principles investigated in Cook 2017 and Fine 2002, Chapter 3), turn out to be quite complex precisely in virtue of the fact that they (like HP) are tracking seemingly natural, but complex cardinality properties.

Naturalness or *ad hocity* seem more promising. We're looking for a *natural* class of acceptable APs so, on the one hand, if Heck principles really are unacceptable, we should expect them to be be ruled out anyway (because they're unnatural). On the other, throwing out APs because they *feel* unnatural is unprincipled and may end up begging the question, depending on how the broader question is framed and naturalness understood. The suggestion that Heck principles are *ad hoc* is on the right track. In fact I think that fishiness is closely related to, if not a species of *ad hocity*. I'll have more to say about this in §§6.5-6.6 below, but first we need to get a better grasp on what exactly is going on with these sorts of APs.

Bigger Fish to Fry

The apparent problem with Heck principles, is that they don't *provide* the abstracts needed to fill out the domains required for their own satisfaction. By this I mean they do nothing to help us pick out the relevant abstract objects. Recall from §6.3 that HP not only requires that the universe be at least countably infinite, but gives us an ω-sequence of abstract objects. Heck principles, on the other hand, only say that the universe is a certain size, on pain of Russell's paradox. In the postscript to (Heck, 2011d) in (Heck, 2011b), Heck suggests that we find a way to impose the requirement that 'enough' abstracts be provided. In their paper "Cardinality and Acceptable Abstraction," Cook and Linnebo (2018) take up the problem of fishiness. In this section, I'll review some of their results with a focus on sharpening the boundaries of what we might call fishy APs.

Cook and Linnebo's first attempt at ruling out Heck principles is to define *conservative stability*, which is meant to guarantee that there are enough of the *right sort* of abstracts.

Definition 2 (Conservative Stability)**.** An AP, Σ, is *conservatively stable* iff it is strongly stable with stabilization point κ and in any model of Σ, there are at least κ-many abstracts of the sort characterized by Σ. A cardinal κ is a *stabilization point* of an AP Σ iff Σ is κ-satisfiable, and there is a $\gamma < \kappa$, for all $\lambda,\ \gamma \leq \lambda < \kappa,\ \Sigma$ is not λ-satisfiable.

This looks to be the sort of criterion we need to weed out Heck principles and their ilk, and it does. It just doesn't weed out all of the principles that look (to Linnebo and me, but maybe not Cook) fishy. Cook and Linnebo (2018, p. 63, theorem 1) prove that for any strongly stable AP, there's what I will call an Apogon principle that is conservatively stable.[17] Before looking at that theorem, we will need a couple of definitions.

Definition 3 (Cardinality Equivalence)**.** Two abstraction principles, Σ and Γ, are *cardinality equivalent* iff, for any cardinal κ, Σ is κ-satisfiable iff Γ is.

Definition 4 (Ramsification)**.** The *Ramsification* of an AP Σ, denoted $\mathcal{R}(\Sigma)$, is the sentence obtained when the abstraction operators are replaced with appropriate variables bound by existential quantifiers.[18]

Theorem 1 (Cook and Linnebo 2018[19])**.** Given a strongly stable AP, Σ, there is an AP, Σ^+, that is conservatively stable and cardinality equivalent to Σ.

$$(\boldsymbol{\Sigma}^+) \qquad \Im^+ F = \Im^+ G \leftrightarrow ((\mathcal{R}(\Sigma) \wedge F \cong G) \vee (\neg\mathcal{R}(\Sigma) \wedge F \equiv G))$$

The equivalence relation denoted by '$\cong$' says that either F and G are co-extensive singleton concepts, or neither is a singleton concept. A singleton

[17] *Apogon* is a genus of the family *Apogonidae*, also know as cardinalfish.

[18] This requires third-order logic, though there's a work-around given in (Linnebo, 2011, lemma 3.6).

[19] They actually prove something a bit stronger, and part of this as a corollary, but this is all we need here.

concept is one with exactly one object falling under it (haecceities are an example).

What an Apogon principle does, then, is provide exactly the number of singleton-abstracts needed to satisfy a given Heck principle. As noted by Cook and Linnebo (notes 5 & 6), there are two ways to interpret this result. On the one hand, you could say: Oh good, we've shown that Heck principles, and (perhaps) by extension all strongly stable APs, are acceptable after all; case closed. On the other hand, you might think that this too looks fishy; we've found a class of APs that will let us choose the size of the universe, but they're based on fishy principles, and look more like a technical trick than legitimate logico-mathematical tools.[20] Cook and Linnebo officially decline to adjudicate, but I'll argue for the second option. Once I establish the fishiness of Apogon principles, we'll have a good starting point for figuring out what exactly it is that makes these sorts of APs fishy.

Two points are worth mentioning. First, Apogon principles guarantee only that there are the right number of Σ^+-abstracts to satisfy the given Heck principle, *not* that there are enough Σ-abstracts. This means that Heck principles are still ruled out if we take conservative stability to be a necessary condition for acceptability. Indeed, if we decide that Apogon principles are acceptable, we can rule out Heck principles with the weaker requirement that APs be *critically full*.

Definition 5 (Critical Fullness)**.** An AP Σ is *critically full* iff, for each critical point κ of Σ, any model of Σ of size κ contains κ-many abstracts of the sort characterized by Σ. κ is a *critical point* of Σ iff Σ is κ-satisfiable, and there is a $\gamma < \kappa$ such that, for all λ, $\gamma \leq \lambda < \kappa$, Σ is not λ-satisfiable.

This brings us to a second, more important point. Conservative stability is a *very* strong criterion. In fact, conservative stability is strictly stronger than strong stability *plus* critical fullness (the combination of which Cook and Linnebo call Heck stability)—all conservatively stable APs are Heck stable, but

[20] Which is not to say formal tricks aren't sometimes useful tools, but presumably a neologicist reduction requires more robustness given their epistemic goals.

the reverse doesn't hold—making it the strongest of the model theoretic criteria that have so far been investigated.[21] After arguing that Apogon principles are indeed fishy, I'll look at ways we might rule them out.

The Fishiness of Big Fish

The reason we thought that Heck principles were fishy was that they allowed us to dictate the size of the universe without telling us anything about the things in the universe, and by extension, *why* the universe is that size. In other words, we expect APs to play a particular explanatory role with respect to the size and composition of the mathematical universe that Heck and Apogon principles don't seem to play.

Since the goal is to ground mathematics in logic and APs, the lower bound on the size of models of the resulting abstractionist mathematical theory should be a consequence of that theory, and not 'chosen' ahead of time.[22] Since we've already required acceptable APs to be satisfiable on infinite domains (in §6.2), our starting point is an absolute lower bound of $\aleph_0$, which is unproblematic as HP already guarantees us that many abstracts. If we accept an Apogon principle however, we're pushing that lower bound up with very little gain in mathematical strength.[23] Likewise, while we might think that HP provides an explanation for there being countably many numbers, either by telling us that every collection of numbers has its own number, or by reconceptualizing our concept of equinumerosity, it is much more difficult to see how Apogon principles could factor in any such parallel explanations.

Perhaps an analogical anecdote will help clarify things.[24]

[21] See appendix B for a list of criteria.

[22] I am talking about the lower bound of the size of the models because our starting point is strong stability. If it turns out that strong stability is too strong we'll have to take another approach.

