The Review of Symbolic Logic
Volume 16, Number 2, June 2023

THE POTENTIAL IN FREGE'S THEOREM

WILL STAFFORD
Department of Logic and Philosophy of Science, University of California

Abstract. Is a logicist bound to the claim that as a matter of analytic truth there is an actual infinity of objects? If Hume's Principle is analytic then in the standard setting the answer appears to be yes. Hodes's work pointed to a way out by offering a modal picture in which only a potential infinity was posited. However, this project was abandoned due to apparent failures of cross-world predication. We re-explore this idea and discover that in the setting of the potential infinite one can interpret first-order Peano arithmetic, but not second-order Peano arithmetic. We conclude that in order for the logicist to weaken the metaphysically loaded claim of necessary actual infinities, they must also weaken the mathematics they recover.

§1. Introduction.

1.1. Potentially infinite models. In the nonmodal setting, Frege (1893; Heck, 1993) essentially proved that second-order Peano arithmetic, PA^2, is interpretable in the theory HP^2, which consists of the Second-order Comprehension Schema and Hume's Principle:

$$\forall X, Y(\#X = \#Y \Leftrightarrow \exists \textit{ bijection } f : X \to Y). \tag{HP}$$

Hume's Principle characterizes the cardinality operator #, read 'the number of' or 'octothorpe', as a type-lowering function that takes equinumerous second-order objects to the same first-order object. This definition can be motivated in the finite case by examples such as checking one has the same number of knives and forks by setting them out in pairs. Formally, Frege's result is:

Theorem 1.1 Frege's Theorem. *There is a translation from the language of* PA^2 *to the language of* HP^2 *that interprets* PA^2 *in* HP^2.

The formal definition of the theories mentioned here can be found in Appendix A. Frege's Theorem has traditionally been regarded as philosophically important because it is supposed to show that we can derive all arithmetical theorems from an epistemically innocent system. This requires that Hume's Principle is analytic. However, on the usual semantics, Hume's Principle is only true on domains with at least a countable infinity of objects. This commits logicists like Frege to the analytic existence of an actual infinity of objects (Boolos, 1998, pp. 199, 213, 233; Hale & Wright, 2001, pp. 20, 292, 309; Cook, 2007, p. 7).

A commitment to a *potential* infinity, in contrast, isn't a commitment to how many things there actually are, just how many are possible. This is a much safer area in which

Received: August 31, 2019.
2020 *Mathematics Subject Classification*: 03-02, 00A30, 03B16, 03B45.
Key words and phrases: logicism, potential infinity, modal logic.

doi:10.1017/S1755020320000349

to make analytic claims. Here we show that some but not all of the mathematics of the actual infinite is recoverable in the setting of the potential infinite. And so, to avoid problematic ontological commitments the logicist must also weaken the mathematics they recover.

To do this we must decide how to represent Hume's Principle. Below we will define 'the number of' operator # in a semantic manner. However, we are convinced that this is simply a convenience and we can think of our models as defining # as satisfying Hume's Principle with the additional criteria that this function is rigid across worlds. An axiomatization would consist of the following modification of Hume's Principle:

$$\Box\forall X, Y(\#X = \#Y \Leftrightarrow \exists\, bijection\ f : X \to Y),$$

plus a principle to rigidify the # operator. This would require working in a hybrid modal logic where worlds could be saved and recalled such as Williamson (2013, 370).[1] However, we leave the details of this approach for future work. As the modification is so minimal, the move to the potentially infinite doesn't undermine the justifications offered for Hume's Principle. The syntactic priority thesis can still be argued for as we can identify the behavior of terms in a modal setting as well as in a nonmodal setting. Similarly if we think that abstraction principles offer implicit definitions then this justification works as well in the modal setting.

The rigidity of the octothorpe is important for the success of the project here. However, by assuming that it is rigid we are presuming that 'the number of' operator is rigid. Whether this is the case in natural language is an empirical question (e.g. Stanley, 1997). We do not address this issue here, but two things are worth noting. First the question of the rigidity of 'the number of' is not the same question as e.g. whether the number of planets varies between worlds. This is because we do not apply the operator to predicates but rather to sets which do not vary their membership across worlds. The second is that this setting does rule out the possibility of multiple different number structures in the different worlds, e.g. the numbers being von Neumann ordinals in one world and Zermelo ordinals in another. This means that a certain kind of referential indeterminacy which has a prominent place in philosophy of mathematics cannot be addressed in this setting as we have presumed against it (Benacerraf, 1965; Button & Walsh, 2018, chap. 2).

To set up our result, we define a set of second-order Kripke models, which we will call *potentially infinite models*. This idea comes from Hodes (1990, p. 379), although he does not place exactly these constraints on the accessibility relation. We want the models to be nearly linear sequences of worlds (if there are two worlds neither of which accesses the other, there is a third world they both access), where later worlds are possible from the perspective of earlier worlds but not the other way around. Each of these worlds should contain only a finite number of objects as we are assuming actual infinities are impossible, and the number of objects should increase from one world to the next. Each world will have its own second-order domain, which as the worlds are finite, will be the full powerset. The octothorpe will implement Hume's Principle by taking sets of the same cardinality to a unique object and this

[1] For those familiar with hybrid systems the axioms needed is $\uparrow \Box\forall X, y \downarrow [\#X = y \to \Box\# X = y]$. However, this will not play a role in what follows.

https://doi.org/10.1017/S1755020320000349 Published online by Cambridge University Press

object will not change from one world to the next. We define the models formally as follows:

DEFINITION 1.2. *A potentially infinite (PI) model is a quadruple* $\mathcal{M} = \langle W, R, D, I \rangle$ *in the modal signature with second-order quantification and with* # *and* $\boldsymbol{a}$ *as the only nonlogical symbols, such that the following conditions are met:*

1.2.1. *W is countably infinite and R is a directed partial order,*[2]
1.2.2. *the first-order domain of w, written* $D(w)$*, is nonempty and finite for all* $w \in W$*,*
1.2.3. *for each* $n \geq 1$*, the range of the second-order n-ary relational quantifiers at w is* $\mathcal{P}(D(w)^n)$ *consisting of all subsets of the nth Cartesian power* $(D(w))^n$ *of* $D(w)$*,*
1.2.4. *if* $w, s \in W$ *such that* $R(w, s)$ *and* $w \neq s$*, then* $D(w) \subsetneq D(s)$*,*
1.2.5. *the function* $\boldsymbol{a} : \omega \to D$ *(where D is* $\bigcup_{w \in W} D(w)$*) assigns to each number n a distinct element* $\boldsymbol{a}_n$ *in one of the first-order domains, and for all* $w \in W$*, the cardinality of X is n if and only if* $\#X = \boldsymbol{a}_n$ *at w. More formally, for* # *and all w the interpretation function is defined as follows:* $I(\#, w) = \{\langle X, \boldsymbol{a}_{|X|} \rangle \mid \exists s \in W \ X \in \mathcal{P}(D(s))\}$.

REMARK 1.3. *Three brief remarks on this definition:*

First, conditions 1.2.1–4 define a PI model as a directed partial order of ever-increasing finite domains. This means that if we have several objects existing in different possible worlds we can always move to a world where they all exist.

Second, condition 1.2.5 defines the cardinality operator # *using metatheoretic cardinality* $|X|$*. It is sufficient for Hume's Principle to hold that* # *picks-out cardinality, and so condition 1.2.5 ensures that all potentially infinite models are models of Hume's Principle. One reason we need* $\mathcal{P}(D(w)^2)$ *from 1.2.3 is because the quantifier over graphs of functions in Hume's Principle ranges over this set.*

Third, condition 1.2.5 also ensures that the interpretation of the octothorpe is rigid. That is, the octothorpe is interpreted as the same relation at every world. Because of this nothing will be lost if we write $\#X = x$ *and don't specify the world of evaluation. In fact, while we define* $\#X$ *using the* $\boldsymbol{a}_i$*'s, we could have instead simply defined it as rigid and satisfying Hume's Principle and this along with directedness would ensure the* $\boldsymbol{a}_i$*'s exist.*

This definition can obscure the simplicity of the idea here, as such it helps to give several examples. The simplest potentially infinite model we can construct is the following:

EXAMPLE 1.4. *The minimal potentially infinite model is* $(\omega, \leq, D, I)$ *where* $D(\boldsymbol{n}) = \{\mathbf{0}, \ldots, \boldsymbol{n}\}$ *and the interpretation function I interprets octothorpe as cardinality in the metalanguage.*[3] *That is,* $I(\#, w)(X) = \boldsymbol{n}$ *if and only if* $|X| = \boldsymbol{n}$*. The minimal model is illustrated in Figure 1A. When working with such a model we see that a number can be missing from a world even if a set of that cardinality is present. So* $I(\#, \mathbf{1})(\{\mathbf{0}\}) = \mathbf{1}$ *and* $\mathbf{1} \in D(\mathbf{1})$*, but* $I(\#, \mathbf{1})(\{\mathbf{0}, \mathbf{1}\}) = \mathbf{2}$ *and* $\mathbf{2} \notin D(\mathbf{1})$ *even though* $\{\mathbf{0}, \mathbf{1}\} \subseteq D(\mathbf{1})$.

A less simple but similarly elementary model makes use of the nonempty finite subsets of the natural numbers. This model helps illustrate a nonlinear *R* relation:

[2] An order *R* is directed if for all $w, s \in W$ there exists an $t \in W$ such that $R(w, t)$ and $R(s, t)$.
[3] I will use bold face numbers for the numbers in the metalanguage.

https://doi.org/10.1017/S1755020320000349 Published online by Cambridge University Press

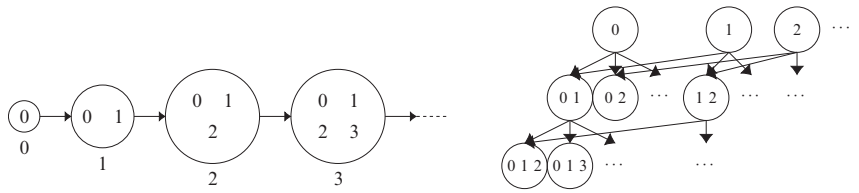


Fig. 1. (A) The minimal model and (B) The subset model. Examples of potentially infinite models.

EXAMPLE 1.5. *Let the subset model be* $(\mathcal{P}(\omega)^{<\omega} - \{\varnothing\}, \subseteq, D, I)$ *where* $D(X) = X$ *and again the octothorpe is cardinality. The subset model is illustrated in Figure 1B. Note that if we have worlds* $X_0, \ldots, X_n$ *we can always find an accessible world whose domain is* $\bigcup_{i=0}^{n} X_i$. *For example,* $\{\mathbf{0}, \mathbf{1}\}, \{\mathbf{3}\}, \{\mathbf{100}, \ldots, \mathbf{200}\}$ *are all finite subsets of the natural numbers, none of which access each other, however, their union* $\{\mathbf{0}, \mathbf{1}, \mathbf{3}, \mathbf{100}, \ldots, \mathbf{200}\}$ *is also a world, which they all access.*

It is easy to generate unintended models from these two cases. Using the minimal model, for example, we can define the **3-0** swap model:

EXAMPLE 1.6. *The* ***3-0*** *swap model takes* ***0*** *and* ***3*** *in the domain of the minimal model and switches them around. So* $D(\mathbf{0}) = \{\mathbf{3}\}, D(\mathbf{1}) = \{\mathbf{3}, \mathbf{1}\}, D(\mathbf{2}) = \{\mathbf{3}, \mathbf{1}, \mathbf{2}\}, D(\mathbf{3}) = \{\mathbf{3}, \mathbf{1}, \mathbf{2}, \mathbf{0}\}$ *and then for all* $\boldsymbol{n} \geq \mathbf{3}$, *we have that* $D(\boldsymbol{n})$ *exactly as it is in the minimal model.*

These models should help illustrate the intuition behind the potentially infinite models. They will also be helpful when we need counterexamples to claims later in the paper.

