Journal for the Philosophy of Mathematics

Volume 1 - 2024

FIRENZE
UNIVERSITY
PRESS

Journal for the Philosophy of Mathematics

Volume 1 - 2024

Firenze University Press

Journal for the Philosophy of Mathematics
Published by
Firenze University Press – University of Florence, Italy
Via Cittadella, 7 - 50144 Florence - Italy
https://riviste.fupress.net/index.php/jpm/index

Enumerative Induction in Mathematics

Alan Baker

1 Introduction

In 1919, Hungarian mathematician George Polya was pursuing questions concerning prime factorization and began noting down, for each of the first few natural numbers, whether a given number had an even number of prime factors or an odd number of prime factors. What Polya was counting was not distinct prime factors, but just the raw number of prime factors. So, for example, $10 = 2.5$ has an even number of prime factors, 11 has an odd number, and $12 = 2.2.3$ has an odd number. Next, Polya recorded, for each given number, n, how many numbers in the set $\{1, 2, 3, \ldots n\}$ have an odd total number of prime factors and how many have an even total number. Looking at all the numbers up to 100, Polya noticed that at every point at least half of members of the set $\{1, 2, 3, \ldots n\}$ had an odd total number of prime factors. Intrigued, Polya extended his calculations up to $n = 1{,}500$ and found that the pattern continued to hold. At this point, Polya conjectured in print that this holds universally. In other words, for all n, at least half of the numbers less than n have an odd total number of prime factors. This came to be known as *Polya's Conjecture* (PC).

Polya presumably made the conjecture because he believed that it was likely to be true. What grounds did he have for this belief? According to the narrative outlined above, Polya's belief in the truth of PC was based exclusively on enumerative induction. 1,500 instances from a larger domain were examined and found to fit the hypothesis, and on this basis the hypothesis was conjectured to hold universally. As such, this seems like a particularly pure case of enumerative induction, with few extraneous complicating factors. But was Polya *rational* in taking the results of the first 1,500 cases to lend support to the universal hypothesis in this way? More generally, can enumerative induction lend genuine support to mathematical claims? It is these questions that I shall be taking up in what follows.

2 Scene Setting

Before proceeding it may be helpful in clarifying the scope of our inquiry to distinguish the questions I am asking from some other, related (and also interesting) questions, and also to make explicit several presuppositions that I will be taking for granted.

Firstly, I am not asking whether enumerative induction in mathematics can yield *knowledge* of general mathematical claims. My focus here is on justified belief, not knowledge, and I will be taking no stand on the potential link between enumerative induction and knowledge. Secondly, I am not asking whether enumerative induction in mathematics can (or should) ever be on a par with *deductive proof*. As it pertains to mathematical methodology, everything I say is compatible with deductive proof continuing to be the gold standard for demonstrative mathematical reasoning.[1] Thirdly, I am not asking whether enumerative induction can provide *compelling* grounds for belief in a general mathematical claim. Rather, I am interested in whether enumerative induction can achieve the more modest goal of providing some positive support.

As for presuppositions, the work being done in this paper will be carried out against the background of a general anti-scepticism about induction. In other words, I will be presuming that enumerative induction in the empirical sciences is generally (other things being equal, *ceteris paribus*) a rationally acceptable tool for acquiring justified belief. Also, although I will continue to talk sometimes about "enumerative induction in mathematics" my focus will almost exclusively be on enumerative induction in number theory. The examples I will be discussing will all involve claims about the totality of the natural numbers. Putting together these two points, the question I asked at the end of Section 1 can be precisified as follows: is there something distinctively problematic about using enumerative induction to bolster beliefs in arithmetical claims?

3 An Argument Against Enumerative Induction in Mathematics

In his 2007 paper, "Is There a Problem of Induction for Mathematics?," Alan Baker rejects the idea that enumerative induction has force for mathematical claims. More specifically, Baker argues for the following two theses, one normative and one descriptive:[2]

[EIM Norm] Enumerative induction *ought not* to increase confidence in a universal mathematical generalization (over an infinite domain).

[EIM Desc] Enumerative induction *does not* (in general) increase confidence in a universal mathematical generalization (over an infinite domain).

Baker is not an inductive sceptic in a more general sense, so his argument for EIM Norm rests on features that he takes to be peculiar to the mathematical case. These distinctive features can also be framed as *disanalogies* between pure mathematics and the empirical sciences.

[1] Relatedly, I am not advocating for any relaxation of the notion of theoremhood, whereby a mathematical conjecture could be elevated to the status of "theorem" given sufficient enumerative inductive support.

[2] Baker (2007, p.73).

In increasing order of importance, the three main disanalogies that Baker points to are the following:

(i) The domain of number theory is infinite.

The challenge here for enumerative induction is clear enough. When dealing with an infinite domain, such as the natural numbers, however big our finite sample of instances of a given universal claim, it will still only represent an infinitesimally small portion of the whole domain. What is less clear is how much difference this makes for enumerative induction in mathematics as compared to the empirical sciences. Firstly, it may be that the universe is in fact infinitely large, in which case fully general scientific claims would also have an infinite domain. Secondly, even if the universe is not infinitely large, most general empirical claims seem to presuppose some sort of indefinitely open-ended domain, and this may be not much different in its implications for enumerative induction than an absolutely infinite domain.[3] In summary, while the presence of an infinite domain certainly poses a problem, it is not enough on its own to justify abandoning enumerative induction in mathematics while retaining it in the empirical sciences.

(ii) The domain of number theory is non-uniform.

Here Baker draws explicitly from Frege's discussion of induction in the *Grundlagen*. The following three quotes from that work give the flavor of Frege's position:

> [In mathematics] the ground [is] unfavorable for induction; for here there is none of that uniformity which in other fields can give the method a high degree of reliability.
>
> An even number can be divided into two equal parts; an odd number cannot; three and six are triangular numbers, four and nine are squares, eight is a cube, and so on.
>
> In ordinary inductions we often make good use of the proposition that every position in space and every moment in time is as good in itself as every other. . . . Position in the number series is not a matter of indifference like position in space.[4]

Following Frege, Baker argues that the domain of the natural numbers is *non-uniform*. In fact, the above quotes suggest two different scales on which this non-uniformity might manifest. At the scale of individual numbers, if we arbitrarily pick two separate numbers, m and n, just because m has property P gives us no *prima facie* reason to think that n also has property P. (For example, just because m is a triangular number gives us no reason to think that n will also be

[3] Note that even scientific claims that have an implicit spatial limitation – for example, many biological claims that are about how genetic code, metabolism, etc. work among organisms found on Earth – are likely to be temporally open-ended.

[4] Frege (1884, Section 10).

a triangular number.) But there is also, potentially, non-uniformity at the scale of different segments of the number line. For example, consider the sequence of numbers between 200 and 300 (call this S_1), and the sequence of numbers between 10^{10} and $10^{10} + 100$ (call this S_2). Next consider some property, P*, that holds for every member of S_1, or which holds collectively for the set. Each of the two sets S_1 and S_2 consists of 100 consecutive numbers. So should we expect P* also to hold for S_2? Not necessarily, and in fact there are many properties that we know to vary according to where in the number line we are sampling. For example, every number in S_1 has eight or fewer prime factors, but this is not true of every number in S_2. And at the level of collective properties, the density of primes, of square numbers, of Fibonacci numbers, and so on are all significantly lower in S_2 than in S_1.

The fact that the properties of one number are not a reliable guide to the properties of other numbers, and similarly for groups of numbers, is undeniably a challenge for the effectiveness of enumerative induction in supporting general claims about all numbers. What is less clear is whether it is enough on its own to justify Frege's wholesale rejection of enumerative induction in an arithmetical context. After all, there is plenty of variability in many of the empirical domains in which enumerative induction is employed, yet this is not considered a fatal flaw. If the issue is simply non-uniformity, then it can be counteracted by selecting a diverse array of instances to examine. If the conjectured hypothesis holds across this diverse array, which will presumably exemplify the same kind of non-uniformity as the domain as a whole, then this enumerative inductive evidence may still provide genuine support for the hypothesis.

(iii) We can only sample from among the very small numbers.

For Baker the infinite size and the non-uniformity of the domain of the natural numbers are not enough on their own to undermine the effectiveness of enumerative induction in mathematics. However, they do become enough when combined with a third feature, which concerns how and where we are sampling the domain.

The basic point is straightforward enough. Whether we are talking about the instances initially collected by Polya in connection with his conjecture, PC (i.e. all the numbers up to 1,500), or the current range of instances checked in connection with the Goldbach Conjecture (i.e. currently, all the even numbers up to about 4×10^{18}), all the instances that we can ever feasibly collect for any general arithmetical conjecture are in an important sense *small.* In his (2007), Baker introduces the term "minute" to make this latter notion more precise, defining a natural number to be *minute* if it can be written down (by us) in ordinary decimal notation, including exponentiation (but not iterated exponentiation). Baker then states that we are practically limited to searching among the minute numbers for instances of general arithmetical claims.

So what? Even if we are in practice restricted in this way, what relevance does this have to the effectiveness of enumerative induction? This is a fair question, and – as with the first two distinctive features of mathematics discussed earlier – it could justifiably be claimed that on its own it is not a serious problem. However it becomes a problem, or so Baker argues, when combined with the two earlier features. Baker's core argument can be reconstructed as follows:

(1) All instances of general arithmetical conjectures that we can feasibly check are drawn from among the minute numbers.

(2) Hence, there are an infinite number of non-minute numbers, none of which we can ever sample.

(3) Size is known to make a difference to the properties of individual numbers and of sets of numbers.

(4) Hence, any sample used as a basis for enumerative induction in arithmetic is unavoidably biased.

(5) Hence, enumerative induction in arithmetic carries (and ought to carry) no weight.

Returning to the example that we began with, the above argument suggests that Polya's examination of the first 1,500 natural numbers provides no substantive support for his conjecture, PC, because there is no reason to think that this initial sample is representative of the number line as a whole.