[23] Actually the question of whether Apogon principles provide a gain in mathematical strength is a complicated one. We know that the addition of large cardinal principles to ZFC raises the consistency strength of the resulting theory, but there are significant differences between ZFC and the sort of abstractionist theory that we're considering here. However, even if it turns out that increasing the size of the mathematical domain significantly increases the consistency strength of our abstractionist theory, that in itself is no reason to accept Apogon principles; we're still left with the same fishiness we had before.

[24] The uninterested reader can feel free to skip the following short story, especially if they

Further adventures of Hero and Hera[25]

After graduating with his doctorate from St. Andrews, Hero landed a couple of short postdocs and teaching appointments, but the Tory government was making it more and more difficult for him to stay in the UK, so he had a tough decision to make. Hero eventually chose stability and landed a job at a prestigious management consultancy firm in New York City that valued his PhD and experience with logic and the philosophy of mathematics. But despite his career change, Hero never stopped caring about neo-logicism, and his regular conversions with his beloved sister—now a tenured assistant professor at the University of Alberta[26]—kept him reasonably up to date with the world of academic philosophy.

After being cooped up for a couple of days due to a freak snowstorm (the second one that year), Hero decided to put his mind to work considering the concept of set. He had lost his copy of Jech's *Lectures in Set Theory* (Jech, 1971), but that was OK, he wanted to work from first principles anyway. So Hero started with a couple axioms he knew he would need: extensionality and infinity (because, still thinking like a neo-logicist, he wanted to create a mathematically fruitful theory). Hero then had the basis of an identity criterion for sets, and had guaranteed the existence of enough sets to at least model arithmetic (this should be sounding familiar). At this point our hero wasn't sure how to proceed, but intuitively thought that the mathematical universe is pretty big (but not superhuge), and as a platonist, also thought that there are sets outside Gödel's constructible universe. Recall that Hero's facility with logic was second only to that of his undergrad advisor at Ohio State University, so he was able to reverse engineer a set-theoretic axiom that guarantees that $V \neq L$. That task took him longer than it took the MTA to get the subways running again, so Hero had little time to devote to set theory for some time after that.

already agree that we ought not adopt principles that allow us to dictate the size of the universe.

[25] For the story of their higher education, see (Rossberg and Ebert, 2007). For a thanksgiving tale, see §5.3.

[26] Despite the winters, the offer of a tenure track position at a research institution was too much to pass up, especially when it came with the opportunity to work with Bernie Linsky.

In the mean time, Hera had been invited to speak at a workshop on set theoretic foundations in the City, and of course jumped at the chance to get out of Edmonton in the middle of winter and see her brother. So a couple of weeks later Hero and Hera were drinking Scotch in Hero's tiny Bushwick apartment and got talking about his set theory project. Hera quickly worked out, despite the Scotch, that Hero had somehow posited the existence of a Ramsey cardinal mostly accidentally (they were, after all, accomplished mathematical logicians).

"Extensionality and infinity make sense (though have you read Aczel?), but why on earth would you take the existence of *any* large large cardinal as axiomatic? And besides, what does that get you other than lots of sets?"

"Well, I'm not entirely sure, or, rather, I haven't the foggiest.." replied Hero, "but the universe must be extra-constructible, right?"

"Well, *I* think so, but. . . " she paused "you can't just *decide* how big the universe is! Why not supercompact, or hell, superhuge?"

"Now hold on!" protested Hero.

"At least posit a measurable Woodin for comedy's sake." (The Scotch was starting to have an effect.)

After a worryingly long tangent involving lots of poor and slightly distasteful jokes about the names of large cardinals, Hera brought them back to Hero's set theory project. "But seriously, what makes you think you can just *declare* that the universe is a certain size, and what good is that going to do anyway?" Hero gestured as if he were about to speak. He sipped his drink instead. Hera continued, "It's not even clear without other axioms, what positing *any* large large cardinal would get you in terms of mathematical strength. Even if your theory does turn out to be fairly strong, surely the size of the universe–"

"Don't call me 'Shirley."

"Oy! the size of the universe should be determined more than your vague intuitions that it's 'pretty big'. We generally think that there are accessibles because that assumption implies the consistency of ZFC which we assume anyway. ZFC, by the way, tells us a lot more about sets than 'there's a Ramsey'. Anyway, think about that while I top us up."

"Don't you have to give a talk tomorrow?" inquired Hero. A couple of minutes passed and Hera looked to be ready to get back to their discussion. Hero dove back in. "First of all, it's name is 'Frank'. More importantly, this is an unfinished project. I intended to add more axioms. I suppose the task is to fill in my theory in a way that justifies Frank's existence."

"I'll grant that that *might* be possible. Still though."

The conversation devolved from there, but rest assured that Hera gave a brilliant talk, and was only slightly late to the workshop, and only due delays on the L.

* * *

The moral of that short tale (that we ought not determine the size of the universe simply by fiat) holds even more so for the neo-logicist considering accepting Apogon principles for two important reasons. The first is that, even from an external perspective, it's not at all clear what benefit adding large cardinal-many singleton abstracts to our abstractionist theory would provide. More to the current point are the philosophical goals of the neo-logicist program. While Hero was trying to axiomatize his intuitive notion of set in a mathematically fruitful way with perhaps some larger ontological or foundational goals in mind, neo-logicists are trying to epistemically ground mathematics on APs. So while Hero might need only to articulate a notion of set that he's trying to capture, along with some other guiding principles, neo-logicists have more to worry about than merely what is mathematically interesting or intuitively plausible.

To put this another way, we (and Hera) think that the mathematical principles we adopt should do more than just pick the minimum size of the universe, they should be motivated by philosophical principles beyond intuitions about the size of the mathematical universe; the mathematical principles we adopt ought play an explanatory role at least in our discussion of mathematical objects. And that is even more important for neo-logicists given the explicit epistemic goals of neo-logicism.

So, the point of the preceding anecdote was to illustrate the *ad hocity* of employing axioms or principles that artificially dictate the size of the mathe-

matical universe—our original criticism of Heck principles. If we then insist that APs also provide the right number and sort of abstracts, say by insisting on critical fullness or conservative stability, Heck principles are ruled out, but we're still stuck with Apogon principles facing our original criticism—at least if we confine ourselves to cardinality considerations. But such considerations appear to be the crux of the issue.

The interesting cases for our purposes are the Apogon principles that are the result of instantiating the AP parameter (Σ) in the Σ^+ schema with Heck principles. That's allowed even if we have already ruled out Heck principles, as the Ramsifications of abstraction principles are not themselves abstraction principles.

We can now ask whether $\Supset^+$-abstracts are "of the sort characterized" by such an Apogon principle.

It is worth reminding ourselves at this point that APs act to associate objects with equivalence classes of concepts. HP associates objects (numbers) with classes of equinumerous concepts, but what equivalence classes of objects are at play in the case of Apogon principles? Of course we need not have a name for the equivalence relation, but we should, upon understanding that equivalence relation, be able to see how the abstracts picked out are related to that principle. Here we continue to have BLV apply in finite cases. However, in infinite cases we have a situation analogous to the problem we had with Heck principles, but with one very important difference. In both cases the size of the universe is determined by the sentence A_κ from the Heck principle on pain of contradiction via BLV. The difference is that a clause is snuck in that means that there are at least the number of singleton abstracts needed to fill out a domain the size κ. Note here though, that the sort(s) of abstracts that are produced by any Apogon principle is the same: singleton abstracts and extensions.