We can now define satisfaction for potentially infinite models using a standard semantics for quantified modal logic, such as in Fitting & Mendelsohn (1998). Three things to note first: (1) Our quantifiers are actualist, but free variables may be assigned to objects in any world. (2) Set variables are interpreted rigidly across worlds. That is the membership of a set doesn't change depending on the world. (3) To simplify the notation, instead of variable assignments, we work as though we had a rigid name for every object in the models. Recall that $\mathcal{M}, w \vDash \varphi$ means that given any replacement of free variables with the added constants we evaluate φ as true in $\mathcal{M}$ at world w. With this in place, the notion of potentially infinite models induces a natural validity relation, which we define as follows:

DEFINITION 1.7. *We say that* φ *is true in all potentially infinite models, or* $\vDash_{\mathsf{PI}} \varphi$, *if for all potentially infinite models* $\mathcal{M}$ *and worlds* $w \in W$ *we have* $\mathcal{M}, w \vDash \varphi$. *We define* $\varphi \vDash_{\mathsf{PI}} \psi$ *as for all models* $\mathcal{M}$ *and worlds* $w \in W$, *if* $\mathcal{M}, w \vDash \varphi$ *then* $\mathcal{M}, w \vDash \psi$.

The consequence relation here is defined locally rather than globally (Fitting & Mendelsohn, 1998, p. 21). This is because the deduction theorem holds for the local consequence relation but not the global one (Fitting & Mendelsohn, 1998, p. 23).

1.2. Main results. We will now state our two main results which together show that we can interpret the first-order theories of first-order Peano arithmetic PA^1 and first-order true arithmetic TA^1, but not the second-order theories of second-order Peano arithmetic PA^2 and second-order true arithmetic TA^2, in theories defined in

https://doi.org/10.1017/S1755020320000349 Published online by Cambridge University Press

terms of potentially infinite models. A deductive theory for second-order modal logic with rigid operators would be unwieldy and the complications caused by it would be likely to obscure the insights provided by the Kripke semantics. Hence, we leave development of a deductive theory for future work. We can define a theory just in terms of the potentially infinite models. This theory will be stronger than anything we could produce deductively because it does not admit nonstandard models of the natural numbers. Because of this we will call it the external theory of the potentially infinite or E_{PI}:

$$\mathsf{E}_{\mathsf{PI}} = \{\varphi \mid \vDash_{\mathsf{PI}} \varphi\}. \tag{1}$$

To capture something closer to what can be deduced from the models we need to use the model-theoretic validity relation defined above, relativized to a weak metatheory. The theory ACA_0 is a subsystem of PA^2 which only has comprehension for first-order formulas. More information about this theory can be found in Appendix A. Since we can code finite sets of natural numbers as natural numbers in ACA_0, we can define the property of being a potentially infinite model in this theory, along with the associated validity notion $\vDash_{\mathsf{PI}}$. This gives us the internal theory of the potentially infinite or I_{PI}:

$$\mathsf{I}_{\mathsf{PI}} = \{\varphi \mid \mathsf{ACA}_0 \vdash \text{`} \vDash_{\mathsf{PI}} \varphi\text{'}\}. \tag{2}$$

Intuitively, this theory is every formula that can be proven valid on potentially infinite models given the weakest metatheory that can formalize the models. A full definition is given in Appendix B.[4] The definition of interpretation is traditionally restricted to theories in the same logic, whereas in this setting E_{PI} and I_{PI} are theories in second-order modal logic but PA^1, PA^2, TA^1, and TA^2 aren't modal theories. So, to state and prove our main results we need a more general notion of *generalized translation,* and *interpretation* which captures those interpretations which involve not just different theories but different logics. This is defined in §5. Our first main result is:

THEOREM 1.8.

(i) *There is a generalized translation from the language of* PA^1 *to the second-order modal language with octothorpe that interprets* TA^1 *in* E_{PI}.
(ii) *There is a generalized translation from the language of* PA^1 *to the second-order modal language with octothorpe that interprets* PA^1 *in* I_{PI}. *Further, this is a* PA^1*-verifiable generalized interpretation.*

This result is proven in §5. The translation used is based on one offered by Linnebo (2013) in the setting of modal set theory. The key difference, compared with the standard notion of translation, is that "for all" is translated as "necessarily for all" and, similarly, "there is" is translated as "possibly there is."

The first theorem shows that the PI models capture a significant amount of mathematics. However, we cannot strengthen the result to second-order theories of arithmetic as our second main theorem shows:

[4] We picked the weakest theory because we are interested in what is deducible from PI models and if we strengthen the metatheory I_{PI} will be strengthened in ways that reflect what the metatheory thinks about finite sets (which can code consistency statements).

https://doi.org/10.1017/S1755020320000349 Published online by Cambridge University Press

THEOREM 1.9.

(i) *There is no generalized translation from the language of* PA^2 *to the second-order modal language with octothorpe that interprets* TA^2 *in* E_{PI}.
(ii) *There is no generalized translation from the language of* PA^2 *to the second-order modal language with octothorpe that* PA^2*-verifiably interprets* PA^2 *in* I_{PI}.

For both E_{PI} and I_{PI}, the results follow from the fact that PI models are Π^1_1 definable. And this follows because all of the worlds are finite. Because of this, PI models are representable in reasonably weak theories of second-order arithmetic. But then limitive results about what theories can represent about themselves will stop theories that can represent E_{PI} and I_{PI} being interpretable into E_{PI} and I_{PI}.

These results are important because they show that less mathematics is analytic on the philosophical perspective which motivates the potentially infinite models than on the traditional perspective. The external theory cannot recover TA^2 but only TA^1. And the internal theory cannot recover PA^2 but only PA^1. Further, PA^2 has traditionally been the target of Fregean interpretation results as it allows for the recovery of analysis and much of mathematics.[5] Analysis can be coded in second-order Peano arithmetic, as real numbers can be coded as sets of rationals, which in turn can be coded as naturals. This means that Frege's theorem already accounts for a larger expanse of mathematics than it might first appear. If we try to avoid the claim that it is analytic that there are actually infinitely many objects, however, it then seems we will not have managed to recover as much mathematics. If we are looking to show that mathematics is analytic, we have moved further from our goal.

However, we have still captured a substantial chunk of our most frequently used mathematics. Feferman (2005, p. 613) has argued that all scientifically applicable analysis can be developed in PA^1 or a conservative extension of it.[6] If this is correct then we can still recover the mathematics for which an explication of its truth is most philosophically fruitful, namely the mathematics which we rely on when we act in the world. One might wonder why a logicist would care about whether or not the mathematics recovered is used. But it seems we should keep an open mind to different parts of mathematics being justified in different ways. Maybe something as fundamental as first-order arithmetic turns out to be analytic, but it seems unlikely that the same is true of the higher reaches of set theory. With this in mind, it should not be damaging that not all mathematics turns out to be analytic.

1.3. A diversity of modal logicisms. The idea of using the potentially infinite as a foundation of logicism has a pedigree in the work of Putnam and Hodes, and more recent work on modal foundations of mathematics and on variants of Frege's theorem in different logics. Putnam suggested that by accepting a modal picture of mathematics we could avoid being Platonists about the numbers or committing to how many objects there actually are. This is stated most clearly when he writes:

[5] (Demopoulos (1994), 238 n26) points out that Frege often uses arithmetic when he means something broader including analysis.

[6] For example, "By the fact of the proof-theoretical reduction of W to $[\mathsf{PA}^1]$, the only ontology it commits one to is that which justifies acceptance of $[\mathsf{PA}^1]$." Feferman (2005, p. 613) works in a system W which contains types for the naturals, the cross product and partial functions. The full classical analysis of continuous functions can be carried out in W (Feferman, 2005, p. 611).

https://doi.org/10.1017/S1755020320000349 Published online by Cambridge University Press

> 'Numbers exist'; but all this comes to, for mathematics anyway, is that (I) ω-sequences are possible (mathematically speaking); and (2) there are necessary truths of the form 'if α is an ω-sequence, then…'[.] (Putnam, 1967, pp. 11–12)

Hodes took on this idea, but he was sceptical of the existence of actual infinities. He thought that '[a]rithmetic should be able to face boldly the dreadful chance that in the actual world there are only finitely many objects' (Hodes, 1984, p. 148). His solution made use of the idea of the potentially infinite rather than the actually infinite. He appealed to modality and in particular the modality that seems to be implicit in our concept of number: the idea that it is always possible to add 1 (Hodes, 1990, p. 378).

However, by 1990, Hodes concluded that the reduction of mathematics to higher-order modal logic had failed. Hodes describes the problem as follows:

> The problem is simple: relative to [a model of Hume's Principle] for a type-0 variable v, $\Diamond(\exists v)(\underline{N}(v)\&\dots)$ "moves us" to other worlds u and then has us seek a witnessing member of [the natural number in the model] in [the domain of u]; we may find one, but then have no way "back" to w to see what hold [sic] for it there. (Hodes, 1990, p. 388)

So we might know that there possibly exists a number with a property, but in Hodes's system, we have no way of returning to our original world to use what we have found. For example, if we find the number of a set in some world, we have no assurance that this number is available for us to talk about in the world the set came from. It is only known that it is the number of the set *in the world the number exists in*. The difficulty identified here is with cross-world predication, which occurs when we want to say something about an object in one world and how it relates to objects in another world (Kocurek, 2016).

In what follows we will show that the problem is not with cross-world predication *per se*. Both by working directly with the models, but also by allowing the octothorpe to be rigid, we can mimic some of the effects of cross-world predication. Yet in this setting we recover some but not all of the arithmetic recovered by Frege's theorem. Indeed, our main results, Theorems 1.8 and 1.9, show that the situation is more complicated than Hodes suggested, and that a partial realization of his project is possible.

There are two recent trends in the study of logicism which this project is connected to. First, Studd (2016) has suggested that the modal setting is an attractive one for the logicist because it would help to solve the bad company objections. Unlike here, Studd's is concerned with inconsistent abstraction principles and in particular set abstraction. This is interestingly connected to the naïve conception of set because one can think of the unrestricted set Comprehension Schema as similar in spirit to a modal version of Basic Law V. While work in this area goes back to Parsons (1983), it has been pursued recently by Linnebo (2013; 2018). Much of Linnebo's work has been on set theory. The concerns there are very different from ours, as it make little sense in set theory to worry about the actual infinite not existing and set theory is generally treated in first-order logic. The work in this paper takes inspiration from the results presented in Linnebo (2013) and (2018) and makes use of a similar method of translating between the modal and nonmodal setting. However, while the dynamic abstraction principles discussed by (2018) resemble the behavior of the *number of* operator, his preferred abstraction

https://doi.org/10.1017/S1755020320000349 Published online by Cambridge University Press

principle for arithmetic is ordinal abstraction (Linnebo, 2018, chap. 10.5), whereas in this paper we work with a modal version of Hume's Principle, a cardinality principle.

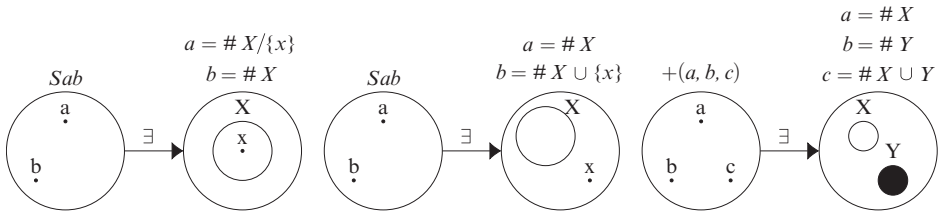
Second, there has been a lot of recent work on whether Frege's Theorem still holds when the logic is modified in certain ways. Bell (1999) and Shapiro and Linnebo (2015) have shown that Frege's Theorem is available in the intuitionistic setting. Burgess (2005) and Walsh (2016) found that a version of Frege's Theorem is possible in a certain predicative setting. Kim (2015) proves a version of Frege's Theorem in a modal setting. This employs an axiomatized version of the 'the number of F's is n' as a binary relation, instead of the traditional type-lowering 'number of' operator. Kim recovers the axioms of PA but finds that a restricted version of HP^2 holds. The modality used is S5 and meant to represent logical possibility, not potentiality. Because of this Kim's system does not have the same structure of our models, where the numbers slowly grow. Closer in spirit to the work here is that on finite models of arithmetic by Mostowski (2001). There he considers initial sequences of the natural numbers and what holds over all such models. These have a clear connection to the minimal model discussed above. Urbaniak (2016) has taken Mostowski's models and worked with them in a modal setting. They have shown that Leśniewski's typed, free logic with modal quantifiers, which proves a predicative version of HP^2, can interpret PA^2. Our setting is quite different from that of Urbaniak's paper as Leśniewski's typed, free logic differs dramatically from the one we work in here. The work in this paper proceeds by looking at whether a version of Frege's Theorem is available in a classical second-order modal setting. Unlike these other results, we find that a modal version of Frege's Theorem for PA^2 is *not* possible, as shown by Theorem 1.9.