Baker's view on induction in mathematics can be seen as a targeted form of inductive scepticism. Paseau (2021) uses the term "size scepticism" to refer to this position, and helpfully distinguishes three different types of size scepticism. According to Paseau's classification scheme, Baker (2007) is a *u-sceptic* (the "u" here stands for "unrepresentative").[5] Sampling only from the minute numbers amounts to not varying the sampling space along an axis (in this case size) that is *known* to make a difference. So, in this sense, any basis for enumerative induction in mathematics is (unavoidably) *unrepresentative.*

It is worth noting that Baker's conclusion here is quite radical (though not historically unprecedented, as seen from the remarks of Frege quoted above), for his anti-inductive thesis is not merely that enumerative induction fails to provide compelling evidence for general arithmetical claims, but that it provides no supporting evidence whatsoever. This 'no-support' thesis can best be made sense of as arising not just from the unrepresentativeness of our sample set but also from the infinite number of objects in the domain (as alluded to in step (2) of Baker's core argument). In drawing exclusively from the minute numbers, we end up with a sample set that is unrepresentative of the domain as a whole and is also an infinitesimally small portion of that whole. If either of these two obstacles was removed then there would be the possibility of enumerative

[5] This is to be distinguished from *s-scepticism* (size scepticism), according to which small numbers are distinctively problematic. i.e. sampling 1,000 non-small numbers would provide better evidence than sampling 1,000 small numbers. And it is also to be distinguished from *c-scepticism* (comparative scepticism), according to which a set of larger numbers provides better evidence than a set of smaller numbers.

inductive evidence yielding substantive support for general conjectures. A representative sample would give grounds for projecting the partial result across the whole domain. And a non-infinitesimal sample would raise our subjective probability for the general conjecture by verifying a substantive proportion of cases.

4 Bolstering the Biased-Sample Claim

Most of the criticism of Baker's core argument for inductive scepticism with respect to mathematics has focused on undermining the intermediate conclusion, (4), that our sampling from the domain of the natural numbers is unavoidably *biased*.[6] Implicit in such attacks is the presumption that the final step in the argument, from sample bias (step (4)) to lack of inductive support (step (5), is unproblematic. My own approach will be rather different. I think that Baker's argument for sample bias is convincing. However, I want to resist the inference that sample bias automatically undermines the force of enumerative induction. To this end, my strategy will be as follows. Firstly, I will sharpen and defend the biased-sample thesis by focusing on an initial subset of the natural numbers that is radically smaller than the set of minute numbers. Secondly, I will argue that the particular kind of bias that afflicts this initial subset actually serves to *strengthen* inductive inferences that are based on it. The counterintuitive-sounding takeaway is that bias in our sampling of the natural numbers provides a rational boost to enumerative induction applied to general arithmetical claims in a way that helps to overcome the challenges of generalizing over a demonstrably infinite domain.

The first task is to formulate and defend a more circumscribed version of the biased-sample claim. I shall begin by zeroing in (no pun intended) on a much smaller initial subset of the natural numbers than features in Baker's core argument. Baker focuses on the "minute numbers," where a natural number is defined as *minute* if it can be written down by us in ordinary decimal notation, including exponentiation (but not iterated exponentiation). It is worth noting that, despite the label, the minute numbers include a lot of numbers that are very far from being small in any ordinary sense. For example, current estimates put the number of atoms in the universe at something on the order of 10^{80}. Included among the minute numbers are numbers vastly larger than this, such

[6] See e.g. Walsh (2014), Waxman (2017). One common line of objection is that in gathering confirming instances for a general scientific conjecture we are also stuck with potentially biased samples, since all instances that we observe must be "close" to us (in space and in time). But induction in empirical contexts is (by presumption) acceptable. Hence it cannot be this "bias" that leads to problems in the mathematical context. However, this objection ignores the fact that the "close to us" bias is very different from the "close to the beginning of the number line" bias. Firstly, being "small" is an objective property of numbers, while being "close to us" is (manifestly) observer-relative. Secondly, we have *antecedent* reason to believe that smallness can make a systematic difference to the behavior of numbers, whereas we have no antecedent reason to believe that being close to us in space and/or time makes a systematic difference to the behavior of physical objects.

as 10^{800}, 10^{8000}, and so on. Presumably, Baker's reason for placing such a relaxed bound on the minute numbers is to make absolutely sure that all putative inductive evidence for unproven general conjectures consists exclusively of instances that count as minute. However, for current purposes this is overkill, and in fact the liberal characterization of minute number has a negative impact on Baker's argument. I consider these two points separately below.

How is this overkill? The fact is that in the vast majority of actual cases where mathematicians gather instances of a general arithmetical conjecture, these instances are not especially large. Historically, when instances were calculated by hand, these rarely exceeded a few thousand (in number, and in magnitude), and this persisted until the advent of the digital computer.[7] Once computers could be used to speed up this checking process, the sampling size massively increased. But even now, the limitations of computing speed and memory place fairly tight bounds on how many cases can feasibly be calculated. Consider the oft-cited example of the Goldbach Conjecture, which has been checked by computer for all even numbers up to around 4×10^{18}. This is among the larger sample sets for any currently open arithmetical conjecture, yet it is nowhere close to pushing the bounds of the minute numbers (which, recall, include anything we can feasibly write down using decimal notation and non-iterated exponentiation, so 10^{180} , 10^{1800}, and 10^{18000} are all minute numbers). The upshot is that we could focus on a much, much smaller set of numbers than the minute numbers and still capture nearly all actual cases of putative enumerative induction in mathematics.

Even if the above point is true, it might seem that there is still no harm in using the minute numbers as a basis for Baker's argument. However, it turns out that there is a clear downside, namely that referencing an excessively large set of numbers risks undermining the biased-sample claim itself. There are a couple of ways in which this happens.

Recall that the basis of Baker's claim that our sampling of the natural numbers is (unavoidably) biased is that the sample space is unrepresentative of the numbers as a whole. So if the focus is on the minute numbers, then this amounts to the claim that the minute numbers behave differently than the non-minute numbers. The (Frege-inspired) intuition that position along the number line matters to what properties a number is likely to have was motivated, back in Section 3, by comparing the sequence of one hundred numbers between 200 and 300 (call this S_1) with the sequence of one hundred numbers between 10^{10} and $10^{10} + 100$ (call this S_2), and noting the sharp difference in the relative occurrence of properties such as being prime, square, Fibonacci, and so on. Notice, however, that both of these sequences fall squarely within the set of minute numbers, so this observation does nothing to support the claim that the minute numbers are themselves systematically different from the non-minute numbers. If instead we take a set of one hundred numbers from further along the sequence of minute numbers and compare it with a set of one hundred non-minute num-

[7] The Polya Conjecture example discussed in Section 1, in which cases were checked up to 1,500, is very typical in this respect.

bers, it is much less clear that the 'systematic difference' claim holds up.[8] The basic problem is that the set of minute numbers includes numbers of such magnitude that maintaining that all minute numbers are unrepresentative becomes much less plausible.

A second problem with the excessive magnitude of the minute numbers is that it becomes harder to find examples where a general conjecture has held for all minute numbers but ended up failing for some non-minute number. In his (2007), Baker comes up with just one putative example where this may happen, and even this is conditional on the unlikely eventuality that the Riemann Hypothesis is false. Under this assumption, the upper bound of the point at which the logarithmic density function is eventually exceeded is given by the second Skewes number, which is a non-minute number. However, as D'Alessandro points out, this lone example no longer holds up.[9] A result of Saouter and Demichel, proved in 2010, shows that the upper bound is at most a little over 10^{316}, which is well within the realm of the minute numbers. The difficulty of finding actual examples of conjectures for which the minute numbers behave differently from the non-minute numbers puts further pressure on Baker's thesis that the minute numbers are "unrepresentative" of the numbers as a whole.

I propose to cut through the difficulties associated with the minute numbers by restricting attention to a drastically smaller initial segment of the natural numbers, and focusing on the biased-sample thesis as applied to this smaller subset. Consider the natural numbers from 1 to 1,000. At risk of multiplying terminology beyond necessity, I shall refer to the numbers in this set as *tiny numbers.* And I hope to make plausible the following two theses associated with this set:

(A) A sample of positive instances selected only from among the tiny numbers is biased, and is unrepresentative of the domain of natural numbers as a whole.

(B) A sample of positive instances of a general conjecture, C, that includes all of the tiny numbers provides substantive support for C, other things being equal, *in virtue* of the fact that the sample is biased.

The broad arguments for the more general thesis (A) have already effectively been made back in Section 3. Now that we are restricting attention to just the first 1,000 numbers, it is very plausible to maintain that sequences of numbers from this set look very different in their individual properties, and distribution of properties, than sequences of much larger numbers. In the next section, I will turn to defending thesis (B), firstly by highlighting three specific ways in which the tiny numbers are unrepresentative, and then by showing how each of these biases actually *increases* the level of inductive support that tiny numbers provide.

[8] For example, consider the (minute) number $M = 10^{1,000,000}$, and compare an arbitrary set of minute numbers, $\{M + 1, M + 2, \ldots, M + 100\}$, with an arbitrary set of non-minute numbers, $\{10^M + 1, 10^M + 2, \ldots, 10^M + 100\}$.

[9] D'Alessandro (2021, p.33).

5 From Bias to Inductive Support (I): Unrepresentative Representativeness

Bias in a sample is not usually considered to be a positive feature for inductive inference. However, certain specific types of bias can boost inductive support, or so I shall argue. Counterintuitive though this sounds, it is easy enough to think up examples where this happens even in a non-mathematical context. Consider some claim about the upper physiological limits on human strength or endurance, and imagine that we test this claim by measuring the capacities of 1,000 professional athletes and finding that they all fall within the conjectured limit. Clearly this is a biased sample, since professional athletes are not representative of human beings as a whole. However, it is equally clear that in this case the bias ought to strengthen our belief in the physiological limit conjecture more than if we had picked 1,000 people at random from the general population. This is because it is *prima facie* more likely for a professional athlete to exceed a postulated performance limit. So here the bias of the sample serves to boost the inductive support that it provides.

Returning to the mathematical case, I will highlight three features that manifest differently among the tiny numbers in comparison to arbitrary larger numbers. For each feature, I will begin with a characterization of what the feature is, then show how the feature plays out differently for the tiny numbers, and finally argue that this difference boosts the inductive force of samples involving the tiny numbers.

The first feature involves *significant properties.* Every natural number has an infinite number of mathematical properties, but some of these properties are more significant than others.[10] What do I mean here by “significant”? I am not going to try to give a precise definition of this term, partly because I do not know how to do so and partly because the larger argument I shall be giving does not require it. For present purposes, I shall proceed with two complementary approaches to characterizing the notion of significance: a loose definition of significance, and then some contrasting cases of *in*significant properties.

Starting with the loose definition, we shall count a mathematical property as *significant* to the extent that it features in the statements and proofs of important mathematical results, and to the extent that it is systematically linked to other significant properties.[11] The boundary between significant and insignificant properties is vague, and significance is itself plausibly a matter of degree.[12]

[10] For example, each natural number, n, has the (infinitely large) set of properties of being $< n+1, < n+2, < n+3$, etc.

[11] In his *Mathematician’s Apology*, G.H. Hardy spends some time exploring the notion of “seriousness” as a property of certain mathematical theorems and results. Hardy writes: “A ‘serious’ theorem is a theorem which contains ‘significant’ ideas and I suppose that I ought to try to analyse a little more the qualities that make a mathematical idea significant.” (Hardy, 1940, p.21) Hardy ends up linking significance (of mathematical ideas) both to generality and to depth. For an interesting recent discussion of Hardy’s notion of seriousness, and how initial judgements of non-seriousness may be overturned by subsequent developments in mathematics, see Weisgerber (2024).