Canonical Equivalence Relations (again)

I still haven't answered the question of whether these abstracts are of the right sort, i.e. whether they satisfy the requirements for conservative stability or

critical fullness. Cook and Linnebo don't go into much detail about which abstracts are the *right sort*, presumably because "the sort charactarised by that principle" is straightforward. But we can be more precise. In the previous chapter, I argued that we can pick out the *canonical equivalence relation(s)* associated with a given AP. That notion will allow us to be precise about what abstracts we're talking about when we're asking whether they're of the right sort.

The rough idea is that the canonical equivalence relation associated with an AP is the equivalence relation whose equivalence classes are associated with the the sort of abstracts identified on the LHS of the AP in question. This definition is obviously question begging as it stands, but a more rigorous route to the identification of canonical equivalence relations is laid out in chapter 5, and I think there are others. In that chapter I also argue that, for many practical purposes, we can pick out canonical equivalence relation(s) informally using our knowledge of APs and a couple of heuristics. So let's have a look at Apogon principles.

We have an excellent starting point since we're working with a schema, and have already identified the sort of abstracts we are ostensibly concerned with—singleton abstracts and extensions. In fact, there's a strong sense in which we've already picked out singleton-equivalence and coextensionality as the canonical equivalence relations of Apogon principles. This is because I (and Cook and Linnebo) am happy to refer to the abstracts associated with Apogon principles as 'singleton abstracts', and 'extensions' despite the fact that each substitution instance of Σ^+ is a different AP, with a different equivalence relation on its RHS. This suggests that what we're interested in with respect to abstracts is not the entirety of the RHS of such a principle, but rather what we can now identify as its canonical equivalence relations. Nevertheless, we should have a closer look, especially as that will help with developing a more general theory of fishiness.

The equivalence relations that characterise $\Supset^+$-abstracts can be seen to be composed of four parts: Ramsified APs that provide that basic material (so to speak); a sentence, A_κ, specifying how many abstracts there are to be; coextensionality terms; and singleton abstraction terms. It is the first two

parts that have problematic consequences, and the latter two that determine the nature of the abstracts. As per the heuristics discussed in the preceding chapter, we need not consider the first two components when determining the canonical equivalence relation(s). Therefore, the canonical equivalence relations associated with Apogon principles are singleton equivalence ($\cong$) and coextensionality ($\equiv$).

There are two takeaways here, the first is an affirmative answer to the question of whether Apogon principles provide abstracts of the right sort to satisfy the definition of critical fullness. The second is that the cardinality constraints imposed by Apogon principles aren't directly related to the sorts of abstracts they produce. This can readily be seen by observing that the sentence 'A_κ' doesn't figure into either canonical equivalence relation.

The former point tells us that we need to do more work to rule out Apogon principles. The latter gives us a key component in both defining fishiness and ruling out fishy principles.

6.5 Casting the Net

What all of this has been leading up to is an attempt to capture what it means for an AP to be fishy. So far we've determined that the problem has to do with how the size of the universe is determined, which in turn relates to what kind of abstracts we have and which equivalence relations those abstracts are canonically associated with, i.e. what the canonical equivalence relations associated with the APs in question are. In other words, fishiness has to do with the relationship between how many objects are required to satisfy an AP, and the equivalence classes (and thus abstracts) associated with the canonical equivalence relations of that AP. Putting all of the pieces together I propose the following definition as a starting point.

> *An abstraction principle, Σ, is* fishy *iff the size of the first-order domain required for its satisfaction is not determined solely by the the canonical equivalence relation or equivalence relations associated with Σ or, by*

the nature of the abstracts associated with those canonical equivalence relations.

What's needed for this definition to be useful is a method of determining whether the size of the requisite domain has been determined by a particular equivalence relation or sort of abstract. As with the determination of canonical equivalence relations, the fishiness of an AP (or AP schema) is often easy to determine by inspection—that's how we identified Heck and Apogon principles in the first place—but since we will have already determined the relevant canonical equivalence relations, there is a ready made test for fishiness that doesn't rely on smell, taste, or a vague uncomfortable feeling.

I'll continue to talk about Apogon principles since that's the most important case we're currently aware of, but what follows will easily generalize. I'll also cover the case of Heck principles briefly below. The following three APs that share canonical equivalence relations with Apogon principles, will be useful our analysis.

SAP: $\$F = \$G \leftrightarrow F \cong G$

NewV: $\varepsilon F = \varepsilon G \leftrightarrow (F \equiv G \vee (F \approx V \wedge G \approx V))$

SV: $\clubsuit F = \clubsuit G \leftrightarrow [(F \cong G \wedge (F < \omega \vee G < \omega)) \vee \ldots$
$\ldots \vee (F \equiv G \wedge (\omega < F < |V| \wedge \omega < G < |V|)) \vee (F \approx V \wedge G \approx V)]$

where 'V' is the universal concept, $[x : x = x]$.

Call these the singleton abstraction principle (SAP), New V, and Singleton V (SV), respectively. SAP just gives us singleton abstracts; how many will depend on the underlying domain. SAP is important because it produces only singleton abstracts. Notice, though, that SAP won't tell us much at all about the size of the universe; you will have at most as many singleton abstracts as there are objects in the underlying domain.[27] In fact it's precisely that feature of singleton abstraction that's being exploited by Apogon principles.

New V has been studied extensively, as it can be used to capture a bit of set theory (see Boolos, 1987b; Shapiro and Weir, 1999). It is just a restriction

[27] On finite domains, singleton abstraction could in fact double the size of the domain, but since we only really care about infinite domains, we need not worry about that. For the rest of this section, I will assume we are working with a domain of size at least $\aleph_0$.

of BLV to concepts that aren't universe-sized, allowing for a consistent theory of Fregean extensions. Notice again that New V puts no requirements on the size of the domain.

Singelton V is perhaps the most interesting of our three examples. It is exactly New V for infinite concepts, but acts like SAP on finite concepts. This then means that we get most of our extensions plus our singleton abstracts, (along with a dummy abstract or two). More than that, FV gives us exactly the same sorts of abstracts as Apogon principles, but doesn't dictate how many abstracts there are in total, and thus doesn't dictate a lower bound on the size of the universe.

The importance of the APs I just introduced is that their canonical equivalence relations are one or both of the canonical equivalence relations associated with Apogon principles. We can then compare any given Apogon principle to these three APs to determine whether its canonical equivalence relations determine the size of the domain needed to satisfy that AP. The reason there are three APs is that there are three possibilities for the behaviour of canonical equivalence relations.

There are equivalence relations that can stand alone on the RHS of an AP; there are those that need a size restriction to be satisfiable at all; and there are cases where an AP has two or more canonical equivalence relations. Apogon principles are of the third type, but it is important to look at SAP and New V to ensure that neither coextensionality nor singleton abstraction are setting the minimum domain size. The key here is to recognize that neither SAP, New V, nor SV require a particular size domain for their satisfaction. Another way to interpret this is to say that neither coextensionality, nor singleton abstraction solely determines the minimum size of the domain needed to satisfy an Apogon principle, and neither does their combination.

For most cases, reasoning directly about the relevant canonical equivalence relations will give us the answer we need. Consider an arbitrary Heck principle. We can easily see that the canonical equivalence relation is coextensionality. We can then ask whether the domain needed to satisfy that Heck principle, namely a domain of size κ, is required by coextensionality. The answer is obviously **no**—no extensions are actually provided, yet the Heck

principle requires a domain of size κ for its satisfaction.