1.4. Outline of paper. This paper is organized as follows. §2 expands the potentially infinite models' language to include the language of arithmetic. In §3 we show that using the expanded language the potentially infinite models satisfy a weak theory of arithmetic equivalent to a modal version of Robinson's Q. In §4 we define the inductive formulas of the language and show that induction holds for them. This allows us to show Theorem 1.8, that TA^1 is interpretable in our external theory and PA^1 is interpretable in our internal theory, in §5. In §6 we show that no natural interpretation of PA^2 is possible by proving Theorem 1.9.

§2. Definitions for a modal Grundlagen. Just as Frege in the *Grundlagen* defined the numbers and the relations on them using only the 'number of' operator, here we show how modified versions of Frege's definitions can do this in the setting of the potentially infinite.[7] Proving that these definitions satisfy the usual arithmetical axioms will occupy us in §3–4. In this section we simply set out the definitions themselves and say a word about their motivation. While entirely rigorous, it is our hope that, as in the *Grundlagen*, the definitions will be intuitive and correspond to our understanding of cardinal numbers.

[7] This has some precedent in Hodes (1990, p. 383). However, whereas we (and Frege) first define successor and then use this to build the other definitions, Hodes takes 'less than or equal to' as his primitive. In his system a number N (understood as a higher-order object) is less than or equal to another number N' just in case it is possible that there are two other second-order objects A and A' each with the same number of objects as N and N' respectively and A is a subset of A'. That this has parallels with the definition of successor offered here will be clear on inspection.

https://doi.org/10.1017/S1755020320000349 Published online by Cambridge University Press



(A) Diagram of when a is succeeded by b.

(B) Diagram of when a is succeeded by b for alternative definition of S.

(C) Diagram of when c is the addition of b and c.

Fig. 2.

The first definition is easy and does not require any of the modal apparatus. We simply let $0 = \#\varnothing$. This follows Frege (1884, sec. 74; p. 87) explicitly, who said that zero is "the Number which belongs to the concept 'not identical with itself'."

Next we must define the successor, as the other definitions rely on it. The definition here is like the one offered by Frege, but it differs by allowing the sets which witness that one object is the successor of another to be merely possible. This is to ensure that if an object is ever the successor of another, then it is the successor of that object in every world where they both exist. This property will be important in the proof of induction. The definition of successor, in plain terms, is: one object is the successor of another just in case it is possible that there are two sets, which differ by one object and the successor is the number of the larger set, and the predecessor is the number of the smaller set. Figure 2A and 2B illustrate the two ways this can be done, resulting in two definitions of the successor:

Definition 2.1.

$$Sxy \equiv \Diamond\exists G, u[Gu \wedge (y = \#G) \wedge (x = \#(G - \{u\}))] \tag{3}$$

$$S'xy \equiv \Diamond\exists F, u[\neg Fu \wedge (x = \#F) \wedge (y = \#(F \cup \{u\}))] \tag{4}$$

The first of these definitions simply adds the possibility operator to the definition of successor suggested by Frege (1884, sec. 76; p. 89). These definitions are equivalent: to see this, simply consider $F = G - \{u\}$ and $G = F \cup \{u\}$.[8] In what follows we will simply use the definition that is most convenient and will write S for both.

The definition of addition is similarly intuitive. The relation $+$ holds between three objects a, b, and c such that it is possible that there are disjoint sets X and Y of cardinality a and b respectively, and c is the cardinality of $X \cup Y$, the union of the two disjoint sets. This is illustrated by Figure 2c and can be written formally as:

Definition 2.2.

$$+(a, b, c) \equiv \Diamond\exists X, Y(a = \#X \wedge b = \#Y \wedge c = \#X \cup Y \wedge (X \cap Y) = \varnothing) \tag{5}$$

For c to be the result of multiplying a and b we need a set B of cardinality b and for each element x of B a set A_x of cardinality a. The A_x's must all be disjoint. And c must

[8] For easy of readability, we will use set theoretic notation as a convenient short hand for concepts formed using the language of the model. So $F \cup \{u\}$ is used for the concept given by $Xx \leftrightarrow (Fx \vee x = u)$.

https://doi.org/10.1017/S1755020320000349 Published online by Cambridge University Press

be the cardinality of the union of all the A_x's. To define the A_x's we define a binary relation P that holds between x in B and all y in A_x. So A_x is $\{y \mid Pxy\}$.

Definition 2.3.

$$\times(a,b,c) \equiv \Diamond\exists X, P[\#X = b \wedge \forall x \in X(\#\{y \mid Pxy\} = a) \qquad (6)$$
$$\wedge\forall x, y \in X(x \neq y \rightarrow \{z \mid Pxz\} \cap \{z \mid Pyz\} = \varnothing) \wedge \# \bigcup_{x \in X} \{y \mid Pxy\} = c]$$

The definition of the natural numbers is more complicated and require us to define the notion that one number follows another in the ordering of the natural numbers. We will make use of Frege's definition from the 1879 *Begriffsschrift* (1967, sec. III; pp. 55 ff; 1884, sec. 79; p. 92 ff). Russell and Whitehead (1910, p. 316) called this relation the *ancestral relation* because a good example of what it does is define the relation 'ancestor of' from the relation 'parent of'. The *strong ancestral* of φ holds between two objects a and b just in case b is contained in every set such that the set is closed under φ and the set contains everything a bears φ to. So, we can define someone's ancestors as everyone who is in every set that contains their parents and the parents of everyone in the set. It is not guaranteed that a bears this relation to itself, and so we also define the reflexive *weak ancestral*.

Definition 2.4 The strong ancestral.

$$\varphi^{+}(a,b) \equiv \forall X[(\forall x, y(Xx \wedge \varphi(x,y) \rightarrow Xy) \wedge \forall x(\varphi(a,x) \rightarrow Xx)) \rightarrow Xb].$$

Definition 2.5 The weak ancestral.

$$\varphi^{+=}(a,b) \equiv \varphi^{+}(a,b) \vee a = b.$$

Using this definition, we define a natural number as an object that is some finite number of successor steps from 0, assuming 0 exists.

Definition 2.6 Natural Number.

$$\mathbb{N}x \equiv S^{+=}0x \wedge \exists y(y = 0).$$

This definition closely parallels Frege's, though the definition of S is different. The existence claim is added because in the modal setting 0's existence cannot be assumed. For example, 0 does not exist at worlds **0**, **1**, and **2** in the **0-3** swap model, and, as **0** is not a member of infinitely many finite subsets of the natural numbers, 0 does not exist at infinitely many worlds in the subset model. In these worlds nothing is a natural number.

2.1. Some useful results. The following six lemmas will help explain the behavior of $\mathbb{N}$ in the models. We omit the proofs as they do not pose any particular difficulty. For the following Lemmas, recall Definition 1.7 where $\vDash_{\mathsf{PI}} \varphi$ was defined as φ is true in all worlds in all potentially infinite models. First, note that the set defined by $\mathbb{N}$ at a world satisfies the antecedent of $S^{+}0x$. Intuitively, the idea here is that if x is in every set containing 0 and closed under S, and Sxy, or $S0y$, then y must also be in every set with these properties.

Lemma 2.7. $\vDash_{\mathsf{PI}} \exists x(x = 0) \rightarrow \forall y(S0y \rightarrow \mathbb{N}y))$

Lemma 2.8. $\vDash_{\mathsf{PI}} \forall x, y(\mathbb{N}x \wedge Sxy \rightarrow \mathbb{N}y)$

https://doi.org/10.1017/S1755020320000349 Published online by Cambridge University Press

It follows immediately from this that if x exists at a world and at that world $\mathbb{N}y$ and Syx then $\mathbb{N}x$. However, that doesn't mean $\mathbb{N}$ is the set of all numbers across all worlds as $\mathbb{N}$ only holds of objects which exist at the world of evaluation. This contrasts with our other definitions where the objects need not exist at the world.

LEMMA 2.9. $\models_{PI} \mathbb{N}x \rightarrow \exists y\ y = x$

This is because the quantifiers in $\mathbb{N}$ are plain rather than having modals in front of them. This is important because if we put the modals in front everything is a number!

We informally extend our definition of the interpretation function I to $I(\mathbb{N}, s) = \{x \in D(s) \mid \mathcal{M}, s \vDash \mathbb{N}x\}$. Note that by Lemma 2.9 we have $\{x \in D(s) \mid \mathcal{M}, s \vDash \mathbb{N}x\} = \{x \in D \mid \mathcal{M}, s \vDash \mathbb{N}x\}$, where D is the domain of the model not the world.

Recall that $\boldsymbol{a}_i$ is the unique element in D such that if $|X| = i$ then $I(\#, w)(X) = \boldsymbol{a}_i$ as defined in 1.2.5. We can now explicitly describe the interpretation of $\mathbb{N}$ at a world w in terms of the $\boldsymbol{a}_i$'s, that is, the set $I(\mathbb{N}, w)$:

LEMMA 2.10. *Let w be a world and let n be the first number such that $\boldsymbol{a}_n \notin D(w)$. Then if $n > 0$, it follows that $\{0, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_{n-1}\} = I(\mathbb{N}, w)$, and further, $n = 0$ iff $I(\mathbb{N}, w) = \varnothing$.*

This result shows us how the differences between our modal setting and the traditional nonmodal setting of the *Grundlagen* become most stark in the case of the interpretation of the natural numbers at a world. Two things are worth highlighting. The first is that $\mathbb{N}$ is finite at every world, since it is a subset of the domain of the world, and the domain of every world is finite. The second is that objects that are not in $\mathbb{N}$ at one world can 'become' numbers at later worlds. This doesn't happen in the minimal model, where $I(\mathbb{N}, \boldsymbol{n}) = D(\boldsymbol{n})$ at every world. But it does in the subset model. For example, $I(\mathbb{N}, \{\mathbf{2}, \mathbf{100}\}) = \varnothing$, $I(\mathbb{N}, \{\mathbf{0}, \mathbf{1}, \mathbf{3}\}) = \{\mathbf{0}, \mathbf{1}\}$ and $I(\mathbb{N}, \{\mathbf{0}, \mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{100}\}) = \{\mathbf{0}, \mathbf{1}, \mathbf{2}, \mathbf{3}\}$. This distinguishes $\neg\mathbb{N}(x)$ from the other relations which have a certain stability; if objects stand in these relations at one world, then they do so in all worlds in which they all exist. The formal definition of stability is given as Definition 8. This difference is caused by there being no possibility operator at the beginning of the definition of $\mathbb{N}$. Despite this, once something is a number it remains one:

LEMMA 2.11. $\vDash_{\mathsf{PI}} S(x, y) \rightarrow \Box S(x, y)$ *holds, as does* $\vDash_{\mathsf{PI}} S^{+}(x, y) \rightarrow \Box S^{+}(x, y)$, $\vDash_{\mathsf{PI}} S^{+=}(x, y) \rightarrow \Box S^{+=}(x, y)$ *and* $\vDash_{\mathsf{PI}} \mathbb{N}x \rightarrow \Box\mathbb{N}x$.

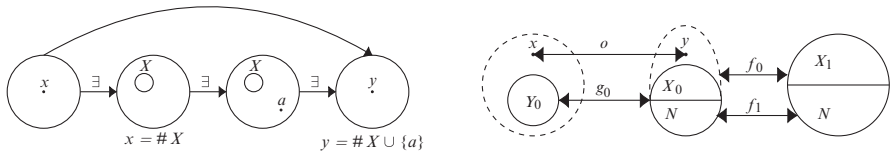
It is also worth noting that even though some cardinalities may not be numbers at ever world, the cardinality of every set eventually becomes a natural number.