[12] I will return to discuss the issue of *degree* of significance in Section 6 below.

Nonetheless, it seems clear that some mathematical properties are significant *simpliciter* and others are not. Any list of canonical significant mathematical properties would presumably include properties such as being prime, being even, being square, and being a Fibonacci number.[13]

As a second aid to drawing the above distinction, it may be helpful to enumerate a few ways in which number properties can fail to be significant. The following list of kinds of insignificant property is obviously not meant to be exhaustive, but it provides a few examples:

(i) properties that relate to non-mathematical facts
e.g. the property of being the number of weeks in a year

(ii) properties that are representation-dependent
e.g. the property of having the decimal digits sum to a prime number

(iii) properties that are highly gerrymandered (or disjunctive)
e.g. the property of being either a square number or a Fibonacci

(iv) properties that are arbitrarily specific
e.g. the property of having exactly 23 distinct prime factors

Most of the canonical significant properties decrease in frequency as numbers get larger. Sometimes this decrease is fairly gradual (as with prime numbers) and other times it is more drastic (as with perfect numbers). This general pattern suggests that the tiny numbers are biased in manifesting a greater frequency of significant properties, but we need to be careful in how this claim is formulated. It is not the case, for instance, that there is a greater density of numbers with significant properties among the tiny numbers. The reason why not is that *every* number has (some) significant properties. How so? Because in addition to the kinds of properties mentioned above, there are also significant properties that do not decrease in frequency among larger numbers. Typically, these are significant properties whose complements are also significant.[14] An example of a significant property whose frequency does not decrease with number size is the property of being an odd number. This property is significant because many other properties are linked to it, and its complementary property, being an even number, is also a significant property, and also does not decrease in frequency as number size increases. Another example is the property of being composite. This is a significant property whose frequency actually increases with number size (since its complementary – and also significant – property, primality, decreases in frequency). Every number is either even or odd, and every number (apart from 1) is either prime or composite, so every number has significant properties.

The upshot of the above discussion is that the tiny numbers are not different from sequences of larger numbers in having fewer elements with significant

[13] For discussion of the – potentially related – notion of "mathematical natural kind," see Corfield (2004) and Lange (2015).

[14] This is not the case for the majority of significant properties. For example properties such as '_ is not a square number' and '_ is not a Fibonacci number' are not significant.

properties. However, it is nonetheless true that most significant properties decrease in frequency with number size. As a consequence, if we compare the tiny numbers with a sequence of 1,000 consecutive much larger numbers, we should expect a greater *range* of significant properties to be instaConsider, for example, the properties of being a square number and of being a prime number. The difference between consecutive squares is $(n+1)^2 - n^2 = 2n+1$, so once we get above 500^2, there will be sequences of 1,000 consecutive numbers none of which has the property of being square. For prime numbers, the density of primes less than n is well approximated by $1/\log n$, and among numbers as 'small' as 10^{15}, sequences of 1,000 consecutive composite numbers are known to occur.[15] For both of these properties, being square and being prime, we don't have to go very much further along the number line beyond the tiny numbers before the chances of either being instantiated in a given 1,000-number sequence is extremely low. The vast majority of numbers will have the bare minimum of significant properties, being either odd and composite or even and composite, and only very occasionally will significant properties such as being prime, being perfect, being square, or being Fibonacci be instantiated.

This, then, is the sense in which the tiny numbers are biased with respect to significant properties. The range of significant properties that is encompassed by the tiny numbers radically exceeds that encompassed by similar-sized samples from among much larger numbers. The tiny numbers are *diverse* in a way that collections of larger numbers are not. What impact does this 'diversity bias' have on the role of tiny numbers in enumerative induction? The tiny numbers are atypical in this respect and so in one sense they are not "representative" of the numbers at large. However, in another sense, the tiny numbers *are* representative: the variability and density of properties instantiated by tiny numbers allows them to effectively stand in (or "represent") a much broader range of numbers.[16]

I shall use the term *unrepresentatively representative* to refer to this – initially paradoxical seeming – aspect of the tiny numbers. Normally, unrepresentativeness correlates with lack of variation. This is what underpins Baker's (2007) u-scepticism, and what decreases the value of unrepresentative samples, other things being equal. However, in this context the unrepresentativeness in question pertains to a precisely contrary feature, namely *increased* variation. My claim is that this makes the tiny numbers *more* effective as a basis for enumerative induction, not less. Why, exactly? If some large number has a significant property, p*, then it is likely that there is some tiny number that is p*. So if the large number is a counterexample to a given conjecture, C, in virtue of the number having property p*, a corresponding counterexample to C will occur among the tiny numbers.

By way of illustration, consider the following non-mathematical analogy,

[15] Caldwell (2021).

[16] Are there any good examples of a significant property that is not instantiated by a tiny number, but only by some large number? If we restrict attention here to individual properties, rather than combinations of two or more properties, it is surprisingly difficult to come up with a good example of a significant property whose first instance is greater than 1,000.

which is based loosely on an actual historical example. In the 1930's, the U.S. medical establishment drew up infant growth charts in order to give guidance to doctors about when a baby's rate of growth might be a cause for concern. These growth charts gave upper and lower bounds of what counted as 'normal' weight, height, and weight to height ratio, for each age. In devising the charts, the medical authorities measured thousands of healthy infants of different ages. The surveys were carried out in the rural Midwest, where most of the infants were bottle-fed and of northern European and Scandinavian origin. How representative was this sample? In one sense it was very representative: the 'typical' or 'average' American baby in 1930 may well have been similar to the sampled population. But in another sense, the sample was not representative: there were many other sub-populations present in the U.S. at the time that were not included in the sample. Imagine a second sample of infants, of similar size to the first sample, taken from New York City. As a sub-region of the U.S. in 1930, New York City was far from typical; most other regions would have looked very different. So in our first sense above, the NYC sample is not representative. However, an important way in which New York City was different was in the diversity of its population. This diversity holds across multiple axes: ethnicity; country of origin; socio-economic class; and so on. If we focus for the moment on the geographic origin of the parents of NYC infants in 1930, it may well be the case that there was significant representation from across all 48 states of the U.S. So in this second sense, the NYC sample is representative. This "unrepresentative representativeness" of the NYC sample would have made it a better basis for drawing up widely applicable infant growth charts than the sample from the Midwest.

6 From Bias to Inductive Support (II): Severe Tests

The second feature that makes the tiny numbers different from larger numbers is the occurrence of *boundary cases*. For any instantiated mathematical property, there is a smallest natural number with that property.[17] Without any restriction on what properties we are considering, a given number being a boundary case is of no particular import. Indeed, it is trivially the case that every number, n, is a boundary case for the corresponding property of being greater than or equal to n. But if we restrict attention to *significant* properties (in the sense characterized in the previous section), then boundary cases carry more weight.

As a starting-point, let us consider where the boundary cases for various significant mathematical properties lie. The boundary case for being prime is 2, for being even is 2, for being square is 1, for being perfect is 6, and for being a Fibonacci number is 1. These examples are all tiny numbers, indeed they are

[17] There may also be a largest number with that property, if there are only a finite number of occurrences of the property, but for many mathematical properties the only boundary is at the lower end.

toward the very beginning of the tiny numbers. We can also consider boundary cases for *combinations* of significant properties. For example, the boundary case for being both odd and prime is 3, for being both even and square is 4, and for being both a square and the sum of two squares is 25. Not all boundary cases for significant properties (and for combinations of significant properties) are tiny numbers, but the vast majority are. In this respect, the tiny numbers are very unlike subsets of larger numbers.

What are the implications of this fact, that there are disproportionately more boundary cases among the tiny numbers, for the value for enumerative induction based on samples involving the tiny numbers? The key point here is that general claims, if they fail, often fail to hold for extreme cases. For general mathematical claims, this may not be readily apparent, because failure to hold for the smallest case often means that the claim is never seriously entertained. However, the fact that we sometimes have to modify general conjectures in order to avoid the pitfalls of boundary cases shows that this phenomenon is fairly widespread. For example, the Goldbach Conjecture (GC) is formulated as the claim that every even number *greater than 2* is expressible as the sum of two primes. 2 is a boundary case, since it is the smallest even number. And GC fails for 2, so it needs to be explicitly excluded in order for the conjecture to (potentially) be true. Or consider the theorem that every number has a unique prime factorization, which is true as long as we exclude the boundary cases of 0 and 1.[18]

Alan Hajek has suggested that it is not just in mathematics where boundary cases are more likely to yield counterexamples to general hypotheses. In a paper on "philosophical heuristics," one of the techniques that Hajek recommends wielding against philosophical claims is to check extreme cases:

> [L]ook for trouble among the extreme cases – the first, or the last, or the biggest, or the smallest, ... It is a snappy way to reduce the search space. Even if there are no counterexamples lurking at the extreme cases, still they may be informative or suggestive.[19]

The upshot of the above observations is that the prevalence of boundary cases among the tiny numbers tends to make any sample containing these numbers an unusually *severe test* of a general mathematical conjecture. It is a familiar point that samples that severely test a conjecture but then end up being positive instances of that conjecture provide stronger inductive support for

[18] A less well-known example, that fails at both its lower and upper boundaries, is the following: For a given number, n, assume that we have n boxes and n objects. Let Q be the property that holds of n if and only if it is possible to assign all the objects to boxes in such a way that fewer than n boxes are used and fewer than n objects are in each box. If we pick an arbitrary number, say 17, then it is easy to see that property Q holds for it. In this case, for example, we could put 13 objects in box 1, 4 objects in box 2, and leave the other 15 boxes empty. So we have used only 2 boxes ($2 < 17$), and each box contains less than 17 objects. Does Q hold for every number? No. It is straightforward to verify that Q fails to hold for the two lower boundary cases of 0 and 1. And it also fails to hold for the upper boundary case of ∞.

[19] Hajek (2014, p.292). As one example, Hajek considers testing the claim "every event has a cause" by looking at the first event as a boundary case.

the conjecture than positive instances that never really put the hypothesis at much risk.[20]

There is also a third feature of the tiny numbers that contributes to the severe testing of general hypotheses, and that arises from the feature discussed in the previous section concerning the wider range of significant properties among the tiny numbers. This wider range increases the chances of two or more significant properties being co-instantiated by a single number, and for more different combinations of significant properties to be co-instantiated. Because most significant properties become less common as the size of numbers increases, the frequency of coinstantiation (especially for properties that are not directly related) dramatically decreases beyond the tiny numbers. To give just one example, consider the question of which numbers are both square and Fibonacci. It turns out that there are only three numbers that co-instantiate these two (significant) properties: 0, 1, 144, and these are all tiny numbers.