Finally, let's have a quick look at our paradigm of acceptable abstraction, and make sure it doesn't get caught in our net. Equinumerosity is the canonical equivalence relation associated with HP (see §5.4 for an explanation), so we can ask whether equinumerosity (or the nature of numbers) requires a domain of a certain size, and whether that is the same size as is required by HP. The answer to both questions is yes, as we know. Note that although HP is only satisfiable on infinite domains, that is because of how equinumerosity works—by carving the world into infinitely many equivalence classes—and not anything else going on on the RHS of HP. Nevertheless, we might want to build infinite satisfiability into the definition of fishiness to rule out cases where equivalence relations exhibit bad behaviour in finite models. Such a move should be unproblematic since we decided back in §6.2 that we need not consider APs that aren't satisfiable on infinite first-order domains.

Our updated definition of fishiness will be the following.

> *An abstraction principle, Σ, is* fishy *iff the size of the smallest infinite first-order domain required for its satisfaction is not determined* solely *by the the canonical equivalence relation or relations associated with Σ, or (equivalently) abstraction principles with only that/those canonical equivalence relations on their right hand sides do not require that particular size of infinite domain for their satisfaction, (modulo size restrictions necessary for satisfiability).*[28]

As we have just seen, that definition classifies Heck principles and Apogon principles as fishy, and HP as not fishy. This leaves us the question of how best to rule out fishy principles.

6.6 Hauling out the Catch

It's all well and good to say we want to rule out APs that don't affect the cardinality of the domain in a particular way. It's harder to find a way to rule

[28]See §5.5 for one possible reason to accept size restrictions necessary for satisfiability.

out fishy principles *en masse*, and this is particularly important if we want to work with systems of APs as with Fine's (2002, Chapt. 4) General Theory of Abstraction, or collections of acceptable principles we've found to be useful as, for example, in the system sketched in chapter 3. In this section I'll look at two ways we might rule out fishy principles based on proposals made by Cook and Linnebo (2018). Before that, I'll summarize some reasons we ought to rule out fishy principles in the first place, and provide a diagnosis of the failure of other proposals to rule out fishiness.

Why Fish Stink

I argued above that it is *ad hoc* or unnatural to accept as true, principles that allow us to arbitrarily set the minimum size of the mathematical universe. That diagnosis of the problem with fishy principles fits well with the definition of fishiness I just gave for the following reason. The definition of fishiness essentially claims that there ought not be an explanatory gap, or at least not a large explanatory gap, between the equivalence relations on the RHS of an AP, and the number of abstracts that AP produces. Similarly, an *ad hoc* statement has little explanatory power, i.e. *ad hoc* stipulations rarely ground robust explanations.

The reason that we ought to ban fishiness then, is that since APs are supposed to be epistemically foundational, allowing us certain kinds of epistemic access to abstract mathematical objects, there ought not be a large explanatory gap between those objects and the equivalence relations they're associated with. That is to say, the reason we ought to ban fishiness is more or less the same the reason I developed the notion of a canonical equivalence relation in the previous chapter. Banning fishiness allows us to rule out at least some of the APs with significant epistemic gaps between their two sides.[29]

[29] Chapter 5 contains a discussion of the general nature of such gaps, and why we ought to be concerned about them from a neo-logicist perspective.

Stability

Recall that I launched my analysis of fishiness from the question of whether the 'right sort' clause of the definition of critical fullness is met by Apogon principles. Here's the definition again.

Definition 5 (Critical Fullness)**.** An AP Σ is *critically full* iff, for each each critical point κ of Σ, any model of Σ of size κ contains κ-many abstracts of the sort characterized by Σ.

Recall too, that adding to the requirement that APs be strongly stable, the requirement that they be critically full, results in an acceptability criterion that Cook and Linnebo call *Heck stability*. Heck stability rules out Heck principles, but not Apogon principles. Finally, recall that there is an even stronger criterion, *conservative stability* that is strictly stronger than Heck stability, but still doesn't rule out Apogon principles.

Definition 2 (Conservative Stability)**.** An AP, Σ, is *conservatively stable* iff it is strongly stable with stabilization point κ and in any model of Σ, there are at least κ-many abstracts of the sort characterized by Σ.

We can now see why these attempts failed. In both cases, we were trying to get enough of the right sort of abstracts, when we should have been trying to ensure we're getting enough abstracts *because they're of the right sort*. The use of 'because' captures the idea that we want there to be the number of abstracts there are to be connected to the nature of the equivalence relations that they're associated with. It also helps enforce the epistemic connection between equivalence relations and abstracts that is a large part of the value of APs.

Before proceeding to my positive proposal, we should consider one more purely formal acceptability criterion. When attempting to rule out Apogon principles model theoretically, Cook and Linnebo (2018, p. 71) suggest monotonicity.

Definition 6 (Monotonicity)**.** An AP, Σ, is *monotonic* iff the equivalence relation $\Phi(X, Y)$ is *intrinsic* where $\Phi(X, Y)$ is *intrinsic* just in case $\Phi(X, Y)$ iff $\Phi^{X \cup Y}(X, Y)$.[30]

This captures the thought that what should matter are only the objects that fall under the concepts quantified over in the AP. The effect is that if an AP is satisfiable on a domain of size κ, if you move to a larger domain at which it's satisfiable, you still have the abstracts from the original domain. Even more simply, if an AP is monotonic, you keep the abstracts you already had/would have had on smaller domains. Note though, that monotonicity is distinct from the usual model theoretic criteria based around notions of conservativeness and stability. Furthermore, monotonicity fails to rule out our paradigmatic example of a bad AP: BLV. So monotonicity on its own won't solve Bad Company. It does, however, rule out Apogon principles and Heck principles, but it also bans size restricted principles like New V.

There are three issues I see with adopting monotonicity as the solution to the problem of fishiness. First, we may not want to rule out New V and similar APs (e.g. for the reasons mentioned at in §5.5). New V does give us *some* sets after all. It seems to me that our solution to Bad Company should be as inclusive as possible while still being able to participate in grounding of the epistemic privilege of acceptable APs.

Second, we don't yet know much about monotonicity. At the end of their paper, Cook and Linnebo have this to say:

> We leave for another occasion the question of whether the bad company problem can be solved by means of the requirement of monotonicity combined with some stability requirement and/or some requirement of fullness. (p. 71)

Not yet knowing how monotonicity will behave in the presence of other acceptability criteria is not a problem in itself, but it does mean that it isn't on better footing than any other new proposal.

[30] I.e. an equivalence relation is intrinsic iff it behaves the same when quantifiers are restricted to objects falling under one or the other of the concepts it's being applied to.

Knowing what we do now about the nature of fishiness, by far the biggest problem with using monotonicity as a way to rule out fishiness is that it bears little connection to the origin of the phenomenon it is being used to rule out. I would even argue that adopting monotonicity solely for the purpose of ruling out Apogon principles is *ad hoc*. That's because monotonicity exploits a feature of Apogon principles—singleton abstraction—that is, at best, tenuously connected to the fishiness of those principles. Even if we adopt monotonicity for independent reasons, it will look like no more than a happy accident that it rules out fishiness.[31]

Ultimately, whether or not we adopt a monotonicity requirement, keeping the formal and epistemic aspects of the neo-logicist project intertwined will be beneficial, and doing so will allow us to rule out fishy APs *because they're fishy*.

The obvious way to modify the definition of conservative stability or critical fullness to capture the explanatory nature of the relationship between canonical equivalence relations and abstracts is simply to add a *not-fishy* clause. For concreteness here is how I would modify the definition of critical fullness.