LEMMA 2.12. *For all $w \in W$ and $X \subseteq D(w)$, there is a world s such that $R(w, s)$ and $\#X \in I(\mathbb{N}, s)$.*

This is because # is a function, first-order converse Barcan holds, and the accessibility relation is directed. With these preliminary results we can now show our definitions satisfy a simple theory of arithmetic.

§3. Proving modalized Robinson's Q. In what follows we will prove that the modalized axioms of Robinson's Q are true on all PI models (cf. Definition 1.7). Robinson's Q is a weak theory of arithmetic that defines successor as an injective function that never returns 0 and gives a recursive definition of addition and multiplication. By "modalized" we mean that we write "necessarily for all" for "for all"

https://doi.org/10.1017/S1755020320000349 Published online by Cambridge University Press



(A) (B)

Fig. 3. (A) Finding the successor of x (B) Proof of the recursion clause for addition.

and "possibly there is" for "there is". In other words, it is what results when we apply the Linnebo translation (mentioned in the introduction) to the axioms of Robinson's Q. The theory PA^1 is obtained by adding the mathematical induction schema to Q. We deal with PA^1 and the proof of the induction schema in §4.[9]

First we will show that our relations define the graphs of functions. The easiest case is successor.

LEMMA 3.1 (S1). $\vDash_{\mathsf{PI}} \Box\forall x, y, z \in \mathbb{N}((Sxy \wedge Sxz) \rightarrow y = z)$.

Proof. Let $s \in W$ and $x, y, z \in I(\mathbb{N}, s)$ satisfy the antecedent. As x is the predecessor in both relations it follows by directedness that there is a $w \in W$, such that $R(s, w)$ where there are $X, X' \subseteq D(w)$ and $\#X = x = \#X'$. As such there is a bijection $g : X \rightarrow X'$. There will also be $a, b \in D(w)$ such that $a \notin X$, $b \notin X'$, and $y = \#X \cup \{a\}$ and $z = \#X' \cup \{b\}$. As $a \notin X$ and $b \notin X'$ we can construct h such that for all $u \in X$, $h(u) = g(u)$ and $h(a) = b$. Clearly h is a bijection, so $y = \#X \cup \{a\} = \#X' \cup \{b\} = z$. □

LEMMA 3.2 (S2). $\vDash_{\mathsf{PI}} \Box\forall x \in \mathbb{N}\Diamond\exists y \in \mathbb{N}\ Sxy$.

Proof. As illustrated in Figure 3A, let $s \in W$ and $x \in I(\mathbb{N}, s)$, it follows that $x = \boldsymbol{a}_n$ for some n and, by Lemma 2.10, $\{0, \ldots \boldsymbol{a}_{n-1}\} \subsetneq D(s)$. Further, $\boldsymbol{a}_n = \#\{0, \ldots \boldsymbol{a}_{n-1}\}$ and $\boldsymbol{a}_n \notin \{0, \ldots \boldsymbol{a}_{n-1}\}$. Thus, there must be a further world w accessible from w_1 and a $y \in D(w)$ such that $y = \#\{0, \ldots \boldsymbol{a}_{n-1}\} \cup \{\boldsymbol{a}_n\}$. It follows that Sxy at w. By Lemma 2.11 $x \in I(\mathbb{N}, w)$. As $\mathbb{N}$ is closed under successor by Lemma 2.8, we have that $y \in I(\mathbb{N}, w)$. And since R is transitive, w is accessible from s. □

These two proofs offer a general outline of the reasoning for addition and multiplication. For A1 and M1 this strategy is to show that whatever x is the sets assigned to y and z will have the same cardinality. Where as for A2 and M2 one simply needs to construct a set of the correct cardinality. For this reason we do not give the proofs for the next four lemmas.

LEMMA 3.3 (A1). $\vDash_{\mathsf{PI}} \Box\forall x, y, z, z' \in \mathbb{N}(+(x, y, z) \wedge +(x, y, z') \rightarrow z = z')$.

LEMMA 3.4 (A2). $\vDash_{\mathsf{PI}} \Box\forall x, y \in \mathbb{N}\Diamond\exists z \in \mathbb{N}\ +(x, y, z)$.

LEMMA 3.5 (M1). $\vDash_{\mathsf{PI}} \Box\forall x, y, z, z' \in \mathbb{N}(\times(x, y, z) \wedge \times(x, y, z') \rightarrow z = z')$.

LEMMA 3.6 (M2). $\vDash_{\mathsf{PI}} \Box\forall x, y \in \mathbb{N}\Diamond\exists z \in \mathbb{N}\ \times(x, y, z)$.

[9] A list of the nonmodalized axioms can be found in Appendix A. While what we show here is that these axioms are in the theory E_{PI}, each of the proofs that follow can be formalized in ACA_0 (cf. Appendix B). That this is possible will ensures that all axioms proven here are also in the theory I_{PI} (from §1.2). This is a key point in the proof of Theorem 1.8.ii which we complete in §5.

https://doi.org/10.1017/S1755020320000349 Published online by Cambridge University Press

We also need to show that 0 meets the right conditions to be a constant.

LEMMA 3.7 (Z1). $\vDash_{\mathsf{PI}} \Diamond\exists x \in \mathbb{N}(x = 0 \wedge \Box\forall y(y = 0 \to y = x))$.

Proof. By the definition of $\mathbb{N}$, it follows that $0 \in I(\mathbb{N}, s)$ for any world s with 0 in the domain. And as $0 = \#\varnothing$ there is some s with 0 in the domain. The second conjunct follows by the transitivity of identity. □

We can now move on to the recursion equations in Q. We separate these into the base steps concerning 0 and the recursive step. For the base steps, because $0 = \#\varnothing$ the proofs of the lemmas are relatively straight forward. As such we list them here without proof.

LEMMA 3.8 (Q1). $\vDash_{\mathsf{PI}} \neg\Diamond\exists x \in \mathbb{N}(Sx0)$.

LEMMA 3.9 (Q3). $\vDash_{\mathsf{PI}} \Box\forall x \in \mathbb{N} \ +(x, 0, x)$.

LEMMA 3.10 (Q5). $\vDash_{\mathsf{PI}} \Box\forall x \in \mathbb{N} \ \times(x, 0, 0)$.

What is left now is to show the recursion steps. He we only prove the case for $+$ as one can use the same stratagy for $\times$ and the proof is simple for S.

LEMMA 3.11 (Q2). $\vDash_{\mathsf{PI}} \Box\forall x, y, z \in \mathbb{N}((Sxz \wedge Syz) \to x = y)$.

The proof simply follows from the fact that if there is a bijection between two sets X and Y then there will be a bijection between $X \cup \{a\}$ and $Y \cup \{b\}$ if a and b aren't in X or Y respectively.

LEMMA 3.12 (Q4).

$$\vDash_{\mathsf{PI}} \Box\forall n, x_0, x_1, y_0, y_1, z \in \mathbb{N}(S(x_0, x_1) \wedge S(y_0, y_1) \wedge +(n, x_0, y_0) \wedge +(n, x_1, z) \to y_1 = z).$$

Proof. As illustrated in Figure 3B, let $s \in W$ and $n, x_0, x_1, y_0, y_1, z \in I(\mathbb{N}, s)$ satisfy the antecedent. We want to show that $y_1 = z$. By directedness, we know there is a world w containing all the objects and sets which the antecedent states possibly exist. As y_1 succeeds y_0 there is a set Y_0 and an object $a \notin Y_0$ at w such that $y_1 = \#Y_0 \cup \{a\}$ and $y_0 = \#Y_0$. We know y_0 to be the addition of n and x_0 so there are disjoint sets N and X_0 such that $n = \#N$, $x_0 = \#X_0$, and $y_0 = \#N \cup X_0$. Further there is a bijection $g_0 : Y_0 \to N \cup X_0$. Now let b be an element not in N or X_0 (we can always pick w so that such an element exists). Clearly we can define a bijection o between the singletons of a and b. Now, using g_0 and o, define the bijection $g : Y_0 \cup \{a\} \to N \cup X_0 \cup \{b\}$, as the union of g_0 and o. Now as x_1 is the successor of x_0, it follows that $x_1 = \#X_0 \cup \{b\}$. As z is the addition of n and x_1 there are disjoint sets N' and X_1 such that $n = \#N = \#N'$, $x_1 = \#X_0 \cup \{b\} = \#X_1$ and $z = \#N \cup X_1$. As such there are bijections $f_0 : X_0 \cup \{b\} \to X_1$ and $f_1 : N \to N'$. So, we can define the bijection $f : N \cup X_0 \cup \{b\} \to N' \cup X_1$ as f_0 on $X_0 \cup \{b\}$ and f_1 on N. Then as $z = \#N \cup X_1$ the composition $f \circ g$ is a bijection proving $y_1 = z$. □

LEMMA 3.13 (Q6). $\vDash_{\mathsf{PI}} \Box\forall n, x_0, x_1, y_0, y_1, z \in \mathbb{N}(S(x_0, x_1) \wedge +(n, y_0, y_1) \wedge \times(n, x_0, y_0) \wedge \times(n, x_1, z) \to y_1 = z)$.

This proof is similar to the above except we end up showing that $y_1 = \#\bigcup_{x \in A_0 \cup \{u\}}\{y \mid Pxy \vee (x = u \wedge y \in N)\} = \#\bigcup_{x \in A_1}\{y \mid Txy\} = z$ where A_0, A_1, and N are of cardinality x_0, x_1, and n respectively and P is the relation given by $\times(n, x_0, y_0)$ and T by $\times(n, x_1, z)$.

https://doi.org/10.1017/S1755020320000349 Published online by Cambridge University Press

These results show that we have successfully defined a modalized version of Robinson's Q in our system. The next section will recover a modalized induction schema.

§4. Proving the modalized induction schema. We have succeeded in giving a weak theory of arithmetic in a potentially infinite setting. However, we can recover more arithmetic by proving that when restricted to appropriate formulas a modalized version of the induction schema is true on all PI models. The modalized induction schema is:

$$[\varphi(0) \wedge \Box\forall x, y \in \mathbb{N}(\varphi(x) \wedge S(x, y) \rightarrow \varphi(y))] \rightarrow \Box\forall x \in \mathbb{N}\, \varphi(x) \tag{7}$$

Modalized induction does not hold for all formulas in our models, as will be shown in Lemma 4.4. So, we need to define a subclass of the formulas in the language of potentially infinite models for which it does hold. These we will call the inductive formulas, and in Lemma 4.3 it will be proven that induction does hold for inductive formulas.[10]

DEFINITION 4.1. *The inductive terms and formulas are defined recursively as follows:*

1. *An inductive term is either 0 or a first-order variable.*
2. *If t_0, t_1, t_2 are inductive terms then $t_0 = t_1$, $S(t_0, t_1)$, $+(t_0, t_1, t_2)$ and $\times(t_0, t_1, t_2)$ are inductive formulas.*
3. *Applications of the propositional connectives to inductive formulas are inductive formulas.*
4. *If φ is an inductive formula then $\Box\forall x \in \mathbb{N}\ \varphi$ and $\Diamond\exists x \in \mathbb{N}\ \varphi$ are inductive formulas.*

The inductive terms and formulas are a subset of the terms and formulas respectively. Any term of the form $\#X$ is not an inductive term, and indeed no term or formula with a free second-order variable is inductive. Likewise $\mathbb{N}0$, $\forall z(x = z)$ and $\exists y(S0y)$ are not inductive formulas, while $\Box\forall z \in \mathbb{N}(x = z)$ and $\Diamond\exists y \in \mathbb{N}(S0y)$ are.

A formula φ is *stable* when:

$$\vDash_{\mathsf{PI}} \varphi \rightarrow \Box\varphi. \tag{8}$$

Stability is taken from Linnebo's (2013, p. 211) work on set theory in a modal setting. It means once a formula has been made true it stays true. As we saw in Lemma 2.11, S, S^+, $S^{+=}$, and $\mathbb{N}$ are all stable and an example of an unstable formula is $\neg\mathbb{N}$. Fortunately, the inductive formulas all have the property of being stable, as we will now prove. This will allow us to prove induction for these formulas.