Call a number *interesting* if it co-instantiates two significant properties such that the proportion of numbers that co-instantiate in this way is much smaller than the proportion of numbers with either property alone. So while significance is a feature of individual properties of numbers, interestingness is a feature of pairs of properties.[21]

Interestingness (like significance) is not a precisely defined notion, but it is clear nonetheless that 0, 1, and 144 are all interesting numbers, in the above sense, in virtue of exhibiting the – very rare – co-instantiation of the properties of being square and being Fibonacci. Or consider the following pair of cases:

> 2 is interesting in virtue of co-instantiating the significant properties of being even and being prime, because this co-instantiation is unique and is therefore much rarer than either being prime or being even.
>
> 3 is not interesting in virtue of co-instantiating the significant properties of being odd and being prime, because this co-instantiation is almost as common as the property of being prime *simpliciter*.[22]

As was the situation with boundary cases, interesting numbers are a potentially rich source of counterexamples to general conjectures about numbers. Interesting numbers combine significant properties in unusual ways, and as such they are more likely to be anomalous and thus to provide severe tests for uni-

[20] For example, following Hempel, both a white shoe and a black raven can be seen as positive instances of the general hypothesis that all ravens are black. However, observing a raven and determining that it is black puts the hypothesis at risk, while observing a shoe and determining that it is white does not.

[21] Note that in a situation in which there are two significant properties and one is a special case of the other, this will not result in the number in question being interesting. Why not? Because if all numbers with property P also have property Q then the proportion of numbers that have the twin properties (P, Q) is the same as the proportion of numbers that have property P.

[22] Which is not to say that 3 is not interesting, it is just that it is not interesting in virtue of *this* combination of significant properties.

versal claims.[23] The excess of interesting numbers among the tiny numbers is therefore another way in which samples drawn from tiny numbers provide disproportionately severe tests of general mathematical conjectures.

Not only is the frequency of interesting numbers much greater among the tiny numbers, in comparison to the non-tiny numbers, so is the typical *degree* of interestingness, or so I shall argue. What makes one interesting number more interesting than another? A couple of features come to mind. Consider a number, n, that has significant properties P and Q. Intuitively, the degree of interestingness of n is determined partly by the significance of P and Q, and partly by how rare the combination of P and Q is relative to these properties considered individually.[24]

In the previous section, we discussed how the range of significant properties is greater among the tiny numbers than among comparable-sized sequences of larger numbers, and how this makes the tiny numbers "unrepresentatively representative." Thinking now of significance as a matter of degree, perhaps a better way to put the above point is that the range of properties of *higher significance* is greater among the tiny numbers.[25] This latter claim seems very plausible, and although I do not have any knockdown argument to establish it, some circumstantial evidence in its favor may be gathered.

The Penguin Dictionary of Curious and Interesting Numbers is, as the title suggests, a (non-exhaustive!) catalog of numbers whose properties are deemed to be interesting for one reason or another.[26] In our terminology, the numbers listed include both those with very significant properties and those with very interesting combinations of properties.[27] We can get some sense of the relative degree of interestingness of numbers of different sizes by looking at snapshots from different places along the number line and seeing what is listed for each sequence of numbers. For reasons of space, I will restrict this survey to looking at three different sequences of three consecutive numbers, picking the most intuitively interesting pairs of properties in each case:[28]

[23] Consider the number 2 in relation to the Goldbach Conjecture, as was previously discussed. Arguably the reason that 2 is a counterexample (and thus needs to be excluded from the domain of application of GC) is because it (uniquely) instantiates both the property of being even and of being prime.

[24] In some sense, then, degree of interestingness involves a trade-off. A pair of highly significant properties whose combination is not rare relative to the occurrence of the properties individually may result in the same level of interestingness as a pair of less significant properties that very rarely occur together.

[25] One way to formulate this more precisely is as follows: for any cut-off that is made between "higher significance" and "lower significance" for mathematical properties, the range of properties above this cut-off is greater among the tiny numbers than among almost all sequences of 1,000 consecutive non-tiny numbers.

[26] Wells (1986).

[27] The catalog also includes many examples of properties that are not significant in our sense, either because they are not internally mathematical (e.g. being related to some physical world phenomenon) or because they are notation dependent (e.g. concerning some feature of the decimal digits of the number in question).

[28] Why confine attention just to combinations of *two* properties in the diagnosis of interestingness? This is a fair question, and I do not mean to rule out analyses of interestingness that go beyond pairwise concatenations of significant properties. However, my worry is that

2 The only even prime number

3 The smallest odd prime number

4 The smallest composite number

102 The smallest 7th power to be the sum of eight 7th powers

103 One of a pair of twin primes (with 101)

104 A semi-perfect number (i.e. equal to the sum of all or some of its proper divisors)

For the third sequence, $\{1002, 1003, 1004\}$, we enter the realm of the non-tiny numbers. It is perhaps no coincidence that, even by this relatively early point on the number line, the proportion of numbers that get a mention in *The Penguin Dictionary* is very low, and that none of the numbers in this third sequence appear. Looking elsewhere for information related to interestingness, the following properties may be noted:[29]

1002 A sphenic number (i.e. the product of 3 distinct primes)

1003 The product of a prime, n, and the n^{th} prime

1004 A heptanacci number[30]

The difference, even just intuitively, between the degree of significance of the properties involved in these three sequences is striking. In addition, the numbers in the first sequence either exhibit combinations of core significant properties that are unique, or are boundary cases of a core significant property, and thus are also unique. This gives all three numbers in this sequence close to a maximal degree of interestingness. In the second sequence, the first number, 102, is a boundary case, but the property involved (being a 7th power that is the sum of eight 7th powers) involves the combination of two properties that are of low significance. The remaining two numbers in the sequence, 103 and 104, exhibit individual properties that are of moderate significance. However, in the absence of a second identified property, the only option is to combine with one of the 'default' significant properties (even, odd, composite) and this will not result in the number being interesting. By the time we get to the third sequence, which is also the first to feature non-tiny numbers, the properties involved have become even more obscure and even less significant, while also not combining with a second, non-trivial significant property to boost any prospects of interestingness. I conclude that this survey, brief and partial though it is,

if we allow interestingness to be determined by combinations of greater numbers of significant properties, it will become too easy to find very rare such combinations and as a result the bar for interestingness will become too low.

[29] *Wikipedia* has entries for individual numbers and their notable properties, which is one useful source of information for non-tiny numbers.

[30] The heptanacci sequence is formed analogously to the Fibonacci sequence, except starting with seven consecutive 1's and then forming the next entry in the sequence by summing the previous seven entries.

lends further support to the claim that the tiny numbers typically exhibit a higher degree of interestingness and thus function especially well as severe tests of general mathematical conjectures.

7 Induction and Mathematical Practice

To summarize where we have got to thus far, the main thesis that I have been defending – in contrast to Baker (2007)'s inductive scepticism about mathematics – is that substantive enumerative inductive support can be accrued for a general mathematical conjecture from successfully testing the conjecture on the tiny numbers (i.e. the natural numbers up to 1,000). In this section, I will argue that this modest pro-inductive stance fits better with mathematical practice than the sceptical view.

Some aspects of mathematical practice connected to enumerative induction are equivocal between the pro- and anti-inductive positions. I have in mind relatively well-known conjectures such as the Riemann Hypothesis, the Goldbach Conjecture, and $P \neq NP$, for which numerous instances have been surveyed, and about which mathematicians tend to have a very high confidence in their truth.[31] The complicating factor is that there are many considerations other than pure enumerative induction that feed into this confidence. Baker highlights one such consideration in the case of the Goldbach Conjecture, namely that the number of ways that an even number, n, can be expressed as the sum of two primes increases steadily with the size of n. In the current context, we are trying to adjudicate between the thesis that induction over the tiny numbers provides substantive inductive support and the contrary thesis that it provides no support. Given the prevalence of factors other than enumerative induction in examples such as the Goldbach Conjecture (GC), each of the above theses is seemingly consistent with mathematical practice.[32] According to the pro-inductivist, mathematicians' high level of confidence in the truth of GC is rooted in partial support from enumerative induction which is then boosted by other (also non-deductive) considerations. According to the inductive sceptic, mathematicians' high level of confidence comes entirely from these other considerations.

Are there other patterns of mathematical practice that give clearer guidance one way or the other? I think that there are, and I will present and discuss one important kind of case. A good illustration of what I have in mind is the Polya Conjecture (PC) example that I introduced right at the start of the paper. Recall that Polya initially verified the claim – about the total number of prime factors of a number being even versus being odd – for the first 1,500 natural numbers, before postulating that it might hold for all numbers. Polya (and, subsequently, other mathematicians) then set about trying to prove PC. A standard way to frame this kind of story is in terms of the distinction between "context of discovery" and "context of justification." The checking of the first

[31] These three examples are mentioned by D'Alessandro in his (2021).

[32] Baker (2007, pp.69-70).

1,500 cases belongs to the discovery phase, while the search for a proof belongs to the justification phase. And this allows Polya's actions to be explained without attributing any justificatory force to enumerative induction over the 1,500 positive instances.

But is this explanation actually adequate? How rational is it to devote time and energy to trying to prove a conjecture if the enumerative induction carried out during the discovery phase provides no substantive evidentiary support for it? Imagine that a mathematician is faced with two *prima facie* equally plausible conjectures, C_1 and C_2. She has verified C_1 for the first 1,000 natural numbers, but she has only verified C2 for the first handful of numbers. If C_1 and C_2 are potentially on a par in all other respects (significance, relevance to her other work, etc.), surely it makes rational sense for the mathematician to devote her energy to trying to prove C_1. James Franklin makes essentially this point in the context of defending non-deductive methods in mathematics more generally:

> Anyone can generate conjectures, but which ones are worth investigating? ... Which might be capable of proof by a method in the mathematician's repertoire? ... [T]o say that some answers are better than others is to admit that some are ... rationally better supported by the evidence.[33]

Interestingly, Polya himself has written about the role of induction in mathematics, and while he does not mention the example of his own conjecture, PC, he does defend the view that enumerative induction can provide substantive evidence for a general mathematical claim:

> [H]aving verified the theorem in several particular cases, we gathered strong inductive evidence for it. The inductive phase overcame our initial suspicion and gave us strong confidence in the theorem. Without such confidence we would have scarcely found the courage to undertake the proof which did not look at all like a routine job. "When you have satisfied yourself that the theorem is true, you start proving it" – the traditional mathematics professor is quite right.[34]

In summary, a view that attributes positive (though not compelling) evidentiary value to enumerative induction over tiny numbers seems to fit much more naturally with the above aspect of mathematical practice than does a blanket inductive scepticism.