Definition 7 (Explanatory Fullness)**.** An AP Σ is *explanatorily full* iff, for each critical point κ of Σ, any model of Σ of size κ contains κ-many abstracts of the sort characterized by the canonical equivalence relation(s), $\Phi(X, Y)$, associated with Σ, and the existence of κ-many abstracts can be explained just in terms of $\Phi(X, Y)$ and the original domain.

All I've done is require that the AP Σ not be fishy, so there's a sense in which there is no need to add *not-fishy* clauses to other criteria rather than simply ban fishiness. However, such is the convention, and laying down individual acceptability criteria will allow us to compare and classify those criteria easily and fruitfully. I propose that we call acceptability criteria that include a *not-fishy* clause *explanatory criteria*.

[31]There are several reasons we might adopt monotonicity anyway. One is simply that it represents an intuitively plausible restriction on acceptability. Another is the apparent close connection it bears to certain invariance properties investigated by Cook (2017)—it may help us bridge the gap between invariance and stability/conservativeness.

6.7 Final Thoughts

The most important take away from this chapter (as well as the last) is that we ought not forget the epistemic role abstraction principles need to play if the neo-logicist project is to be successful. More specifically I provided a explanation of the fishiness of certain APs based on the explanatory connection between equivalence relations and abstracts. If that definition is adopted or applied, we will be much closer to a solution to bad company.

Strong stability has been defended as a necessary condition for the acceptability of abstraction principles, and its sufficiency only called into question by the existence of the Apogon principles I looked at here (Cook, 2012, 2017). That suggests that adding an explanatory clause banning fishiness, or even better explanatory fullness, to the definition of strong stability might be just what we need to put Bad Company to bed.

Chapter 7

Conclusion

I've brought together important and closely connected themes in an attempt to shift our approach to neo-logicism in directions I have argued are more promising than many of the approaches that have recently been popular. The last chapter brings two of those themes to the fore. There I argued that, to solve the Bad Company problem, we ought to (re)introduce epistemic thinking into our proposed solutions. This fits well with the neo-logicist goal of grounding mathematical knowledge via abstraction principles (APs). The focus on Bad Company, and the focus on the epistemological goals of neo-logicism are present throughout the dissertation.

Chapter 5 brings out an aspect of the relationship between the equivalence classes on the RHSs of APs and the abstracts identified on the LHSs that is central to my discussion of fishiness and Bad Company and chapter 6. That notion—of a canonical equivalence relation—explicitly brings together metaphysical and epistemic aspects of the neo-logicist understanding of APs. In particular, I appeal to a certain understanding of *content carving*, a popular way of understanding how abstraction works. I also suggests new ways of looking at abstract-identity and Good Company using canonical equivalence relations, opening up space for epistemologically informed approaches to those other problems.

Bad Company is also important in chapters 3 and 4. In the latter we (Darnell and I) argue that HP isn't analytic according to Frege's conception

of analyticity *unless* we adopt a very specific sort of solution to Bad Company that privileges HP above other APs. We also show that such a position requires pulling together threads that have implication for abstract identity and Good Company. More generally, chapter 4 is concerned with establishing the nature of the epistemic privilege attributable to APs. Although our results in that chapter are mostly negative, we also narrow the range of possibilities for the epistemic significance of APs, and show the importance of a holistic understanding of the neo-logicist project.

Chapter 3 is concerned mostly with methodology and metaphysics, though the implications of solutions to Bad Company are discussed at several points. My goal in that chapter is urge neo-logicists who are looking for unified foundations to broaden their approach to include theories other that ZF set theory as the target of neo-logicist reductions. I also argue that we might be better off avoiding looking for reductions of particular theories at all, and sketch a metaphysics that supports that. That metaphysical system is then built upon in the more speculative parts of later chapters.

Chapter 2 takes up another foundational epistemic issue—the status of higher-order logic—from a historical perspective. This is important because without the sharpening of the boundaries between logic and set theory on the one hand, and the recognition that higher-order logic could end up on the 'logic' side of that distinction on the other, there would hardly have been space for neo-logicism to get off the ground at all. I look particularly at the period between 1945 and 1983 because that was the lead up to the birth of neo-logicism, and that period in the history of (neo)-logicism hasn't been extensively discussed.

Each chapter takes or proposes a new direction for neo-logicist research. I hope that some of those new directions are pursued and lead to positive developments for the neo-logicist program.

Appendix A

List of Abstraction Principles

A.1 Single Abstraction Principles

Note: not all of these are satisfiable.

Direction Principle

(dir) $$d(l_1) = d(l_2) \leftrightarrow l_1 // l_2$$

Basic Law V

(BLV) $$\varepsilon F = \varepsilon G \leftrightarrow F \equiv G$$

where $\equiv$ denotes coextensionality.

Hume's Principle

(HP) $$\#F = \#G \leftrightarrow F \approx G$$

where $\approx$ abbreviates a second-order sentence asserting the existence of a bijection between the Fs and Gs.

Singleton Abstraction Principle

(SAP) $$\$F = \$G \leftrightarrow F \cong G$$

where $\cong$ holds if either F and G are singleton concepts, or neither is.

New V

(NewV) $$\varepsilon F = \varepsilon G \leftrightarrow (F \equiv G \vee (F \approx V \wedge G \approx V))$$

Finite Hume's Principle

(FHP) $$\#F = \#G \leftrightarrow ((F < \omega \wedge G < \omega) \wedge F \approx G) \vee (F \geq \omega \wedge G \geq \omega))$$

where $F < \omega$ abbreviates a second-order formula that says F is Dedekind-finite, and $F \geq \omega$ abbreviates a second-order formula that says that F is Dedekind-infinite.

Singleton V

(SV) $$\begin{aligned} \clubsuit F = \clubsuit G \leftrightarrow [&(F \cong G \wedge (F < \omega \vee G < \omega)) \\ &\vee (F \equiv G \wedge (\omega < F < |V| \wedge \omega < G < |V|)) \\ &\vee (F \approx V \wedge G \approx V)] \end{aligned}$$

where $F < |V|$ abbreviates a second-order formula that says F has cardinality strictly less than that of the universe.

Fishy V

(FV) $$\spadesuit F = \spadesuit G \leftrightarrow [((F \approx V \vee G \approx V) \wedge F \cong G) \vee (\neg(F \approx V \vee G \approx V) \wedge F \equiv G))]$$

Hume's V

(HV) $$\begin{aligned} \diamondsuit F = \diamondsuit G \leftrightarrow [&(F \approx G \wedge (F < \omega \vee G < \omega)) \\ &\vee (F \equiv G \wedge (\omega < F < |V| \wedge \omega < G < |V|)) \\ &\vee (F \approx V \wedge G \approx V)] \end{aligned}$$

A.2 Abstraction Principle Schemata

Nuisance Principles

(NP) $\ddagger F = \ddagger G \leftrightarrow (\{x : Fx\} \backslash \{y : Gy\}) \cup (\{y : Gy\} \backslash \{x : Fx\}) < \omega$

which says that the symmetric difference of the Fs and Gs is finite. (Set notation has its usual meaning.)

Distraction Principles

(DP) $\partial F = \partial G \leftrightarrow [(\Upsilon(F) \wedge \Upsilon(G)) \vee (F \equiv G)]$

where Υ is a cardinality property

Heck Principles

(Heck) $\eth F = \eth G \leftrightarrow ((F \equiv G) \vee A_\kappa)$

Where 'A_κ' is a purely logical second-order sentence that asserts that the universe is at least size κ.