LEMMA 4.2. *If φ is an inductive formula then $\vDash_{\mathsf{PI}} \varphi \rightarrow \Box\varphi$.*

Proof. In what follows we prove by induction on the complexity of the inductive formulas that both $\varphi \rightarrow \Box\varphi$ and $\Diamond\varphi \rightarrow \varphi$. The second condition is included to deal with the case of negation.

Base case: $x = y$ and $x = 0$: The result follows from the evaluation of $\#\varnothing$ being rigid and the identity relation being interpreted as the identity from the metalanguage. Note that for S, $+$, and $\times$ that $\Diamond\psi \rightarrow \psi$ follows simply because R is transitive and they

[10] This terminology is used to distinguish between these formulas and other for which induction does not hold. Hopefully no confusion will be caused by the distinct uses of the term inductive formulas elsewhere in the literature.

https://doi.org/10.1017/S1755020320000349 Published online by Cambridge University Press

start with a $\Diamond$. $S(x, y)$: See Lemma 2.11. $+(x, y, z)$: Assume that $\mathcal{M}, w \vDash +(a, b, c)$. It follows that there exists a world w' accessible from w and nonintersecting sets $A, B \subseteq D(w')$ satisfying $+$. Let s be a world such that $R(w, s)$. Then by directedness, there is a world s' such that $R(s, s')$ and $R(w', s')$, and $A, B \subseteq D(s')$. So $+(a, b, c)$ holds at s. $\times(x, y, z)$: The reasoning is essentially the same as that used for $+$.

Now we proceed to the induction step. We will only show the case of the quantifier as $\neg$ and $\wedge$ proceed as one would expect. $\Diamond\exists x \in \mathbb{N}\ \psi$: Assume $\mathcal{M}, s \vDash \Diamond\Diamond\exists x \in \mathbb{N}\ \psi$. It follows by transitivity that $\mathcal{M}, s \vDash \Diamond\exists x \in \mathbb{N}\ \psi$. Now we show that $(\Diamond\exists x \in \mathbb{N}\ \psi) \rightarrow (\Box\Diamond\exists x \in \mathbb{N}\ \psi)$. First take a world w such that $\Diamond\exists x \in \mathbb{N}\ \psi$ holds at w. Then take worlds s, w' such that $R(w, s)$, $R(w, w')$, $\exists x \in \mathbb{N}\ \psi$ holds at w' and we want to show $\Diamond\exists x \in \mathbb{N}\ \psi$ holds at s. At w' there is an $a \in D(w')$ such that $a \in I(\mathbb{N}, w')$ and $\psi(a)$ holds at w'. So, by Lemma 2.11, $\mathbb{N}a \rightarrow \Box\mathbb{N}a$ holds at w' and by the induction hypothesis, $\psi(a) \rightarrow \Box\psi(a)$. Let s' be such that $R(s, s')$ and $R(w', s')$, such a world exists by directedness. It follows that $\mathbb{N}a$ and $\psi(a)$ hold at s' and as s' is accessible from s we have proven $\Diamond\exists x \in \mathbb{N}\ \psi$ holds at s. □

We can now prove that the modalized induction schema holds for all inductive formulas. We do this by showing the more general result that induction holds for all stable formulas.

LEMMA 4.3. *If φ is stable, then*

$$\vDash_{\mathsf{PI}} [\varphi(0) \wedge \Box\forall x, y \in \mathbb{N}(\varphi(x) \wedge S(x, y) \rightarrow \varphi(y))] \rightarrow \Box\forall x \in \mathbb{N}\ \varphi(x).$$

Proof. Let w be a world. Further, we assume the antecedent of the induction schema holds so let $\varphi(0)$ and $\Box\forall x, y \in \mathbb{N}(\varphi(x) \wedge S(x, y) \rightarrow \varphi(y))$ hold at w. Let s be a world accessible from w and let $a \in I(\mathbb{N}, s)$. We will show that $\varphi(a)$ at s. If $a = 0$ then, as φ is stable, we are done so assume not.

As $a \in I(\mathbb{N}, s)$, if we prove $\forall x, y(\varphi(x) \wedge \mathbb{N}x \wedge S(x, y) \rightarrow \varphi(y) \wedge \mathbb{N}y)$ and $\forall x(S(0, x) \rightarrow \varphi(x) \wedge \mathbb{N}x)$ hold at s then we have satisfied the antecedent of S^+0a and so it follows that $\varphi(a) \wedge \mathbb{N}a$ at s.

At s we have $\forall x, y \in \mathbb{N}(\varphi(x) \wedge S(x, y) \rightarrow \varphi(y))$. We also have that if $x \in I(\mathbb{N}, s)$, and $S(x, y)$ hold at s then by Lemma 2.8 that $y \in I(\mathbb{N}, s)$. This proves $\forall x, y(\varphi(x) \wedge \mathbb{N}x \wedge S(x, y) \rightarrow \varphi(y) \wedge \mathbb{N}y)$ at s.

From $a \in I(\mathbb{N}, s)$ it follows that $0 \in D(s)$. Assume $x \in D(s)$ and $S0x$, as $0 \in D(s)$ it follows by Lemma 2.7 that $x \in I(\mathbb{N}, s)$. It then follows by the stability of φ that $\varphi(0)$ at s. As such we have the antecedent of $\forall x, y \in \mathbb{N}(\varphi(x) \wedge S(x, y) \rightarrow \varphi(y))$ so we get $\varphi(x)$. And from this it follows that $\forall x(S(0, x) \rightarrow \varphi(x) \wedge \mathbb{N}x)$ holds at s. □

We have proven the modalized induction axiom restricted to inductive formulas. But we cannot prove modalized induction for all formulas in the language of potentially infinite models, as the following counterexample shows.

LEMMA 4.4. *If $\varphi(x)$ is $\forall z(z = x)$, then*

$$\nvDash_{\mathsf{PI}} [\varphi(0) \wedge \Box\forall x, y \in \mathbb{N}(\varphi(x) \wedge S(x, y) \rightarrow \varphi(y))] \rightarrow \Box\forall x \in \mathbb{N}\ \varphi(x).$$

Proof. It is sufficient to show there is a model and a world in the model where this statement is false. Take the minimal model from Example 1.4 and world $\mathbf{0}$, where $D(\mathbf{0}) = \{\mathbf{0}\}$. Clearly $\mathcal{M}, \mathbf{0} \vDash \forall z(z = 0)$. Let $w \in W$ be such that $R(\mathbf{0}, w)$ and assume that for all $x, y \in I(\mathbb{N}, w)$, that $\forall z(z = x)$ and $S(x, y)$ hold at w. As everything in the domain is equal to x it follows that $y = x$ and so $\forall z(z = y)$ at w. So $\mathcal{M}, \mathbf{0} \vDash \Box\forall x, y \in$

https://doi.org/10.1017/S1755020320000349 Published online by Cambridge University Press

$\mathbb{N}(\forall z(z = x) \wedge S(x, y) \rightarrow \forall z(z = y))$. But it does not follow that $\Box\forall x \in \mathbb{N}\,\forall z(z = x)$, because $\mathbf{1} \in W$ is a counter example as $D(\mathbf{1}) = \{\mathbf{0}, \mathbf{1}\}$. □

§5. Proof of Theorem 1.8. We now have almost all the pieces needed to prove Theorem 1.8. However, before we do that we need to discuss what a translation and interpretation are in our setting because we are moving between logics.

Intuitively, a *translation* between two languages starts with instructions on how to rewrite atomic formulas in one language into the other language. It does not make any changes to the propositional connectives but can restrict the quantifiers to objects meeting some conditions. In the current setting, however, we need a formal definition of what is to count as a translation when the underlying logics are different. This notion should, at the very least, capture the Linnebo translation. We offer the following definition as a minimal condition on any translation, though more will need to be done to ensure a widely applicable definition of translation and interpretation between logics.

DEFINITION 5.1. *Let L_A and L_B be two logics extending first-order predicate logic, defined by the languages $\mathcal{L}_A$ and $\mathcal{L}_B$ and derivability relations $\vdash_{L_A}$ and $\vdash_{L_B}$ respectively. A generalized translation is given by a recursive map $(\cdot)^{\mathcal{G}} : \mathcal{L}_A \rightarrow \mathcal{L}_B$ which preserves free variables and a domain formula $\delta(x) \in \mathcal{L}_B$, such that the map is compositional on the propositional connectives and where for all unnested formulas*[11] *$\varphi_1, \ldots, \varphi_n, \psi$ containing free variables $x_1, \ldots, x_m$ one has the following:*

$$\varphi_1, \ldots, \varphi_n \vdash_{L_A} \psi \Rightarrow \delta(x_1), \ldots, \delta(x_m), \varphi_1^{\mathcal{G}}, \ldots, \varphi_n^{\mathcal{G}} \vdash_{L_B} \psi^{\mathcal{G}} \tag{9}$$

What we have done so far is an informal translation from the first-order language of arithmetic into the signature of the potentially infinite models. In §2 we showed how the atomic formulas could be translated. Further, the modalized versions of the axioms of PA^1 proven in §3 and §4 are the translations of PA^1's axioms via the translation found in §2 and the Linnebo translation for the quantifiers.

While it has been set out in previous sections, for the sake of definiteness we here record the translation explicitly. We will call this translation $(\cdot)^{\mathcal{F}}$, as it is a Fregean translation. Three things are worth noting before we lay out the translation. The first is that the domain formula associated to this interpretation is $\mathbb{N}$ from Definition 2.6. The second is that the range of this translation is the inductive formulas from Definition 4.1. The third is that 11–13 are the same definitions given in 3, 2.2, and 2.3. We have not changed the definitions we are working with. Rather, we merely show how these definitions can be used to define the interpretation function $(\cdot)^{\mathcal{F}}$.

$$0^{\mathcal{F}} \equiv \#\varnothing, \tag{10}$$

$$Sab^{\mathcal{F}} \equiv \Diamond\exists G\exists u[Gu \wedge (b = \#G) \wedge (a = \#G \cup \{u\})], \tag{11}$$

$$+(a, b, c)^{\mathcal{F}} \equiv \Diamond\exists X, Y(a{=}\#X \wedge b{=}\#Y \wedge c{=}\#X \cup Y \wedge X \cap Y = \varnothing), \tag{12}$$

[11] An unnested formula is one where the atomic subformulas of a formula contain at most one constant, function or relation (Hodges, 1993, p. 58). We only give conditions for unnested formulas. So, for example, Sxy and $+(x, y, z)$ are unnested but $S0x$ and $+(0, 0, z)$ are nested. Every formula is equivalent to an unnested one (Hodges, 1993, p. 59, Cor 2.6.2). As such the translation can be expanded to unnested formulas using this equivalence.

https://doi.org/10.1017/S1755020320000349 Published online by Cambridge University Press

$$\times(a,b,c)^{\mathcal{F}} \equiv \Diamond\exists X, P[\#X = b \wedge \forall x \in X(\#\{y \mid Pxy\} = a) \wedge \forall x, y \in X(x \neq y \rightarrow \{z \mid Pxz\} \cap \{z \mid Pyz\} = \varnothing) \wedge \# \bigcup_{x \in X} \{y \mid Pxy\} = c], \tag{13}$$

$$(\psi \wedge \chi)^{\mathcal{F}} \equiv \psi^{\mathcal{F}} \wedge \chi^{\mathcal{F}}, \tag{14}$$

$$(\neg\psi)^{\mathcal{F}} \equiv \neg\psi^{\mathcal{F}}, \tag{15}$$

$$(\forall x\psi)^{\mathcal{F}} \equiv \Box\forall x(\mathbb{N}(x) \rightarrow \psi^{\mathcal{F}}), \tag{16}$$

$$(\forall X^n\psi)^{\mathcal{F}} \equiv \Box\forall X^n(\forall x_1, \ldots, x_n(X^n x_1 \ldots x_n \rightarrow \mathbb{N}(x_1) \wedge \cdots \wedge \mathbb{N}(x_n)) \rightarrow \psi^{\mathcal{F}}). \tag{17}$$

To see that this is a generalized translation all that remains to be shown is that deduction is preserved by our translation. We need this result for both $\mathsf{E_{PI}}$ and $\mathsf{I_{PI}}$.[12]

LEMMA 5.2. *Let $\varphi_0, \ldots, \varphi_n, \psi$ be unnested formulas in the language of PA^1 with free variables $v_0, \ldots, v_m$, it follows that if $\varphi_0, \ldots, \varphi_n \vdash \psi$, then $\mathbb{N}(v_0), \ldots, \mathbb{N}(v_m), \varphi_0^{\mathcal{F}}, \ldots, \varphi_n^{\mathcal{F}} \vDash_{\mathsf{PI}} \psi^{\mathcal{F}}$. Further, it is PA^1-provable that if $\varphi_0, \ldots, \varphi_n \vdash \psi$ then $\mathsf{ACA}_0 \vdash$ "$\mathbb{N}(v_0), \ldots, \mathbb{N}(v_m), \varphi_0^{\mathcal{F}}, \ldots, \varphi_n^{\mathcal{F}} \vDash_{\mathsf{PI}} \psi^{\mathcal{F}}$".*

The first part of this Lemma is similar to (Linnebo, 2013, Thm. 5.4.). But he proves a version of this which does not restrict the quantifiers to a domain. The modification to our case is simple and so we omit the proof.