The philosophical debate here, however, is not just between inductivism and anti-inductivism. Part of the thesis that I am defending is that the claim of positive bias for tiny numbers is a key factor in what allows enumerative induction in mathematics to carry evidentiary weight. In other words, the pro-inductivist position that I have been arguing for above is to be distinguished

[33] Franklin (1987, p.1).

[34] Polya (1954, pp.83-4). See also Zeng (2022) for an interesting discussion of the contrast between Polya's positive views about enumerative induction in mathematics and the anti-inductivist views of Lakatos.

from more mainstream versions of inductivism that are based on downplaying or dismissing size bias as a potential factor.

There is no standard terminology here, but *prima facie* there are six basic positions that one could adopt concerning size bias and enumerative induction in arithmetic. One could take size bias to weaken enumerative induction, to be neutral, or to strengthen enumerative induction. And one could combine this viewpoint with an overall stance that either sees enumerative induction in arithmetic to be rationally justified or does not. Using the terms “positive bias,” “neutral,” and “negative bias” for the first component, and “pro-inductivist” and “anti-inductivist” for the second component, we can generate labels for all six positions. I don’t think that every position has actually been defended in the philosophical literature, but three of them certainly have.

Negative Bias Pro-Inductivism is the closest to the mainstream position. This is the view that size bias negatively affects the force of enumerative induction, but not to a degree that undermines induction carrying significant evidentiary weight. Among defenders of this position, I include Walsh (2014), Waxman (2017), D’Alessandro (2021). Negative Bias Anti-Inductivism is the view articulated and defended in Baker (2007), according to which the inevitability of size bias undermines the legitimacy of enumerative induction for arithmetic. And Positive Bias Pro-Inductivism is the position that I am defending in the current paper, according to which size bias can positively affect the force of enumerative induction such that samples including the tiny numbers can carry significant evidentiary weight.

Is there anything in mathematical practice that might help differentiate between the plausibility of the two alternative versions of Pro-Inductivism mentioned above? One very widespread feature of the way that mathematicians gather information about instances of a general conjecture is that they tend to try the smallest numbers first. In itself this has no particular implications for the issue of size bias, because there are obvious non-bias-related reasons for this pattern of practice. Typically, smaller numbers are easier to do calculations on, so it makes sense to try these numbers first. Also, for many conjectures, the calculation of the $n+1$ case requires that the n case be calculated first.[35]

Nonetheless, if Negative Bias Pro-Inductivism is true then any sample of smaller numbers is biased in a way that negatively impacts the evidentiary force of that sample, because in sampling from just one very small corner of the domain of natural numbers it is less representative than it could be. So if there were a relatively easy way to diversify the size range of the sample, we would expect to see mathematicians doing this. For example, it would be much better to randomly select one thousand numbers between 0 and 1,000,000 to sample, in support of a general conjecture, rather than simply sampling the first one thousand numbers (i.e. the tiny numbers). Yet we almost never see this happening, even when the relevant properties of the larger numbers are easy enough to calculate.

Another way of putting the above point is that mathematicians do not try

[35] This is true of the Polya Conjecture, for example.

to counteract the size bias in their sampling, even when doing so would be relatively straightforward. Indeed mathematicians actually seem to act in ways that *increase* the size-bias effect, by always starting at the very beginning of the number line, and by plodding through successive small numbers rather than randomly sampling larger numbers. As Baker (2007) argues, the fact that a sample has *some* size bias is not something that can ever be overcome: any finite sample will inevitably include only numbers that are "small" relative to the set of natural numbers as a whole. Nonetheless, there are *degrees* of size bias, and if the size bias in question were in fact a weakening factor for enumerative induction, we would expect straightforward measures that could be taken in order to overcome it to be manifested in actual mathematical practice. The fact that we rarely see this is further evidence that Positive Bias Pro-Inductivism is the more plausible of the two versions of pro-inductivist argument that we have canvassed.

8 Objections

In this section I briefly survey, and respond to, several objections that might be raised against the position that I am arguing for. This position, which I have been calling Positive Bias Pro-Inductivism, maintains that substantive inductive support can be provided for a general arithmetical conjecture through sampling the tiny numbers (the natural numbers up to 1,000), and that this support stems crucially from the fact the tiny numbers are a biased sample.

Objection 1: 1,000 positive instances is not very many. How plausible is it that such a small number of instances yields substantive inductive support for a general mathematical conjecture?

It is important to re-emphasize that what I have been arguing for is that induction over the tiny numbers provides some substantive positive evidence for a general conjecture, and not that it ever (on its own) provides *compelling* evidence.[36] How high should the bar be set for evidence to count as "substantive"? I don't have a ready answer to this question, but a couple of approaches come to mind. One could define the evidence to be substantive if it leads mathematicians to the belief that the conjecture is more likely than not to be true. Or one could take a more operational approach, and deem evidence to be substantive if it makes it rational for mathematicians to pursue a proof of the given conjecture.

Objection 2: Cutting off the tiny numbers at 1,000 seems arbitrary. Why think that there is some significant difference between numbers slightly smaller than 1,000 and numbers slightly larger than 1,000?

[36] Let alone that it provides *knowledge*, which we have already set aside as not being an issue that we are addressing here.

No privileged status is being claimed for the cut-off of 1,000 between tiny numbers and non-tiny numbers. The thesis being defended is simply that enumerative induction over the tiny numbers is *sufficient* (other things being equal) to provide substantive evidentiary support. Obviously if this basis is sufficient then so is a larger basis that goes beyond 1,000 (for example, the basis of the first 1,500 numbers that Polya established in support of his Conjecture). It also may well be the case that for many conjectures a smaller initial basis is also sufficient. The cut-off point of 1,000 was chosen not because it has any specific, special features. The motivation was to pick a cut-off that is large enough so that the tiny numbers clearly provide a sufficient basis for substantive inductive support, and small enough that the bias claim (for tiny numbers being different from numbers at large) still has intuitive force.[37]

Objection 3: The 'positive bias' thesis is based largely on the claim that the range of significant properties is much greater among the tiny numbers than it is within comparably-sized stretches of non-tiny numbers. But how can we rule out there being lots of more complex significant properties that we have failed to recognize and which are well represented among the non-tiny numbers? In other words, perhaps the larger numbers just *look to us* to be relatively impoverished with respect to significant properties.

This is an important objection, since the basic possibility that it raises seems hard to rule out. However, even if we grant the assumption that we may be failing to recognize an array of significant mathematical properties that are too complex for our limited minds to grasp, I think that it is still possible to defend Positive Bias Pro-Inductivism. The reason is that the general conjectures that we – as human mathematicians – put forward will of practical necessity feature significant mathematical properties that are graspable by us. Mathematicians seek to prove theorems that are important, and the clearest way for a theorem to be important is through relating significant mathematical properties. A property that is too complex for us to grasp will be unlikely to feature in the theorems that we prove, and similarly for those properties that we can grasp but whose significance escapes us. Thus, for the purposes of defending the core thesis of Positive Bias Pro-Inductivism, we can simply relativize to the significant mathematical properties that are graspable by us. The general conjectures that we formulate feature significant *and* graspable mathematical properties, and significant and graspable mathematical properties occur across a much wider range among the tiny numbers than among the non-tiny numbers. In other words, even if we do show a bias towards significant-to-us mathematical properties, this does not prevent induction involving these properties from supporting important-to-us general conjectures.

[37] As mentioned in Section 4, one problem with Baker (2007) delineation of the "minute numbers" is that many minute numbers are so large that it is difficult to discern systematic differences between minute numbers and non-minute numbers.

One might worry that this line of response to Objection 3 introduces a problematic circularity into the defense of enumerative induction. Are we not now making the relatively unsurprising claim that induction over the tiny numbers supports important-to-us conjectures, where "important-to-us" is defined as involving properties that are densest among the tiny numbers? I think that there are a couple of things that the defender of Positive Bias Pro-Inductivism can say here. Firstly, "important-to-us" is not being *defined* in reference to the tiny numbers. Rather an empirical claim is being made that the mathematical conjectures that we make tend to involve significant properties that are densest among the tiny numbers. Secondly, as was mentioned in Section 6, we are thinking of significance for mathematical properties as a matter of degree. While it seems plausible that there may be hitherto unrecognized significant properties that are well represented among the non-tiny numbers, it seems much less plausible that many (or any) of these will be as highly significant as properties such as being prime, being square, and being even, which feature in the conjectures that we have tested up to this point.

Objection 4: There is never any such thing as "pure" enumerative induction in mathematics: when induction is utilized, there are always other, conjecture-specific considerations that affect the degree of belief in the general conjecture. Hence there are no compelling grounds, based on mathematical practice, for attributing substantive evidentiary weight to enumerative induction.

Even if the factual claim is conceded, it is not clear why the ubiquitous presence of non-inductive considerations should undermine the Positive Bias Pro-Inductivism position. On a Kuhn-style theory choice model, enumerative induction is just one element among many that feeds into the overall non-deductive evidence for a given conjecture. But if enumerative induction never provides any significant positive weight, it remains puzzling why mathematicians devote time and energy to checking small instances of general conjectures. In other words, the 'no pure enumerative induction' claim does not weaken the argument from mathematical practice that was presented in Section 7 above.

Objection 5: The motivating example for the value of tiny numbers in lending inductive support for a general conjecture was Polya's Conjecture, and George Polya's survey of the first 1,500 numbers. But Polya's Conjecture turns out to be false! Doesn't this undermine the claim of inductive support in this case?

Before responding to the above objection, I will start with a brief overview of what transpired following Polya's original conjecture in 1919. In addition to the putative inductive evidence, there was also a heuristic argument in favor of Polya's Conjecture (PC). Recall that the claim is that, for every natural number, n, at least as many numbers less than n have an odd number of prime factors as have an even number of prime factors. The heuristic argument in favor of PC is as follows. All prime numbers have an odd number of prime factors, while composite numbers seem *prima facie* equally likely to have either an even number of prime factors or an odd number of prime factors. Hence we should expect numbers with an odd number of prime factors to predominate. On the other hand, there were also theoretical considerations that spoke against the truth of PC. It was discovered fairly early on that PC implies the Riemann Hypothesis. This fact was considered evidentially neutral,[38] however PC also implies linear dependence relations among the positive imaginary parts of the non-trivial zeroes of the Riemann zeta function. This latter claim is considered unlikely to be true, and this cast doubt in turn on the truth of PC.

This is how things stood until 1958, when Haselgrove not only proved that PC is false, but also that it fails for some $n < e^{832}$. Two years later the first specific counterexample to PC was found, by Lehman: PC fails for $n = 906,180,359$. In 1980, Tanaka proved that $906,150,257$ is the smallest number for which PC fails.