Apogon Principles

(Σ^+) $\eth^+ F = \eth^+ G \leftrightarrow ((\mathcal{R}(\Sigma) \wedge F \cong G) \vee (\neg\mathcal{R}(\Sigma) \wedge F \equiv G))$

The *Ramsification* of an AP Σ, denoted $\mathcal{R}(\Sigma)$, is the sentence obtained when the abstraction operators are replaced with appropriate variables bound by existential quantifiers.

Appendix B

Definitions of Model-theoretic Criteria

B.1 Stability & Conservativity

Definition 1 (Purely Logical)**.** An AP, Σ, is *purely logical* iff the equivalence relation on the RHS is expressible in the language of pure higher-order logic.

Definition 2 (κ-Satisfiable)**.** Σ is κ-satisfiable iff it is satisfiable on a domain of size κ.

Definition 3 (Unbounded)**.** Σ is *unbounded* iff it is κ-satisfiable for an unbounded sequences of cardinals κ.

Definition 4 (Stable)**.** Σ is *stable* iff there is a κ such that it is λ-satisfiable for all $\lambda \geq \kappa$. (The least such κ is the *stabilization point*.)

Definition 5 (Strongly Stable)**.** Σ is *strongly stable* iff there is a κ such that it is λ-satisfiable iff $\lambda \geq \kappa$.

Definition 6 (Weakly Conservative)**.** Σ is *weakly conservative* iff for any theory T and sentence ϕ, IF $T^{P(x)}, \Sigma \models \phi^{P(x)}$, THEN $T \models \phi$.
OR
Σ is *weakly conservative* iff for every $\mathcal{L}$-model $\mathcal{M}$, there is an $\mathcal{L}_{\S}$-model $\mathcal{N}$ sth $\mathcal{M} \subseteq \mathcal{N}$ and $\mathcal{N} \models \Sigma$.

Definition 7 (Strongly Conservative)**.** Σ is *strongly conservative* iff for and theory T, and sentence ϕ, IF $T^{\neg\exists F(x=\S F)}, \Sigma \models \phi^{\neg\exists F(x=\S F)}$, THEN $T \models \phi$.
OR
Σ is *strongly conservative* iff for every $\mathcal{L}$-model $\mathcal{M}$, there is an $\mathcal{L}_\S$-model $\mathcal{N}$ sth (i) $\mathcal{M} \subseteq \mathcal{N}$, (ii) $\mathcal{N} \models \Sigma$, and (iii) $\mathcal{N} \setminus \mathcal{M}$ contains all and only the §-abstracts.

Definition 8 (Irenic)**.** Σ is *weakly (resp. strongly) irenic* iff it is weakly (strongly) conservative and co-satisfiable with any other weakly (strongly) conservative AP.

Definition 9 (Critical Point)**.** A cardinal κ is a *critical point* of an AP Σ iff Σ is κ-satisfiable, and there is a $\gamma < \kappa$ sth, for all λ sth $\gamma \leq \lambda < \kappa$, Σ is not λ-satisfiable.

Definition 10 (Critically Full)**.** Σ is *critically full* iff, for each each critical point κ of Σ, any model of Σ of size κ contains κ-many abstracts of the sort characterized by Σ.

Definition 11 (Heck-*)**.** Σ is *Heck conservative/irenic/stable* iff it is strongly conservative/irenic/stable and critically full.

Definition 12 (Conservatively Stable)**.** Σ is *conservatively stable* iff it is strongly stable with stabilization point κ and in any model of Σ, there are at least κ-many abstracts of the sort characterized by Σ.

B.2 Monotonicity

Definition 13 (Cardinality Montonic)**.** Σ is *cardinality monotonic* iff for any κ, γ, $\kappa \leq \gamma$ where Σ is both κ- and γ-satisfiable, the size of the class of Σ-abstracts on a domain of size κ is no larger than the class of Σ-abstracts on a domain of size γ.

Definition 14 (Monotonic)**.** Σ is *monotonic* iff the equivalence relation $\Phi(X, Y)$ is *intrinsic* where $\Phi(X, Y)$ is *intrinsic* just in case $\Phi(X, Y)$ iff $\Phi^{X \cup Y}(X, Y)$.

Appendix C

Copyright Permission

I, Eamon Darnell, hereby give my permission for significant material from the manuscript co-authored with Aaron R. Thomas-Bolduc entitled "Is Hume's Principle Analytic?", to be included in Thomas-Bolduc's PhD thesis, "New Directions for Neo-Logicism", and for that material to be archived in the institutional repository at the University of Calgary and the Library and Archives Canada.

Signed:

Date:

Bibliography

Aczel, Peter. 1988. *Non-wellfounded Sets*. CSLI Lecture Notes 14. Stanford, CA: CSLI.

Aczel, Peter and Micheal Rathjen. 2010. Constructive Set Theory . Book draft.

Antonelli, G. Aldo. 2010. Notions of Invariance for Abstraction Principles. *Philosophia Mathematica* 18(3): 276–292.

Benacerraf, Paul. 1960. Logicism, Some Considerations. Ph.D. thesis, Princeton.

Benacerraf, Paul. 1965. What Numbers Could Not Be. *Philosophical Review* 74: 47–73. Reprinted in Benaceraff and Putnam (1983), pp. 272–294.

Benacerraf, Paul. 1973. Mathematical Truth. In Benacerraf and Putnam (1983), 403–420. First published in the *Journal of Philosophy*, vol. 70, pp. 661-680.

Benacerraf, Paul and Hilary Putnam, eds. 1983. *Philosophy of Mathematics: Selected Readings*. Cambridge: Cambridge University Press, 2nd ed.

Benci, Vieri and Mauro Di Nasso. 2003. Numerosities of labelled sets: a new way of counting. *Advances in Mathematics* 173: 50–67.

Blanchette, Patricia. 2014. The Frege-Hilbert Controversy. In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Metaphysics Research Lab, Stanford University, Spring 2014 ed. `https://plato.stanford.edu/archives/spr2014/entries/frege-hilbert/`.

Boccuni, Francesca. 2011. On the Consistency of a Plural Theory of Frege's Grundgesetze. *Studia Logica* 97(3): 329–345.

Boccuni, Francesca. 2013. Plural Logicism. *Erkenntnis* 78(5): 1051–1067.

Boolos, George. 1975. On Second-Order Logic. In Boolos (1998), 37–53. First published in *The Journal of Philosophy* vol. 72 pp. 509–527.

Boolos, George. 1984. To Be is to Be a Value of a Variable (or to Be Some Values of Some Variables). In Boolos (1998), 54–72. First published in the *Journal of Philosophy*, vol. 81 pp. 430–449.

Boolos, George. 1987a. On the Consistency of Frege's *Foundations of Arithmetic*. In Boolos (1998), 183–201. Originally in J.J. Thomson (ed) *On Being and Saying* Cambridge: MIT Press, pp. 3–20.

Boolos, George. 1987b. Saving Frege from Contradiction. In Boolos (1998), 171–182. First published in the *Proceedings of the Aristotelian Society Volume 87*, pp. 137–151.

Boolos, George. 1990. The Standard Equality of Numbers. In Boolos (1998), 202–219. Orginally in *Meaning and Method* Boolos (ed) Cambridge: Cambridge University Press. pp. 261–277.

Boolos, George. 1993. Whence the Contradiction? In Boolos (1998). Originally published in *Aristotelian Society Supplementary Volume 67* pp. 213–233.

Boolos, George. 1994. 1879? In Boolos (1998), 237–254. Originally in Clark and Hale (eds), *Reading Putnam*, Oxford: Blackwell,.