On its own a translation is not very interesting. However, a translation is an *interpretation* if the translations of the axioms of the interpreted theory can be proven in the interpreting theory.

DEFINITION 5.3. *Let T_A and T_B be L_A and L_B theories respectively, where a theory is a set of sentences not necessarily closed under deduction. A generalized translation $(\cdot)^{\mathcal{G}} : \mathcal{L}_A \rightarrow \mathcal{L}_B$ interprets T_A in T_B, if for all $\mathcal{L}_A$ unnested sentences χ:*

$$\mathsf{T}_A \vdash_{L_A} \chi \Rightarrow \mathsf{T}_B \vdash_{L_B} \chi^{\mathcal{G}} \tag{18}$$

It is a recursive interpretation if the collection of $\mathcal{L}_A$ and $\mathcal{L}_B$ formulas are recursive, T_A and T_B are also recursive, as is $(\cdot)^{\mathcal{G}}$, and there are recursive maps from proofs to proofs which witness the truth of Equations (9) and (18). Given a theory, if T extends PA^1, then say that the interpretation is T-verifiable if the recursive functions are provably total in T and if the universal closures of the arithmetized versions of 9 and 18 are provable in T.

So, the proofs of §3 and §4 show our translation is an interpretation of PA^1 in $\mathsf{E_{PI}}$. However, to show it is an interpretation in $\mathsf{I_{PI}}$ a certain level of caution is needed because $\mathsf{I_{PI}}$ does not have a background derivability relation. To resolve this, we take $\varphi_0, \ldots, \varphi_n \vdash_{L_{PI}} \varphi$ to be $\mathsf{ACA}_0 \vdash$ "$\varphi_0, \ldots, \varphi_n \vDash_{\mathsf{PI}} \varphi$", where this is as defined in Appendix B. And, of course $\mathsf{I_{PI}}$ is just as defined in (2) of §1, namely the set of sentences φ such that $\mathsf{ACA}_0 \vdash$ "$\vDash_{\mathsf{PI}} \varphi$". We then need to show the following:

LEMMA 5.4. *For all sentences φ in the language of PA^1, if $\mathsf{PA}^1 \vdash \varphi$ then $\mathsf{ACA}_0 \vdash$ "$\vDash_{\mathsf{PI}} \varphi^{\mathcal{F}}$". Further, it is PA^1-provable that if $\mathsf{PA}^1 \vdash \varphi$ then $\mathsf{ACA}_0 \vdash$ "$\vDash_{\mathsf{PI}} \varphi^{\mathcal{F}}$".*

[12] Recall that we formalized $\mathsf{I_{PI}}$ in ACA_0, and those interested in the nuts and bolts are directed to Appendix B.

https://doi.org/10.1017/S1755020320000349 Published online by Cambridge University Press

Proof. By Lemmas 3.2–3.6 and 4.2 and 4.3 we know that if φ is an axiom of PA^1 then $\mathsf{ACA}_0 \vdash$ " $\vDash_{\mathsf{PI}} \varphi^{\mathcal{F}}$". Assume $\mathsf{PA}^1 \vdash \varphi$ not an axiom, then there are n axioms of PA^1, $\varphi_0, \dots, \varphi_n$, such that $\varphi_0, \dots, \varphi_n \vdash \varphi$. Then as we can always take the universal closure of axioms and φ is a sentence it follows by Lemma 5.2 that $\mathsf{ACA}_0 \vdash$ "$\varphi_0^{\mathcal{F}}, \dots, \varphi_n^{\mathcal{F}} \vDash_{\mathsf{PI}} \varphi^{\mathcal{F}}$". Given that the axioms are PI valid, it follows that $\mathsf{ACA}_0 \vdash$ " $\vDash_{\mathsf{PI}} \varphi^{\mathcal{F}}$". □

This final piece gives us the proof of:

THEOREM 1.8. ii. *There is a generalized translation from the language of* PA^1 *to the second-order modal language with octothorpe that interprets* PA^1 *in* I_{PI}. *Further, this is a* PA^1*-verifiable generalized interpretation.*

To prove the first half of Theorem 1.8 we need to define formulas that pick out the numbers in PA^1 and E_{PI}. In PA^1 let $\tau_0(x) \equiv (x = 0)$ and $\tau_{n+1}(x) \equiv \exists y(\tau_n(y) \wedge Syx)$. In E_{PI} let $\sigma_0(x) \equiv (x = 0)$ and $\sigma_{n+1}(x) \equiv \Diamond\exists y \in \mathbb{N}(\sigma_n(y) \wedge Syx)$. Note that $(\tau_0(x))^{\mathcal{F}} \equiv (x = 0)^{\mathcal{F}} \equiv \sigma_0(x)$ and $(\tau_{n+1}(x))^{\mathcal{F}} \equiv (\exists y(\tau_n(y) \wedge Syx))^{\mathcal{F}} \equiv \Diamond\exists y \in \mathbb{N}((\tau_n(y))^{\mathcal{F}} \wedge Syx) \equiv \sigma_{n+1}(x)$. With this we can state the following preliminary Lemma; we omit the proof which is long but not illuminating:

LEMMA 5.5. *For every* $k \geq 0$ *and every unnested formula* $\theta(x_1, \dots, x_k)$ *in the signature of* PA^1 *and every k-tuple of natural numbers* $n_1, \dots, n_k$ *one has that :*

$$\mathbb{N} \models \theta(n_1, \dots, n_k) \quad \Longrightarrow \quad \vDash_{\mathsf{PI}} \forall x_1, \dots, x_k \in \mathbb{N} \left(\bigwedge_{i=1}^{k} \sigma_{n_i}(x_i) \to \theta^{\mathcal{F}}(x_1, \dots, x_k) \right) \tag{19}$$

In the case of $k = 0$, *this is to say: for every unnested sentence* θ *in the signature of* PA^1 *one has that*

$$\mathbb{N} \models \theta \quad \Longrightarrow \quad \vDash_{\mathsf{PI}} \theta^{\mathcal{F}} \tag{20}$$

Theorem 1.8.i follows from (20) of Lemma 5.5. This give us our proof of:

THEOREM 1.8. (i). *There is a generalized translation from the language of* PA^1 *to the second-order modal language with octothorpe that interprets* TA^1 *in* E_{PI}.

§6. Proof of Theorem 1.9. It has been shown by Linnebo & Shapiro (2019, §7) that the Linnebo translation cannot interpret comprehension because modalized comprehension requires the existences of a set of all possibly existing things. However, this leaves open the question of whether there is a different translation which can interpret PA^2. Here we will demonstrate that there is no translation from TA^2 to E_{PI} nor from PA^2 to I_{PI} by proving Theorem 1.9, our second main theorem. The first part of Theorem 1.9 follows from relatively simple Tarskian considerations:

THEOREM 1.9.I. *There is no generalized translation from the language of* PA^2 *to the second-order modal language with octothorpe that interprets* TA^2 *in* E_{PI}.

Proof. Assume for a contradiction that there is an interpretation $(\cdot)^{\mathcal{G}}$ that interprets TA^2 in E_{PI}. Note that as TA^2 is complete it follows that this is a faithful interpretation; i.e. if $\vDash_{\mathsf{PI}} \varphi^{\mathcal{G}}$ then $\mathbb{N} \vDash \varphi$. As E_{PI} is Π^1_1-definable it follows that there is a predicate P such that for all φ in the second-order modal language with octothorpe we have $\vDash_{\mathsf{PI}} \varphi$ if and only if $\mathbb{N} \vDash P(\text{“}\varphi\text{”})$. (Here we use quotation marks for Gödel numbering for both the language of PA^2 and the second-order modal language with octothorpe.) But

https://doi.org/10.1017/S1755020320000349 Published online by Cambridge University Press

then as generalized translations are recursive we can represent $(\cdot)^{\mathcal{G}}$ in $\mathbb{N}$ as g. It follows that $P(g(\text{“}\psi\text{”}))$, where ψ is in the language of PA^2, is a truth predicate for TA^2. But this contradicts Tarski's theorem. □

The proof of the second part of the theorem is trickier and requires Gödelian considerations. Recall the definition of T-verifiable generalized translation and interpretation from Definitions 5.1 and 5.3 in Section 5. There we proved that we have a PA^1-verifiable interpretation of PA^1 in I_{PI} by Lemma 5.4. Given that we defined $\mathsf{I}_{\mathsf{PI}} \vdash \varphi$ as $\mathsf{ACA}_0 \vdash \text{“} \vDash_{\mathsf{PI}} \varphi\text{”}$, that is $\mathsf{PA}^1 \vdash \forall\varphi[\text{“}\mathsf{PA}^1 \vdash \varphi\text{”} \rightarrow \text{“}\mathsf{ACA}_0 \vdash \text{“} \vDash_{\mathsf{PI}} \varphi^{\mathcal{F}}\text{””}]$. Here we show that there is no PA^2-verifiable interpretation of PA^2 in I_{PI}. We can write this as: there is no generalized translation $(\cdot)^{\mathcal{G}}$ from the language of PA^2 to the second-order modal language with octothorpe such that $\mathsf{PA}^2 \vdash \forall\varphi[\text{“}\mathsf{PA}^2 \vdash \varphi\text{”} \rightarrow \text{“}\mathsf{ACA}_0 \vdash \text{“} \vDash_{\mathsf{PI}} \varphi^{\mathcal{G}}\text{””}]$.

THEOREM 1.9.II. *There is no generalized translation from the language of* PA^2 *to the second-order modal language with octothorpe that* PA^2*-verifiably interprets* PA^2 *in* I_{PI}.

Proof. The systems Π^1_k-CA_0 are subsystems of PA^2 that have comprehension for Π^1_k formulas. As proofs are finite and so can only use finitely many instances of the comprehension schema any interpretation which is PA^2-verifiable will also be Π^1_k-CA_0-verifiable for some $k \geq 1$. Let $\varphi_1, \ldots, \varphi_n$ be a finite axiomatization of Π^1_k-CA_0 for some $k \geq 1$ (Simpson, 2009, pp. 303, 311-2). We will show, from the assumption that there is a Π^1_k-CA_0-verifiable translation $(\cdot)^{\mathcal{G}}$ from the languge of PA^2 to the second-order modal language with octothorpe that interprets PA^2 in I_{PI}, that Π^1_k-CA_0 proves its own consistency. This contradicts Gödel's second incompleteness theorem and so shows that no such $(\cdot)^{\mathcal{G}}$ can exist.