According to my main thesis, Polya was justified in looking for a proof of PC because he had surveyed the first 1,500 numbers, thus including all of the tiny numbers as part of his sample. In this case, as we have seen, PC was eventually proven to be false, but this in itself does nothing to undermine my claim. For it is perfectly possible to accrue substantive inductive support for a general conjecture that turns out to be false.

9 Conclusions

I concede that the main thesis I am defending may seem counterintuitive. It certainly sounds odd to cite the bias of our sampling methods as a crucial source of strength for enumerative induction in mathematics. However, in other respects the thesis is fairly conservative. Firstly, only a comparatively modest level of support is claimed for induction over the tiny numbers: they provide evidence that is significant rather than conclusive or compelling. Secondly, even the modest level of support holds only *ceteris paribus.* When might other things *not* be equal? Here are a couple of ways:

(i) Very few testable instances of the given general conjecture occur among the tiny numbers.

> e.g. Consider perfect numbers, and the conjecture that all perfect numbers are even. There are only three instances of perfect numbers among the tiny numbers (6, 28, and 496). So testing this conjecture just on the tiny numbers does not accrue substantive inductive support.

(ii) There is some clear size-relative component to the given general conjecture that makes it more likely for larger numbers to provide counterexamples.

[38] Since the Riemann Hypothesis is generally considered to be true.

e.g. Consider the claim that no number is expressible as the sum of two cubes in three different ways. Plausibly, the larger the number the more different ways it is likely to be possible to express it (and the more cubic numbers there are that are smaller than it). And, indeed, the first example of such a number is 87,539,319, which is well beyond the range of the tiny numbers.[39]

In addition to cashing out the *ceteris paribus* condition, another way that the main thesis can be sharpened is by getting clearer on what it is saying about necessary and sufficient conditions for enumerative induction in the mathematical context to be effective. Thus far, all that I have claimed is that an inductive basis that includes all of the tiny numbers is *sufficient* (absent any conjecture-specific contra-indications) to provide substantive inductive support for a general conjecture. Is the inclusion of all of the tiny numbers also *necessary* for this support to accrue?

In considering this question, it may be helpful to distinguish between two different features of the set of tiny numbers. Firstly, the set includes an *initial segment* of the natural numbers. Secondly, the set includes *one thousand* different potential instances. Once we separate out these two features, we can consider alternative evidential bases that lack one but not the other. Compare, for example, the following two samples:

(A) A sample consisting of the natural numbers from 0 to 10.

(B) A sample consisting of the natural numbers from 11 to 1,000.

How much evidence for a generic general conjecture is provided by each of these samples? Intuitively, neither the A-sample nor the B-sample provides sufficient evidence for reasonable confidence in the truth of a general conjecture about the natural numbers. For the A-sample, there are simply too few positive instances to make a reasonable inductive inference. All sorts of trivially false conjectures happen to hold for the first handful of numbers.[40] What about the B-sample? This sample has almost as many instances as the entirety of the set of tiny numbers, so it does fine with respect to the second feature mentioned above. However, the B-sample excludes the first eleven numbers, and so it does not comprise an initial segment of the natural numbers. This is a problem because of one of the key aspects of the tiny numbers, mentioned back in Section 6, namely the high frequency of *boundary cases*.[41] Excising the first eleven numbers from the tiny numbers yields a sample that has drastically fewer boundary cases,

[39] In a way, this case illustrates the converse of what Baker (2007) argues for the Goldbach Conjecture case. For GC, according to Baker, there is zero support from 'pure' enumerative induction, but there are other considerations that strongly suggest that counterexamples are more likely to occur among smaller cases rather than larger cases.

[40] To give just one example: "All odd numbers are either prime or square" holds for all numbers in the A-sample.

[41] Also, mathematicians seem to be aware (at least implicitly) of the importance of including an initial segment as part of any enumerative inductive basis, since the initial segment is almost always surveyed when considering any general conjecture.

because the first handful of numbers include boundary cases for a multiplicity of significant properties.[42]

What does consideration of the A-sample and the B-sample tell us about necessary conditions for substantive inductive support? There are unlikely to be sharp boundaries around what counts as a necessary, minimal evidential basis, but what we can say is that any such basis needs to include a reasonable sequence of cases from the very beginning of the natural number line (say at least the first ten numbers) and a reasonable proportion of the tiny numbers (say at least a few hundred numbers).[43] At the end of the day, I think that establishing the sufficiency claim is more important, dialectically, than honing the necessity claim. The key question concerning enumerative induction in mathematics is whether it ever provides substantive evidentiary support. The thesis that the tiny numbers are sufficient to provide such support answers this question in the affirmative.

References

[1] Baker, A. (2007). Is There a Problem of Induction for Mathematics?. In M. Leng, A. Paseau, & M. Potter (Eds.), *Mathematical Knowledge* (pp. 59-72). Oxford: Oxford University Press.

[2] Baker, A. (2008). Experimental Mathematics. *Erkenntnis*, 68, 331-344.

[3] Caldwell, C. (2021). Table of Known Maximal Gaps between Primes. *Prime Pages*. OL at https://primes.utm.edu/notes/GapsTable.html

[4] Corfield, D. (2004). Mathematical Kinds, or Being Kind to Mathematics. *Philosophica*, 74(2), 37-62.

[5] D'Alessandro, W. (2021). Mathematical Philosophy. *The Reasoner*, 15(4), 32-33.

[6] Franklin, J. (1987). Non-Deductive Logic in Mathematics. *British Journal for the Philosophy of Science*, 38, 1-18.

[7] Frege, G. (1884). *Die Grundlagen der Arithmetik*. Breslau: W. Koeber. Translated by J. L. Austin as *The Foundations of Arithmetic*, Oxford: Blackwell, second revised edition 1953.

[42] The numbers from 0 through 10 include the (lower) boundary cases for properties such as being odd, being even, being prime, being square, being perfect, and being a Fibonacci number.

[43] Note that the necessary conditions that I am suggesting here are conditions for providing *substantive* positive evidence. A sample that does not include the first few numbers, or does not include a reasonable proportion of the tiny numbers may still provide *some* evidence in favor of a given conjecture.

[8] Hajek, A. (2014). Philosophical Heuristics and Philosophical Creativity. In E. Paul & S. Kaufman (Eds.), *The Philosophy of Creativity: New Essays* (pp. 288-318). Oxford: Oxford University Press.

[9] Hardy, G. (1940). *A Mathematician's Apology.* Cambridge: Cambridge University Press.

[10] Lange, M. (2015). Explanation, Existence, and Natural Properties in Mathematics – A Case Study: Desargues' Theorem.

[11] Paseau, A. (2021). Arithmetic, Enumerative Induction, and Size Bias. *Synthese*, 199, 9161-9184.

[12] Polya, G. (1954). *Mathematics and Plausible Reasoning: Volume 1: Induction and Analogy in Mathematics.* Princeton, NJ: Princeton University Press.

[13] Walsh, S. (2014). Empiricism, Probability, and Knowledge of Arithmetic. *Journal of Applied Logic*, 12, 319-348.

[14] Waxman, D. (2017). Deflation, Arithmetic, and the Argument from Conservativeness. *Mind*, 126, 429-464.

[15] Weisgerber, S. (2024). Value Judgments in Mathematics: G. H. Hardy and the (Non-) seriousness of Mathematical Theorem. *Global Philosophy*, 34.

[16] Wells, D. (1986). *The Penguin Dictionary of Curious and Interesting Numbers.* Penguin Books: UK.

[17] Zeng, W. (2022). Lakatos and Quasi-Empiricism. *Kriterion*, 36, 227-246.

Construction Theorems and Constructive Proofs in Geometry

John B. Burgess

Abstract

Given Tarski's version of Euclidean straightedge and compass geometry, it is shown how to express construction theorems, and shown that for any purely existential theorem there is a construction theorem implying it. Some related results and open questions are then briefly described.

Keywords— Euclidean geometry, straightedge and compass, construction problems, existence theorems

1 Introduction

I will be concerned with comparing and contrasting three types of mathematical assertions, illustrated by these:

(1) Existence claim:
There exists a regular heptadecagon.

(2) Constructibility claim:
It is possible to construct a regular heptadecagon.

(3) Construction claim: It is possible to construct a regular heptadecagon in the following way ...

where the ellipsis in the last item would be completed with the specification of Gauss's construction or some other (as in Weisstein, 2021).

The kind of constructions meant here are those familiar from the early books of Euclid's *Elements*, traditionally called *straightedge and compass* constructions. What would have to be meant are an *unmarked* straightedge, not usable as a ruler, and a *collapsing* compass, not usable as dividers, since a ruler or dividers would make it a trivial matter to transfer a length, while Euclid takes doing so to require the sort of substantive material one finds in his Proposition 2. Actually, what Euclid assumes is that a line can be drawn through two points A and B in the plane (in two steps, first joining the points to form the segment AB as provided for by Postulate 1, then extending the segment indefinitely in a straight line at either end, as provided for by Postulate 2), and also that a circle can be drawn of given center and given radius (as per Postulate 3). And he does not specify any method or tools for doing such things; in particular, he does not mention the use of straightedge and compass. Let us nonetheless for convenience retain the traditional label for the class of constructions in question.[1]

Many propositions of Book I, beginning with the very first, are "problems," whose solutions end with words amounting to "which was to be done" or QEF, in contrast to "theorems," whose demonstrations end with "which was to be shown" or QED. Both kinds of propositions have in Euclid a very stylized form of exposition, according an analysis of Proclus (see Netz 1999) consisting of the same sequence of a half-dozen parts: *protasis, ekthesis, diorismos, kataskeve, apodeixis, symperasma.* In a problem, the *kataskeve* does what is to be done, and the *apodeixis* shows that it has successfully done it. In a theorem, the *kataskeve* introduces auxiliary points, lines, circles, or whatever, and the *apodeixis* uses them to show what was to be shown. The term "construction" gets used in two ways in discussing these matters, first as a term for problems as opposed to theorems, second as a translation of *kataskeve.*

Philosophers of mathematics have often contrasted the ancient style of axiomatic geometry represented by Euclid with the modern style represented by David Hilbert (1902). Paul Bernays (1964, p. 275) explicitly draws the contrast as one between statements of type (2) and statements of type (1):

> Euclid postulates: One can join two points by a straight line; Hilbert states the axiom: Given any two points, there exists a straight line on which both are situated.

Thomas Heath, however, in his classic translation (1968), renders Euclid's formulations as infinitival phrases ("to join two points. . . ") rather than complete sentences ("one can join two points. . . ") of type (2). But probably it would not matter much to Bernays if his formulation of Euclid had to be confessed to be not quite accurate, since a closer look at the context shows that Bernays is much less concerned with differences between Hilbert and Euclid (who indeed is mentioned only briefly very near the beginning of the paper) than with differences between Hilbert and twentieth-century "constructivists."