Boolos, George. 1995. Frege's Theorem and the Peano Postulates. In Boolos (1998), 291–300. First published in *The Bulletin of Symbolic Logic*, vol. 1, pp. 317–326.

Boolos, George. 1996. On the Proof of Frege's Theorem. In Boolos (1998), 275–290. First published in Morton and Stich (eds) *Benacerraf and His Critics*, Oxford: Blackwell.

Boolos, George. 1997. Is Hume's Principle Analytic? In Boolos (1998), 301–314. First published in R. Heck Jr. (ed), *Logic, Language, and Thought*, Oxford: Oxford University Press.

Boolos, George. 1998. *Logic, Logic, and Logic*. Cambridge, MA: Havard University Press.

Burgess, John. 1984. Review of Crispin Wright's *Frege's Conception of Numbers as Objects*. *The Philosophical Review* 93(4): 638–640.

Burgess, John. 2005. *Fixing Frege*. Princeton, NJ: Princeton University Press.

Carnap, Rudolf. 1931. Symposium on the Foundations of Mathematics 1. The Logicist Foundations of Mathematics. In Benacerraf and Putnam (1983), 41–52. Erna Putnam and Gerald J. Massey (trans).

Church, Alonzo. 1962. Mathematics and Logic. In *Logic, Methodology and Philosophy of Science. Proceedings of the 1960 International Congress.*, eds. Ernest Nagel, Patrick Suppes, and Alfred Tarski, 181–186. Stanford, CA: Stanford University Press.

Cook, Roy T. 2012. Conservativeness, Stability, and Abstraction. *British Journal of Philosophy of Science* 63: 673–696.

Cook, Roy T. 2017. Abstraction and Four Kinds of Invariance (Or: What's So Logical About Counting?). *Philosophia Mathematica* 25(1): 3–25.

Cook, Roy T. and Philip A. Ebert. 2005. Abstraction and Identity. *Dialectica* 59(2): 121–139.

Cook, Roy T. and Øystein Linnebo. 2018. Cardinality and Acceptable Abstraction. *Notre Dame Journal of Formal Logic* .

Cruz-Filipe, Luís and Fernando Ferreira. 2015. The Finitistic Consistency of Heck's Predicative Fregean System. *Notre Dame Journal of Formal Logic* 56(1): 61–79.

Demopoulos, William, ed. 1997. *Frege's Philosophy of Mathematics*. Harvard, Mass.: Harvard University Press, 1st paperback ed.

Dudman, Victor H. 1976. From Boole to Frege. In *Studies on Frege I: Logic and Philosophy of Mathematics*, ed. Matthais Schirn, Problemata, 109–138. Stuttgard-Bad Cannstatt: Friedrich Frommann.

Dummett, Micheal. 1991. *Frege: Philosophy of Mathematics*. Cambridge, MA: Harvard University Press.

Ebels-Duggan, Sean. 2015. The Nuisance Principle in Infinite Settings. *Thought: A Journal of Philosophy* 4(4): 263–268.

Eilenberg, Samuel and Saunders Mac Lane. 1942. Group extensions and homology. *Annals of Mathematics* 43(4): 757–831. URL `http://www.jstor.org/stable/1968966`.

Eilenberg, Samuel and Saunders Mac Lane. 1945. General Theory of Natural Equivalences. *Transactions of the American Mathematical Society* 58: 231–294.

Ferreira, Fernando and Kai F. Wehmeier. 2002. On the Consistency of the Δ^1_1-CA Frament of Frege's *Grundgesetze*. *Journal of Philosophical Logic* 31: 301–311.

Fine, Kit. 2002. *The Limits of Abstraction*. Oxford: Oxford University Press.

Frege, Gottlob. 1879. *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle: Nebert. Facsimile published by Hachette Livre.

Frege, Gottlob. 1884. *Die Grundlagen der Arithmetik*. Breslau: Koebner.

Frege, Gottlob. 1893. *Grundgesetze der Arithmetik. Begriffsschriftlich abgeleitet*, vol. I. Jena: Pohler.

Frege, Gottlob. 1903. *Grundgesetze der Arithmetik. Begriffsschriftlich abgeleitet*, vol. II. Jena: Pohler.

Frege, Gottlob. 1980. *The Foundations of Arithmetic*. Evanston: Northwestern University Press, second revised ed.

Frege, Gottlob. 1997a. Grundgesetze der Arithmetic, Volume I. In *The Frege Reader*, ed. Michael Beaney. Oxford: Blackwell Publishing.

Frege, Gottlob. 1997b. Logic in mathematics. In *The Frege Reader*, ed. Michael Beaney. Oxford: Blackwell Publishing.

Frege, Gottlob. 2013. *Basic Laws of Arithmetic. Derived using concept-script.*, vol. I & II. Oxford: Oxford University Press. P. Ebert and M. Rossberg with C. Wright (trans, eds.).

Geach, Peter. 1955. Class and Concept. In Klemke (1968), 284–294. First published in *The Philosophical Review*, vol. 64, pp. 561–570.

Geach, Peter. 1956. On Frege's Way Out. In Klemke (1968), 502–504. First published in *Mind*, vol. 65, pp. 408–9.

Geach, Peter. 1976. Critical Notice of Micheal Dummett, *Frege: Philosophy of Language*, 1973. *Mind* 85(339): 436–449.

Hale, Bob. 1987. *Abstract Objects*. Oxford: Basil Blackwell.

Hale, Bob. 1997. *Grundlagen* § 64. In Hale and Wright (2001). First published in *Proceedings of the Aristotelian Society* 97.

Hale, Bob. 2011. The Bearable Lightness of Being. *Axiomathes* 21(4): 597–597.

Hale, Bob. 2015. Second-order Logic: Properties, Semantics, and Existential Commitments. *Synthese* Published online first.

Hale, Bob and Crispin Wright. 2000. Implicit Defininition and the A Priori. In Hale and Wright (2001), 117–152. First published in P. Boghossian and C. Peacocke (eds.) *New Essays on the A Priori*, Oxford: Clarenden.

Hale, Bob and Crispin Wright. 2001. *The Reason's Proper Study: Essays Towards a Neo-Fregean Philosophy of Mathematics*. Oxford: Clarendon Press.

Hazen, Allen. 1985. Review of Crispin Wright's *Frege's Conception of Numbers as Objects*. *Australasian Journal of Philosophy* 63(2): 251–254.

Heck, Richard. 1996. The Consistency of Predicative Fragments of Frege's *Grundgesetze der Arithmetik*. *History and Philosophy of Logic* 170: 209–220.

Heck, Richard. 2011a. Finitide and Hume's Principle. In Heck (2011b), 237–266. First published in the *Journal of Philosophical Logic*, vol. 25, pp.589–617.

Heck, Richard. 2011b. *Frege's Theorem*. Oxford: Oxford University Press.

Heck, Richard. 2011c. Frege's Theorem: An Overview. In Heck (2011b), 1–39.

Heck, Richard. 2011d. On the Consistency of Second-order Contextual Definitions. In Heck (2011b), 227–236. First published in *Noûs* 26, 1992, pp. 491–494.

Hempel, Carl G. 1945. On the Nature of Mathematical Truth. In Benacerraf and Putnam (1983), 377–393. Orginally in *The American Mathematical Monthly*, vol. 52, pp. 543–556.

Hewitt, Simon Thomas. forthcoming. Tuples All the Way Down? *Thought: A Journal of Philosophy* .