Note that $\mathsf{PA}^2 \vdash \varphi_1, \ldots, \varphi_n$ as all Π^1_k-CA_0 are subsystems of PA^2. We are assuming that $(\cdot)^{\mathcal{G}}$ interprets PA^2 in I_{PI}, so it follows that $\mathsf{ACA}_0 \vdash \text{“} \vDash_{\mathsf{PI}} \varphi_1^{\mathcal{G}}, \ldots, \varphi_n^{\mathcal{G}}\text{”}$. Let $\mathcal{A}$ be a model of Π^1_k-CA_0 for some k. So, we have $\mathcal{A} \vDash \text{“} \vDash_{\mathsf{PI}} \varphi_1^{\mathcal{G}}, \ldots, \varphi_n^{\mathcal{G}}\text{”}$. If $\mathcal{M}$ is the minimal model from Example 1.4 relative to $\mathcal{A}$ then we have then we have $\mathcal{A} \vDash \text{“}\mathcal{M} \vDash \varphi_1^{\mathcal{G}}, \ldots, \varphi_n^{\mathcal{G}}\text{.”}$

Now we show that $\mathcal{A} \vDash \neg Prv_{\varphi_1,\ldots,\varphi_n}(\psi \wedge \neg\psi)$, that is the consistency of Π^1_k-CA_0. Assume for a contradiction that $\mathcal{A} \vDash \exists\pi Prf_{\varphi_1,\ldots,\varphi_n}(\pi, \psi \wedge \neg\psi)$. Then as $(\cdot)^{\mathcal{G}}$ is a Π^1_k-CA_0-verifiable interpretation it follows $\mathcal{A} \vDash Prf_{\mathsf{ACA}_0}(\pi^{\mathcal{G}}, \text{“}\varphi_1^{\mathcal{G}}, \ldots, \varphi_n^{\mathcal{G}} \vDash_{\mathsf{PI}} \psi^{\mathcal{G}} \wedge \neg\psi^{\mathcal{G}}\text{”})$.

Recall that Π^1_1-CA_0 proves Σ^1_1-reflection for ACA_0 (cf. Simpson (2009) Theorem VII.6.9.(4) p. 298 and Theorem VII.7.6.(1) p. 305). As Π^1_1-$\mathsf{CA}_0 \subseteq \Pi^1_k$-$\mathsf{CA}_0$, this means that for any Π^1_1 statement ψ we know Π^1_k-CA_0 proves $Prv_{\mathsf{ACA}_0}(\psi) \rightarrow \psi$. For all ψ, we know that $\text{“} \vDash_{\mathsf{PI}} \psi\text{”}$ is Π^1_1 and similarly for the local derivability relation (see Appendix B). It follows that $\mathcal{A} \vDash \text{“}\varphi_1^{\mathcal{G}}, \ldots, \varphi_n^{\mathcal{G}} \vDash_{\mathsf{PI}} \psi^{\mathcal{G}} \wedge \neg\psi^{\mathcal{G}}\text{”}$ and as $\mathcal{A} \vDash \text{“}\mathcal{M} \vDash \varphi_1^{\mathcal{G}}, \ldots, \varphi_n^{\mathcal{G}}\text{”}$. It follows that $\mathcal{A} \vDash \text{“}\mathcal{M} \vDash \psi^{\mathcal{G}} \wedge \neg\psi^{\mathcal{G}}\text{”}$. And so $\mathcal{A} \vDash \text{“}\mathcal{M} \models \psi^{G}\text{”}$ and $\mathcal{A} \vDash \text{“}\mathcal{M} \models \neg(\psi^{G})\text{”}$. □

We have now shown the two main results set out in the introduction.

§7. Conclusion. We started with the worry that Hume's Principle had only infinite models and so any claim that it was analytic would mean that the claim that there are infinitely many objects is analytic. This worry has been noted before in the literature on neo-logicism, but little has been done to address it. Hale and Wright (2001) state that without this the neo-logicist project cannot even get off the ground:

https://doi.org/10.1017/S1755020320000349 Published online by Cambridge University Press

> To require of an acceptable abstraction that it should not be (even) weakly inflationary [that is require a countable infinity] would stop the neo-Fregean project dead in its tracks, before it even got moving (as it were). It will be clear that I think there is no good ground to impose such a requirement, and I shall not discuss it further. (Hale and Wright, 2001, pp. 417–8)

In this paper we have explored the potentially infinite as one way to address this worry. The move to the potentially infinite does not rid us of posited infinities. We still require there to be an infinity of worlds and an infinity of objects across the worlds. But these infinities are less metaphysically questionable. So, for example, while Putnam and Hodes objected to the positing of actual infinities they allowed for possible infinities. And one could always try to further avoid the commitment by adopting an instrumentalist attitude towards the metatheory.

We have shown that the theory of potentially infinite models interprets first-order Peano arithmetic or first-order true arithmetic, depending on the strength of our meta-language. But we cannot interpret the equivalent second-order arithmetic theory. The difficulty seems to be the nonexistence of a set of all the numbers across all the worlds. As our models are supposed to capture the idea of the potential infinite, we do not want the set of all the numbers across all the worlds to exist. It makes sense that the potential infinite does not capture the infinite progression of the natural numbers as well as actual infinity and this might go some way to explaining why we get the weaker first-order theory.

This allows a fuller understanding of the role of the potentially infinite in the foundation of mathematics. Unlike Hodes, we see that a certain amount of mathematics can be recovered, though some other story would need to be told about more advanced mathematics. It also offers evidence that the ontological commitments that come with Hume's Principle, and which make some reject the claim that its truth is analytic, cannot be avoided by moving to the modal setting if one wants full second-order Peano arithmetic. For in weakening our ontological commitments, we also weakened the mathematical theory which we can recover.

§Appendix A. Formal theories. Here we will spell out the theories other than E_{PI} and I_{PI} which are used in the proofs above. Unlike E_{PI} and I_{PI} none of these are modal theories, however, most are second-order theories.

The weakest theory we consider is first-order Robinson's Q. For a more complete reference see, for example, (Hájek & Pudlák, 1998, p. 28).

Definition A.1. Q *is the usual formalization of Robinson's arithmetic. It consists of the universal closure of the following axioms:*

$$s(x) \neq 0; \quad (Q1) \qquad s(y) = s(z) \to y = z; \quad (Q2)$$
$$x + 0 = x; \quad (Q3) \qquad x + s(y) = s(x + y); \quad (Q4)$$
$$x \times 0 = 0; \quad (Q5) \qquad x \times s(y) = (x \times y) + y. \quad (Q6)$$

Note that in the body of the text we do not use this formulation but rather one with relations instead than functions.[13] We have offered this formulation for readability. The

[13] We use a capital S for the relational successor and lower case s for the functional.

https://doi.org/10.1017/S1755020320000349 Published online by Cambridge University Press

relation formulation gives you the obvious translation of the above, plus an additional 6 axioms ensuring that the relations $S, +, \times$ are the graphs of functions.

We also consider the extensions of Q to PA^1 by the addition of the first-order induction schema, and PA^2 by the addition of the second-order induction axiom and Comprehension Schema. PA^1 is a first-order theory, but PA^2 is a second-order theory.

DEFINITION A.2. PA^1 *is* Q *plus the induction schema, where* φ *is a first-order formula:*

$$(\varphi(0) \wedge \forall x[\varphi x \rightarrow \varphi(s(x))]) \rightarrow \forall x \varphi(x) \qquad \text{(Induction Schema (IS))}$$

PA^2 *is* Q *plus the induction axiom and Comprehension Schema:*

$$\forall P[(P0 \wedge \forall x[Px \rightarrow P(s(x))]) \rightarrow \forall x Px] \qquad \text{(Induction Axiom (IS))}$$

$$\forall \bar{y}, \bar{Y} \exists X \forall x(X(x) \leftrightarrow \varphi(x, \bar{y}, \bar{Y})) \qquad \text{(Comprehension Schema (CS))}$$

In the Comprehension Schema φ *can be any formula of the language of* PA^2 *in which* X *does not occur free.*

Again in the body of the text we use the natural adaptation to the setting of relations rather than functions. There are also two theories we use that are second-order and between PA^2 and PA^1 in strength. They both restrict comprehension. So, we first need to define the formulas we restrict to:

DEFINITION A.3. (Simpson, 2009, I.3.1, p. 6) *An Arithmetical formula is a formula in the language of* PA^2 *which does not contain any set quantifiers, though it may contain free set variables.*

With this we can state ACA_0:

DEFINITION A.4. (Simpson, 2009, I.3.2, p. 7) ACA_0 *is* Q *plus the Induction Axiom and Arithmetical Comprehension:*

$$\forall \bar{y}, \bar{Y} \exists X \forall x(X(x) \leftrightarrow \varphi(x, \bar{y}, \bar{Y})) \quad \text{(Arithmetical Comprehension Schema (ACS))}$$

Where φ *has to be an arithmetical formula and* X *may not occur free.*

Note that as every formula of PA^1 is arithmetical, and ACA_0 contains the second-order induction axiom, every instance of the first-order induction schema is provable in ACA_0.

The next theories of arithmetic to be considered here are the Π^1_k-CA_0 which are used in the proof of Theorem 1.9. To define this theory, we first need to define Π^1_k (and Σ^1_k) formulas:

DEFINITION A.5. (Simpson, 2009, I.5.1, p. 16) *A* Π^1_1 formula *is a formula in the language of* PA^2 *of the form* $\forall X_1, \ldots, X_n \varphi$ *where* $X_1, \ldots, X_n$ *are set variables and* φ *is an arithmetical formula.*

A Σ^1_1 *formula is a formula in the language of* PA^2 *of the form* $\exists X_1, \ldots, X_n \varphi$ *where* $X_1, \ldots, X_n$ *are set variables and* φ *is an arithmetical formula.*

A Π^1_k *formula is a formula in the language of* PA^2 *of the form* $\forall X_1, \ldots, X_n \varphi$ *where* $X_1, \ldots, X_n$ *are set variables and* φ *is a* Σ^1_{k-1} *formula.*

A Σ^1_k *formula is a formula in the language of* PA^2 *of the form* $\exists X_1, \ldots, X_n \varphi$ *where* $X_1, \ldots, X_n$ *are set variables and* φ *is* Π^1_{k-1} *formula.*

The definition of Π^1_k-CA_0 is much like the definition of ACA_0, except that the restriction on the comprehension axiom is broadened to include all Π^1_k formulas:

https://doi.org/10.1017/S1755020320000349 Published online by Cambridge University Press

DEFINITION A.6. *(Simpson, 2009, I.5.2, p. 17)* Π^1_k-CA_0 *is* Q *plus the Induction Axiom and* Π^1_k *Comprehension:*

$$\forall \bar{y}, \bar{Y} \exists X \forall x (X(x) \leftrightarrow \varphi(x, \bar{y}, \bar{Y})) \qquad (\Pi^1_k \text{ Comprehension Schema } (\Pi^1_k \text{ CS}))$$

Where φ *has to be a* Π^1_k *formula and* X *may not occur free.*

We can define the intended model of these theories. Let $\mathbb{N}^1$ be $\{\omega, 0, s, +, \times\}$ where each term is interpreted as it is in the metatheory and $\mathbb{N}^2$ be $\mathbb{N}^1$ with $\mathcal{P}(\omega^n)$ as the domain of the second-order quantifiers. $\mathbb{N}^1$ is the intended model of Q and PA^1, while $\mathbb{N}^2$ is the intended model of PA^2, ACA_0, and Π^1_k-CA_0 for all k. As is well known, by Gödel's incompleteness theorems none of the theories we have seen so far are complete. We can define the complete theories of these models:

DEFINITION A.7. *Let* TA^1 *be* $\{\varphi \mid \mathbb{N}^1 \vDash \varphi\}$ *and* TA^2 *be* $\{\varphi \mid \mathbb{N}^2 \vDash \varphi\}$.

For the sake of completeness, we here define Hume's Principle (HP^2). This system is second-order also and consists of the cardinality principle displayed in Equation HP on page 4, the full Comprehension Schema, as in PA^2, and full comprehension for binary relations:

$$\forall \bar{y}, \bar{Y} \exists X \forall x, z (X(x, z) \leftrightarrow \varphi(x, z, \bar{y}, \bar{Y})) \quad \text{(Binary Comprehension Schema (BCS))}$$

Comprehension for binary relations is required because the definition of HP^2 quantifies over bijections and when spelt out fully this turns out to be the claim that there is a second-order binary relation which is the graph of a bijection between the two sets.

§Appendix B. Formal definition of I_{PI}. In the introduction we gave I_{PI} as the set $\{\varphi \mid \mathsf{ACA}_0 \vdash \text{“} \vDash_{\mathsf{PI}} \varphi\text{”}\}$. Here we will layout explicitly what we mean by defining the arithmetization of $\vDash_{\mathsf{PI}}$ in ACA_0.