Chief among these was L.E.J. Brouwer, the founder of mathematical intuitionism and Hilbert's main adversary in the Grundlagenstreit or foundational dispute of the years between the world wars, which was just beginning to wind down in 1934 when Bernays produced the original French version of the paper just quoted. And Euclid and Brouwer differ from Hilbert in different ways.

[1] But it may be not inappropriate to recall parenthetically that there are other ways to draw lines and circles, some antedating the oldest surviving examples of compasses, which come from Roman times. Proclus describes geometry as arising out of the practice of Egyptian surveyors, with which Euclid as a resident of Alexandria might be expected to have been familiar. These surveyors were called "rope stretchers" and are depicted in wall paintings as carrying a long rope, which stretched taut can be used to trace a straight line on the ground. It is no coincidence that our word "straight" is etymologically linked to "stretched' and that "line," which is still the nautical term for "rope," is etymologically linked to "linen," flax being one of the materials out of which ropes were produced. A method something like this, involving a taut cord tied at both ends to stakes, which I learned as a child from my grandfather, a landscaper, is still used today in gardening, and a web search on the key words "garden" and "straight line" will turn up several videos giving a demonstration. A taut string could be used to draw a straight line on paper or papyrus or parchment in an analogous way. A string can also be used to draw a circle. It can even be used to draw an ellipse, though Euclid makes no provision in the *Elements* for ellipses and other conic sections.

2 Two Kinds of Constructivity in Geometry

Where Hilbert would have existence theorems, Euclid would have construction problems, while Brouwer would still want existence theorems, but would impose a requirement rejected by Hilbert, that a proof of a theorem to the effect that there exists a mathematical object satisfying a given condition not be accepted unless it is "constructive" in the sense that there is implicit in it a method of specifying a particular example of such an object (as for instance Euclid's proof in the *Elements*, Book IX, Proposition 20 that for any given number n of primes there is a further prime, has implicit in it the method for finding such an additional prime, namely, taking the product of the given primes, adding one to obtain a number m, and checking all numbers up through m until a prime dividing m is found.)

Brouwer *identified* mathematical existence with the possibility of being constructed (or in his more extreme formulations, with the actuality of having been constructed). But in this Brouwer avows influence not from Euclid but from Kant's view that mathematical proof involves "constructions in intuition" (or as Brouwerians sometimes put it, "mental mathematical construction"), though the Kantian view itself was clearly influenced by the role of construction in ancient mathematics, so there would be a Euclidean influence on Brouwer after all, at least at one remove.

Their requirement of constructivity in proofs makes it impossible for intuitionists to accept the classical logical laws of the *excluded middle*, $p \vee \neg p$, or equivalently the law of *double negation* $\neg\neg p \rightarrow p$, as used in proofs by contradiction or *reductio ad absurdum*. Brouwer's disciple Arend Heyting presented a formal version of the principles intuitionists *do* accept, and these two are conspicuously absent, and have to be, because using them one can easily produce existence proofs where no specific example of the sort of thing claimed to exist is provided even implicitly.

What is perhaps the simplest case of this phenomenon is shown in the classical proof, intuitionistically unacceptable, of the following:

(4) $\exists x \forall y (\Phi(y) \rightarrow \Phi(x))$

"There is something that Φs if anything does." For excluded middle tells us that either there exists something that Φs or there doesn't. If there does, any such thing may be taken for x in (4), and the conditional will be true because its consequent is true. If there does not, then any x at all may be used in (4), and the conditional will be true because its antecedent is false. This clearly does not give us a specific example of an x of the kind asserted to exist by (4), unless we already either know an example of something that Φs or know that there isn't any.

It would be anachronistic to say that Euclid reasons by "classical" logic in the sense of the logic expounded in orthodox present-day textbooks, but he certainly does *not* conform to the constructivist restrictions of intuitionistic logic. Notoriously he uses *passim* the method of proof by *reductio* (see for instance Book I, Proposition 29). And according to the early modern commentator Christoph Schlüssel, he uses the equally intuitionistically unacceptable law $(\neg p \rightarrow p) \rightarrow p$, sometimes called the *consequentia mirabilis*, but dubbed by Jan Łukasiewicz the *law of Clavius* ("Clavius" being Schlüssel's Latinization of his Germanic family name, which like the Latin *clavis* means *key*) and perhaps better known under that label. So we must distinguish Euclid's emphasis on *construction problems* from Brouwer's insistence on *constructive proofs*, though the full significance of the difference between the two will only gradually become clearer in what follows.

It appears to have been Bernays' paper that first introduced something like the (historically dubious) use of "platonism" current in philosophy of mathematics during the last half-century or more, during which period there has been a vigorous debate between so-called platonists and so-called nominalists over the existence of abstract entities, and specifically mathematical objects such as numbers or sets. Allusions to Euclidean constructions have played an occasional role in this debate.

Notably, Charles Chihara (1973), one of the earliest of a group of writers who have attempted to reconstruct or reconstrue mathematics nominalistically, pursues the following sort of strategy. First, statements about mathematical objects such as numbers are replaced by statements about linguistic expressions, mathematical notations such as numerals. Second, assertions of the actual existence of linguistic expressions as abstract types are replaced by assertions of the possible existence of linguistic expressions as concrete tokens. Third, the modal notion of "possible existence" in this context, which some have criticized as unacceptably "metaphysical," is explained as potential *physical* "constructibility," which Chihara claims is a notion that should be familiar from ancient mathematics and considered philosophically respectable.

And in this last connection he quotes with approval an expression of a similar sentiment from his colleague Ernest Adams (1973, p. 406), who alludes to Euclid thus:

> Euclidean geometry cannot be criticized for lack of rigor simply on the grounds of its modal formulation, whatever other faults it may have in this respect, since the logical laws to which the constructibility quantifier conforms are quite clear.

It is clear enough, though, from the writings of the contemporary and recent philosophical thinkers mentioned so far—Adams, Bernays, Brouwer, Chihara, Heyting, and the list could be extended—that even if all are inspired in some way and to some degree by Euclid's *Elements*, none is primarily concerned with arguing over the correct interpretation of that ancient text. But apart from appropriations for other philosophical purposes, the correct interpretation of Euclid on construction *has* been a topic of interest for its own sake, and a subject of discussion among scholars of ancient Greek philosophy and historians of ancient Greek mathematics.

Recently Silvia De Toffoli and I conducted a joint graduate seminar on mathematical rigor in theory and practice, and among our guest speakers we were fortunate enough to have Benjamin Morison, who gave an account of his work with Jonathan Beere (unpublished manuscript) in precisely this area. The present note is a kind of scholium to the Beere-Morison paper, describing some relevant logical facts about formalized systems of geometry, in a way that I hope will be readable by and informative to students of classical philosophy and ancient mathematics who have a modicum of knowledge of modern logic but are not specialists.

The main thesis of the joint authors concerns the *purpose* of construction problems. (They also offer the opinion that the infinitival phrases "to do this or that" should be taken to have a kind of imperatival or jussive force; but philological issues are beyond me.) Their view is that, just as theorems and their proofs aim to impart an especially solid kind of knowledge-*that* something is the case, *episteme* or *scientia*, so construction problems and their solutions aim to impart an especially solid kind of knowledge-*how* to do something. It is when one tries to express this knowledge-how as a kind of knowledge-that that one may arrive at something like a construction theorem of type (3), as contrasted with a constructibility claim of type (2).

Constructibility claims may still play a role, though a limited one, for those whose primary interest is in construction claims. How is perhaps best brought out by comparison with the views of finitists on arithmetic. For finitists, existence statements in number theory of the type "condition $\Phi(n)$ holds for some natural number n" are not really meaningful or *inhaltich*. Nonetheless, finitists allow that such a form of words may be used as a *partial communication* of meaningful results of the type "condition $\Phi(n)$ holds for the following natural number n..." where would follow the specification of a relevant n. In the same way, something like (2) may be admitted as a *partial communication* of something like (3).

As Beere and Morison make clear, though existence and constructibility assumptions or assertions are not entirely absent from the *Elements*, they are rare. By contrast, construction problems are ubiquitous in the geometrical books, which famously culminate in Book XIII with the construction of the regular polyhedra or Platonic solids. The situation is different in present-day mainstream mathematics. There questions of existence are prior, and it is only when one is answered in the affirmative that a question of constructibility then arises, or rather a variety of them for various kinds of constructibility (of which straightedge and compass constructibility in plane geometry is only the best known). Similarly with numerical functions: Existence comes first, computability in this or that sense (say recursive or primitive recursive or polynomial-time) being a distinct and subsequent issue or series of issues.

Moreover, though modal idioms such as "it is possible to..." or "one can..." or -*ible* and -*able* suffixes do occur all over present-day mathematics, including the present note, in formal contexts such locutions get explained away in terms of existence statements, rather than the other way around. Thus constructib*ility* and comput*ability* are defined in terms of the existence of programs for constructing or computing. This circumstance would affect the understanding of (2) and (3), turning them into "There is a program for doing such-and-such," and "The following is a program for doing such-and-such...," or something of the sort.

3 Tarskian Rigor

But the most important difference of present-day from ancient mathematics is that mathematicians now hold themselves to a higher standard of rigor. They have done so since the later nineteenth century when Frege opened the body of his epochal *Grundlagen der Arithmetik* with the following much-cited remark, which I quote in the English translation of J. L. Austin (see Frege, 1953, p.1):

> After deserting for a time the old Euclidean standards of rigor, mathematics is now returnng to them, and even making efforts to go beyond them.

For Euclid is indeed not the ultimate or last word on rigor, and there are notoriously what from our modern standpoint must be regarded as lapses rigor in the *Elements* (as seems to be conceded, somewhat grudgingly perhaps, by Adams in the passage quoted by Chihara), many of which were being repaired by contemporaries of Frege.

Hilbert's standard is that each theorem must follow logically from postulates stated in advance, where this means that logical form alone, regardless of subject matter or content, guarantees that *if* the postulates are true, *then* so is the theorem. Such is the meaning of the much-quoted statement attributed to Hilbert by his oldest graduate student in a biographical sketch at the end of the third and last volume of Hilbert's collected papers: One must at all times in place of "points, lines, planes" be able to say "tables, chairs, beermugs" (*man muß jederzeit an Stelle „Punkte, Gerade, Ebene"*

,,Tische, Stühle, Bierseidel" sagen können, Blumenthal 1935, p.403). Despite the great heuristic and pedagogical value of inspection of geometrical diagrams and deployment of visual intuition, Hilbert's standards rule out any appeal to such things *in proofs*, and not just because pictures can sometimes be misleading (as in the "proofs" that every triangle is isosceles and some obtuse angles are right in Collingwood, 1899, pp. 900-902).