Horsten, Leon and Øystein Linnebo. 2016. Term Models for Abstraction Principles. *Journal of Philosophical Logic* 45(1): 1–23.

Jech, Thomas J. 1971. *Lectures in Set Theory with Particular Emphasis on the Method of Forcing*. Lecture Notes in Mathematics. Heidelberg: Springer.

Klemke, E. D., ed. 1968. *Essays on Frege*. Urbana, Chicago, IL: University of Illinois Press.

Komjáth, Péter and Vilmos Totik. 2008. Ultrafilters. *The American Mathematical Monthly* 115: 33–44.

Landry, Elaine and Jean-Pierre Marquis. 2005. Categories in Context: Historical, Foundational, and Philosophical. *Philosophia Mathematica* 13(1): 1–43.

Lawvere, F. William. 1963. Functorial Semantics of Algebraic Theories and Some Algebraic Problems in the Context of Functorial Semantics of Algebraic Theories. Ph.D. thesis, Columbia University.

Leach-Krouse, Graham. 2017. Structural-Abstraction Principles. *Philosophia Mathematica* 25(1): 45–72.

Linnebo, Øystein. 2011. Some Criteria for Acceptable Abstraction. *Notre Dame Journal of Formal Logic* 52(3): 331—338.

Linnebo, Øystein. 2017. Plural quantification. In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Metaphysics Research Lab, Stanford University, summer 2017 ed. `https://plato.stanford.edu/archives/sum2017/entries/plural-quant/`.

Linnebo, Øystein. 2018. *Thin Objects: An Abstractionist Account*. Oxford: Oxford University Press. Forthcoming.

Linnebo, Øystein and Richard Pettigrew. 2011. Category Theory as an Autonomous Foundation. *Philosophia Mathematica* 19(3): 227–254.

Linnebo, Øystein and Gabriel Uzquiano. 2009. Which Abstraction Principles are Acceptable? Some Limmative Results. *British Journal for the Philosophy of Science* 60: 293–252.

Logan, Shay Allen. 2015. Abstractionist Categories of Categories. *Review of Symbolic Logic* 8(4): 705–721.

Logan, Shay Allen. 2017. Categories for the Neologicist. *Philosophia Mathematica* 25(1): 26–44.

MacBride, Fraser. 2003. Speaking with Shadows: A Study of Neo-Logicism. *British Journal for the Philosophy of Science* 54: 103–163.

Mancosu, Paolo. 2009. Measuring the Size of Infinite Collections of Natural Numbers: Was Cantor's Theory of Infinite Number Inevitible? *The Review of Symbolic Logic* 2: 612–646.

Mancosu, Paolo. 2015. In Good Company? On Hume's Principle and the Assignment of Numbers to Infinite Concepts. *The Review of Symbolic Logic* 8(2): 370–410.

Mancosu, Paolo. 2016. *Abstraction and Infinity*. New York, Oxford: Oxford University Press.

Marquis, Jean-Pierre. 2015. Category theory. In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Metaphysics Research Lab, Stanford University, winter 2015 ed.

Parsons, Charles. 1965. Frege's Theory of Number. In Demopoulos (1997), 182–207. First published in Max Black (ed.), *Philosophy in America*, pp. 180–203.

Parsons, Terence. 1987. On the Consistency of the First-Order Portion of Frege's Logical System. *Notre Dame Journal of Formal Logic* 28(1): 161–169.

Payne, Jonathan. 2013. Abstraction Relations Need Not Be Reflexive. *Thought: A Journal of Philosophy* 2(2): 137–147.

Quine, W. V. O. 1955. On Frege's Way Out. In Klemke (1968), 485–501. First published in *Mind*, vol. 64, pp. 145–159.

Quine, W. V. O. 1970. *Philosophy of Logic*. Englewood Cliffs, NJ: Prentice-Hall, 1st paperback ed.

Ramsey, F. P. 1926. The Foundations of Mathematics. *Proceedings of the London Mathematical Society* s2-25(1): 338–384.

Rayo, Augustín. 2013. *The Construction of Logical Space*. Oxford: Oxford University Press.

Resnik, Micheal. 1980. *Frege and the Philosophy of Mathematics*. Ithaca, London: Cornell University Press.

Rossberg, Marcus and Philip A. Ebert. 2007. What is the Purpose of Neologicism? *Traveaux de Logique* 18: 33–61.

Schirn, Matthias. 2006. Hume's Principle and Axiom V Reconsidered: Critical Reflections on Frege and His Interpreters. *Synthese* 148(1): 171–227.

Shapiro, Stewart. 2002. *Foundations without Foundationalism. A Case for Second-order Logic*. New York, NY: Clarendon Press, 2nd ed. First published in 1991.

Shapiro, Stewart and Alan Weir. 1999. New V, ZF and Abstraction. *Philosophia Mathematica* 3(7): 293—321.

Sher, Gila. 1991. *The Bounds of Logic: A Generalized Viewpoint*. Cambridge,MA: MIT Press.

Simons, Peter M. 1987. Frege's Theory of Real Numbers. In Demopoulos (1997), 358–385. First published in *History and Philosophy of Logic*, vol. 8, pp. 25–44.

Sobocinski, B. 1949. L'Analyse de l'antinomie russellienne par Leśniewski: IV. La correction de Frege. *Methodos* 1: 220–228.

Studd, James P. 2015. Abstraction Reconceived. *British Journal for the Philosophy of Science* 67(2): 579–615.

Univalent Foundations Program. 2013. *Homotopy Type Theory: Univalent Foundations of Mathematics*. Institute for Advanced Study: `https://homotopytypetheory.org/book`.

van Heijenoort, Jean. 1967. *From Frege to Gödel*. Cambridge: Harvard University Press.

Walsh, Sean. 2012. Comparing Peano Arithmetic, Basic Law V, and Hume's Principle. *Annals of Pure and Applied Logic* 163: 1679–1709.

Walsh, Sean. 2016a. Fragments of Frege's *Grundgesetze* and Gödel's Constructible Universe. *The Journal of Symbolic Logic* 81(2): 605–628.

Walsh, Sean. 2016b. The Strenth of Abstraction with Predicative Comprehension. *Bulletin of Symbolic Logic* 22(1): 105–120.

Walsh, Sean and Sean Ebels-Duggan. 2015. Relative Categoricity and Abstraction Principles. *Review of Symbolic Logic* 8: 572–606.

Weir, Alan. 2003. Neo-Fregeanism: An Embarassment of Riches. *Notre Dame Journal of Formal Logic* 44(1): 13–48.

Wenmackers, Sylvia and Leon Horsten. 2013. Fair infinite lotteries. *Synthese* 190: 37–61.

Whitehead, Alfred North and Bertrand Russell. 1927. *Principia Mathematica*. Cambridge: Cambridge University Press, 2nd ed. Three Volumes.

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

Wright, Crispin. 1997. On the Philosophical Significance of Frege's Theorem. In Hale and Wright (2001), 272–306. First published in R. Heck Jr. (ed), *Logic, Language, and Thought*, Oxford: Oxford University Press.

Wright, Crispin. 1998. Response to Dummett. In Hale and Wright (2001), 256–271. Originally in M. Schirn (ed.) *Philosophy of Mathematics Today*, Oxford: Clarendon Press.

Wright, Crispin. 1999. Is Hume's Principle Analytic? In Hale and Wright (2001), 307–332. Originally in the *Notre Dame Journal of Formal Logic, vol. 40, pp. 6–30*.