It is important to note that the second-order variables in I_{PI} are taken to first-order variables in ACA_0. If all the first-order variables of I_{PI} are of the form x_i and all the second-order variables of I_{PI} are of the form Y_j then let all the first-order variables of ACA_0 be of the form x_i and Y_j, and the second-order variables of ACA_0 be of the form Z_v. In practice we will not stick to this strict distinction, but it can always be implemented by renaming the variables.

We do not restrict the domain of the first-order variables of I_{PI}; there is no need to pick out a subset of the domain of a model of ACA_0. However, the second-order variables of I_{PI} need to be restricted to codes for finite sets of numbers ordered by strict less than. This isn't difficult, we can simply borrow the coding found in the proof of incompleteness. A more complete explication can be found in (Simpson, 2009, Ch. 2.2). The second-order variables are required to be to some sequence $\pi(0)^{n_0} + \cdots + \pi(m)^{n_m}$ where $\pi(i)$ gives the ith prime and $n_0 < n_1 < \cdots < n_m$. Let $Seq(Y)$ be the name of the relation that ensures Y has the above properties. Further, let $nSeq(Y)$ mean that Y codes n-tuples of numbers. We will use this to code relations and relational variables. If x is the number of a sequence then let $[x]_i$ be the ith element and $ln(x)$ is the length of x.

We want to code PI models as sets of natural numbers. We know that we can always combine countably many countably infinite sets (just code n a member of the ith set as $2^i + 3^n$). As such we will just show how to code $W, R, D, \#, \boldsymbol{a}$ as separate sets of

https://doi.org/10.1017/S1755020320000349 Published online by Cambridge University Press

natural numbers. Further, with $R, D, \#, \boldsymbol{a}$ we will talk about pairs (x, y), this should be understood as standing for the code $2^x + 3^y$.

(B.1) Let W be infinite $(\forall x \in W \exists y \in W(y > x))$,[14]
(B.2) let R be such that

(a) for all $(i, j) \in R$ we have that $i, j \in W$,
(b) $\forall x \in W\ R(x, x)$ (reflexive),
(c) $\forall x, y, z \in W(R(x, y) \wedge R(y, z) \rightarrow R(x, z))$ (transitive),
(d) $\forall x, y \in W(R(x, y) \wedge R(y, x) \rightarrow x = y)$ (anti-symmetric),
(e) $\forall x, y \in W\ \exists z \in W(R(x, z) \wedge R(y, z))$ (directed),

(B.3) let D be such that

(a) $D(w, Y)$ implies that $w \in W$ and $Seq(Y)$,
(b) $\forall w \in W \exists Y \in Seq(D(w, Y) \wedge ln(Y) > 0)$ (every world has at least one element),
(c) D is the graph of a function from W to Seq,
(d) if $R(i, j)$ and $i \neq j$ and $D(i, X)$ and $D(j, Y)$ then $\exists u \forall v([X]_v \neq [Y]_u)$ (there is something in Y not in X) and $\forall v < ln(X) \exists u([X]_v = [Y]_u)$ (everything in X is in Y),

(B.4) let $\boldsymbol{a}$ be such that for each n there is exactly one x such that $\boldsymbol{a}(n, x)$ and if $\boldsymbol{a}(n, x)$ and $\boldsymbol{a}(m, x)$ then $n = m$, we then define $\#(Y, x)$ as $Seq(Y) \wedge \boldsymbol{a}(ln(Y), x)$.

Given a set of numbers $\mathcal{M}$ we will write $\mathcal{M} \in PIM$ to signify the set meets (B.1)–(B.4). We define sb (subset) as follows $Y \in sb(X)$ iff $Seq(Y) \wedge \forall i < ln(Y) \exists j([X]_j = [Y]_i)$. In defining the arithmetization note that we add free-variables for the model and the world, we will use $W_M, R_M, D_M, \#_M$, but these can be defined in terms of the model. So, if φ is a formula in the modal second-order language with octothorpe we translate it to some $\psi(w, W_M, R_M, D_M, \#_M)$ in the language of arithmetic. We define the arithmetization as follows:

$$(x_i = x_j)^* \equiv x_i = x_j \tag{21}$$

$$(x_i = \#Y_j)^* \equiv \#_M(Y_j, x_i) \tag{22}$$

$$(Y_j x_i)^* \equiv \exists u(x_i = [Y_j]_u) \tag{23}$$

$$(\forall x \varphi)^* \equiv \forall x(\exists Y \in Seq(D_M(w, Y) \wedge \exists u(x = [Y]_u)) \rightarrow (\varphi)^*) \tag{24}$$

$$(\forall Y \varphi)^* \equiv \forall Y \in Seq(\exists X \in Seq(D_M(w, X) \wedge Y \in sb(X)) \rightarrow (\varphi)^*) \tag{25}$$

$$(\forall P^n \varphi)^* \equiv \forall P^n \in nSeq \tag{26}$$

$$(\exists X \in Seq(D_M(w, X) \wedge \forall(x_1, \ldots, x_n) \in P^n(\bigwedge_{1 \leq i \leq n} \exists j[X]_j = x_i)) \rightarrow (\varphi)^*)$$

$$(\Box \varphi)^* \equiv \forall s \in W_M(R_M(w, s) \rightarrow (\varphi)^*[w/s]) \tag{27}$$

[14] Recall that our definition demanded that our set of worlds be countable. We cannot capture this in ACA_0 in the sense that ACA_0 has none standard models but we will have that we do not have more worlds than ACA_0 thinks there are natural numbers, which is sufficent for the role this plays in the proofs.

https://doi.org/10.1017/S1755020320000349 Published online by Cambridge University Press

where we commute over the logical connectives. This means that every formula arithmetized is arithmetical as defined in Appendix A. For example, $\Box\forall v\Diamond\exists Z(v = \#Z)$ becomes

$$\forall s \in W_M(R_M(w,s) \rightarrow \forall v(\exists Y(D_M(s,Y) \wedge \exists u(v = [Y]_u) \rightarrow \exists s' \in W_M(R_M(s,s') \wedge \exists Z \in Seq(\exists X(D_M(w,X) \wedge Z \in sb(X) \wedge \#_M(Z,v))))). \tag{28}$$

Note "$\vDash_{\mathsf{PI}} \varphi$" means $\forall M \in PIM \forall w \in W_M(\varphi)^*$. It follows that this is then a Π^1_1 formula. Hence, if one were proceeding very formally, we would define I_{PI} as the set of all the φ such that $\mathsf{ACA}_0 \vdash \forall M \in PIM \forall w \in W_M(\varphi)^*$.

Acknowledgements I would like to thank the audience at the Logic Colloquium 2016 in Leeds and the UCI Logic Seminar 2017 for their questions and comments, and Tim Button, Jeremy Heis, Richard Mendelsohn, Stella Moon, Sean Walsh, Kai Wehmeier, and the anonymous reviewers for their helpful feedback.

BIBLIOGRAPHY

Bell, J. L. (1999). Frege's theorem in a constructive setting. *Journal of Symbolic Logic*, **64**(2), 486–488.

Benacerraf, P. (1965). What numbers could not be. *Philosophical Review,* **74**(1), 47–73.

Boolos, G. (1998). *Logic, Logic, and Logic*. Cambridge, MA: Harvard University Press.

Burgess, J. P. (2005). *Fixing Frege*. Princeton, NJ: Princeton University Press.

Button, T., & Walsh, S. (2018). *Philosophy and Model Theory*. Oxford, UK: Oxford University Press.

Cook, R. T. (2007). Introduction. In Cook, R. T., editor. *The Arché Papers on the Mathematics of Abstraction*. The Western Ontario Series in Philosophy of Science, Vol. 71. Berlin, Germany: Springer, pp. xv–xxxvii.

Demopoulos, W. (1994). Frege and the rigorization of analysis. *Journal of Philosophical Logic*, **23**(3), 225–245.

Feferman, S. (2005). Predicativity. In Shapiro, S., editor. *The Oxford Handbook of Philosophy of Mathematics and Logic*. Oxford, UK: Oxford University Press, pp. 590–624.

Fitting, M., & Mendelsohn, R. L. (1998). *First-Order Modal Logic*. Dordrecht, Netherlands: Kluwer Academic Publishers.

Frege, G. (1884). *Die Grundlagen der Arithmetik: eine logisch mathematische Untersuchung über den Begriff der Zahl* (translated as *The Foundations of Arithmetic: A logico-mathematical enquiry into the concept of number*, by J.L. Austin, Oxford: Blackwell, second revised edition, 1974). Breslau: W. Koebner.

———. (1893). *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet* (translated as *Basic Laws of Arithmetic: Derived using concept-script* by P. Ebert and M. Rossberg (with C. Wright), Oxford, UK: Oxford University Press, 2013). Jena: Verlag Hermann Pohle.

Hájek, P., & Pudlák, P. (1998). *Metamathematics of First-Order Arithmetic*. Perspectives in Mathematical Logic. Berlin, Germany: Springer.

Hale, B, & Wright, C. (2001). *The Reason's Proper Study*. Oxford, UK: Oxford University Press.
Heck, R. K. (1993). The development of arithmetic in Frege's Grundgesetze der Arithmetik. *Journal of Symbolic Logic*, **58**(2), 579–601 (originally published under the name "Richard G. Heck, Jr").
Hodes, H. (1984). Logicism and the ontological commitments of arithmetic. *The Journal of Philosophy,* **81**(3), 123–149.
Hodes, H. (1990). Where do the natural numbers come from? *Synthese,* **84**(3), 347–407.
Hodges, W. (1993). *Model Theory*. Cambridge, MA: Cambridge University Press.
Kim, J. (2015). A logical foundation of arithmetic. *Studia Logica,* **103**(1), 113–144.
Kocurek, A. W. (2016). The problem of cross-world predication. *Journal of Philosophical Logic*, **45**(6), 697–742.
Linnebo, Ø (2013). The potential hierarchy of sets. *The Review of Symbolic Logic*, **6**(2), 205–228.
———. (2018). *Thin Objects: An Abstractionist Account*. Oxford, UK: Oxford University Press.
Linnebo, Ø., & ShapiroS. (2019). Actual and potential infinity. *Noûs*, **53**(1), 160–191.
van Heijenoort, J. (1967). *From Frege to Gödel : A Source Book in Mathematical Logic, 1879–1931*. Cambridge, MA: Harvard University Press.
Linnebo, Ø., & ShapiroS. (2019). Actual and potential infinity. *Noûs*, **53**(1), 160–191.
Mostowski, M. (2001). On representing concepts in finite models. *Mathematical Logic Quarterly*, **47**(4), 513–523.
Parsons, C. (1983). Sets and modality. *Mathematics in Philosophy: Selected Essays*. Ithaca, NY: Cornell University Press, pp. 298–341.
Putnam, H. (1967). Mathematics without foundations. *Journal of Philosophy*, **64**(1), 5–22. reprinted in Putnam 1979.
Shapiro, S., & Linnebo, Ø. (2015). Frege meets Brouwer. *Review of Symbolic Logic,* **8**(3), 540–552.
Simpson, S. (2009). *Subsystems of Second Order Arithmetic. Perspectives in Logic*. Cambridge, MA: Cambridge University Press.
Stanley, J. (1997). Names and rigid designation. In Hale, B ., and Wright, C., editors. *A Companion to the Philosophy of Language.* Malden: Blackwell, pp. 555–585.
Studd, J. P. (2016). Abstraction reconceived. *The British Journal for the Philosophy of Science*, **67**(2), 579–615.
Urbaniak, R. (2016). Potential infinity, abstraction principles and arithmetic (leniewski style). *AXIOMS*, **5**(2), 20.
Walsh, S. (2016). The strength of abstraction with predicative comprehension. *The Bulletin of Symbolic Logic,* **22**(1), 105–120.
Whitehead, A., & RussellB. (1910). *Principia Mathematica*, Vol. 1. Cambridge, MA: Cambridge University Press.
Williamson, T. (2013). *Modal Logic as Metaphysics*. Oxford, UK: Oxford University Press.

DEPARTMENT OF LOGIC AND PHILOSOPHY OF SCIENCE
UNIVERSITY OF CALIFORNIA, IRVINE
IRVINE, 92617 CA, USA
E-mail: will.stafford@uci.edu

https://doi.org/10.1017/S1755020320000349 Published online by Cambridge University Press