Euclid's practice does not always conform to these restrictions of Hilbert's. Most notoriously, Euclid lists the primitive constructions and assertions he is supposing to be allowed in Postulates 1-5, but then it seems immediately assumes another without explicit statement or acknowledgment in Proposition 1. This, it will be recalled, is the problem, given a segment AB, to construct an equilateral triangle ABC having it as a side. Euclid considers two circles with radius AB, the one with center at A and the one with center at B, whose construction is provided for by Postulate 3. But in subsequent reference to these circles he speaks of each not as the circle with such-and-such center and radius, but rather by mentioning three points on it. Now for any three non-collinear points there is indeed a unique circle passing through them, but Euclid does not deal with such matters until Book IV, so he may seem to be getting ahead of himself in Book I. Beere and Morison allude to this sort of thing as an illustration of the gap between existence and construction: by definition, every circle *has* a center, but to *find* the center when the circle is presented by naming three points is another matter. Crucially, by labeling his two circles "the circle BCD" and "the circle ACE" Euclid insinuates, without any logical justification on the basis of his announced postulates, that the two circles have a point C in common. In modern treatments this can be proved using a circle axiom to be stated shortly, but Euclid acknowledges no need for any such thing among his postulates.[2]

Besides the difficulty under discussion, which occurs not only in Proposition 1 but well beyond it, there are similar lapses involving the intersection of a line and a triangle, which by modern standards necessitate another missing axiom, introduced by Moritz Pasch in 1882. As the remark of Bernays hints, it was in the process of rigorization, filling in such gaps in Euclid's arguments as was done by Pasch and other nineteenth century figures culminating in Hilbert, that the explaining away of modal locutions in terms of existence statements occurred.

Euclidean straightedge and compass plane geometry was given its final rigorous form only in the twentieth century, by Alfred Tarski and coworkers, in a formal theory here to be called $\mathbf{G}_0$. (It is called $\mathrm{CG}^{(2)}$ in the authoritative survey of Tarski and Gi-

[2] Is Euclid's practice in Proposition 1 a lapse not merely from the standards of Hilbert, but also a lapse from the standards of Euclid himself (of which we unfortunately have no quotable capsule formulation comparable to Hilbert on beermugs)? The existence of the needed point of intersection seems evident from the diagram if not from the text of Proposition 1, but leaving something that cannot be deduced from the text of a proof to be inferred from an accompanying diagram is not permitted by Hilbertian standards. What about Euclidean standards? For present purposes we may set aside this question, and more generally the whole contentious issue of the role of diagrams in ancient Greek mathematics, where they certainly play a much larger role than in Hilbert's work (with as an extreme case Euclid seemingly feeling obliged to attach a diagram to each proposition even in his arithmetical books, where they appear hardly more than doodles in the margin). A great deal discussion of considerable interest (see in particular Netz, 1998) has been written about Euclidean diagrams, but this discussion is not of direct relevance to the concerns of the present note. What we must acknowledge here is that *if* "Euclidean standards of rigor" do permit the sort of steps Euclid takes in Proposition 1, *then* Hilbert's standards do not only, as Frege's remark on Euclidean standards suggests, "go beyond them," but go *a very substantial distance indeed* beyond them.

vant, 1999, p. 191.) The austere formalism of $\mathbf{G}_0$ is to begin with a *first-order* theory, one whose logical notions comprise those of standard classical logic as in present-day textbooks *and those only*, without any "higher-order" apparatus as in Hilbert, let alone intuitionist logical operators as in Brouwer and Heyting, or modalized quantifiers as in Chihara or Adams.

Moreover, the variables of the language are to be thought of as ranging only over points of the plane: there are no variables for lines or circles, for instance. And there are only two primitive predicates or relations symbols for two geometric relations:

(5a) $\mathbf{B}xyz$ or *betweenness*: x, y, z lie in a line, with y between x and z

(5b) $\mathbf{C}wxyz$ or *congruence*: w lies as far from x as y lies from z

(Here betweenness is to be understood inclusively, so $\mathbf{B}xxz$ and $\mathbf{B}xzz$ always hold, and identity of points $x = y$ is definable as $\mathbf{B}xyx$.)

Nonetheless, because any pair of distinct points may be regarded as *coding* a line (the one passing through both points) and a circle (the one having as center the first point and passing through the second point), one can express indirectly through coding the basic geometry of lines, circles, as well as many kinds of composite figures. (Propositions about areas, including the Pythagorean theorem, admittedly present further challenges.) In particular, $\mathbf{G}_0$ has a *circle axiom* to the effect that if a line has both a point inside a given circle (of distance from the center less than the radius) and a point outside that circle (of distance from the center greater than the radius), it will have a point *on* the circle.

(Ellipses, too, can be coded, not by pairs but by triples of points, the two foci plus any point on the circumference. But in $\mathbf{G}_0$ there is no ellipse axiom comparable to the circle axiom, and arguments assuming the existence of intersections of conic sections could not have been carried out on the basis of Euclid's postulates—not that Euclid wished to go into such matters.)

Returning to (1)-(3), because there is no modal apparatus in the language of $\mathbf{G}_0$, (2) cannot be expressed if the modal locutions are understood literally or primitively, in terms of a possibility operator $\Diamond$. And if (2) is explained in terms of the existence of a program, it still cannot be expressed—not directly, since the variables do not range over programs; nor yet indirectly, since programs cannot be coded by pairs of points, as can lines and circles, or by configurations of any other fixed finite number of points.

But two more positive remarks can be made. There is a slight modification or small adjustment $\mathbf{G}$ of $\mathbf{G}_0$ of which the following may be said:

(A) By adding to the language of $\mathbf{G}$ symbols for certain functions definable in $\mathbf{G}$, corresponding to the familiar operations of finding the intersections of lines and circles, one can obtain a notation for straightedge and compass construction programs permitting the systematic expression of assertions of type (3).

(B) It can then be shown that, for any purely existential assertion of type (1) provable in $\mathbf{G}$, there is a construction assertion of type (3) that implies the given existenstial assertion and is provable in $\mathbf{G}$ taken together with the definitions of the new symbols

In the next two sections I will first sketch the construction justifying (A), then outline the proof justifying (B).

These results should not be surprising. On the contrary, they are just what one might expect (in both the factive and the normative sense) of a formalism in modern language that aspires to embody so far as possible the spirit of Euclid's construction-oriented geometry. But the departures from the *ipsissima verba* of ancient formulations needed to secure a level of rigor up to modern standards are numerous enough and large enough that (A) and (B) are not obvious without proof, either. Their proof is, so to speak, a check on the fidelity of Tarski and his school to something like the Euclidean point of view.

Before proceeding further let me indicate the "slight modification or small adjustment" needed in the formal geometry under consideration. For $\mathbf{G}_0$ as thus far described conspicuously fails to satisfy the most basic kind of "constructivist" requirement that every proof of the existence of a configuration of points with a certain feature has to have at least implicit in it a specification of a particular example of such a configuration.

The most trivial existential theorem is simply that there exists a point that is self-identical, or more simply still, that there exists a point. But it is not possible to specify any particular point using only the betweenness and congruence relations. This can be seen by looking at the most familiar model of $\mathbf{G}_0$, the *Cartesian plane* as studied in middle school. The points of the plane are pairs (a, a') of real numbers, and betweenness and congruence are defined by the usual formulas. Thus if $x = (a, a')$, $y = (b, b')$, $z = (c, c')$, $u = (d, d')$, $v = (e, e')$, then

(6a) $\mathbf{B}xyz$ holds if and only if

$a \leq b \leq c$ or $c \leq b \leq a$ and

$(b - a) \cdot (c' - a') = (b' - a') \cdot (c - a)$

(6b) $\mathbf{C}xyuv$ holds if and only if

$(b - a)^2 + (b' - a')^2 = (e - d)^2 + (e' - d')^2$

In this plane any condition Φ satisfied by any one point x will equally be satisfied by any other point x', so that the condition cannot be used to pick out a distinguished point. This is because there is a *translation* carrying x' to x, and translation preserves relations of betweenness and congruence, in terms of which Φ would be stated.

If we add a constant $\boldsymbol{a}$ to denote some specific, privileged or destinguished point a, the "constructivist" condition will still fail to hold. For it will be a trivial theorem that there exists some point y other than a, while again it is not possible to specify any particular example. This is because, given any point y distinct from a and any other such point y' there will be a rotation around a, keeping a fixed, that will carry y', if not yet to y itself, then at least to some other point y'' lying on the line ay on the same side of a as y. And then there will then be a *dilatation* (uniform expansion or contraction) leaving a fixed and carrying y'' to y. And rotation and dilatation, just like translation, preserve betweenness and congruence relations.

If we add a constant $\boldsymbol{b}$ to denote some specific, privileged or distinguished point b other than a, then there will be infinitely many points of the plane that *can* be uniquely specified in terms of the two distinguished points a and b. One of these will be the midpoint of the segment ab, characterizable as being between a and b and at the same distance from either. But all the specifiable points will, like this example, be on the line ab. And it is, of course, a theorem of plane geometry that the plane is more than one-dimensional, and so contains a point z not collinear with a and b. But for any such point z there is another such point z', though only one, satisfying all the same conditions involving betweenness, congruence, and the two distinguished point a and b, namely, the mirror image of z on the opposite side of ab. For there will be a transformation, namely, *reflection* in the line ab, leaving a and b fixed, that will carry z' to z, and reflections as much as translations, rotations, and dilatations preserve betweenness and congruence.

But if we add a constant c to denote some specific point c not on the line ab, then it is a theorem of plane geometry that the plane is less than three-dimensional, and there will be no two points z and z' satisfying all the same conditions involving betweeness and congruence and the three distinguished points; in particular there will be no two points simultaneously equidistant from a and from b and from c. From any pair of points z, z' we may therefore always specify one as what may be called *proximal* and the other *distal* relative to the three distinguished points a, b, c:

Namely, if the two points are at different distances from first point a, then the one nearer to a should be designated the proximal one. If the two points are equidistant from a but of different distances from the second point b, then the one nearer to b should be designated the proximal one. If the two points are equidistant both from a and from b, then they cannot be equidistant also from the third point c, and the one nearer to c should be designated the proximal one.

In what follows it will be convenient to pick as the point c one of the two points of intersection of the circle with center a passing through b and the line through a perpendicular to the line ab, thus obtaining what in the terminology of Burgess (1982) would be called a system of *benchmarks* a, b, c. In connection with such as system the point a may suggestively be termed the *origin*, and the lines ab and ac the *horizontal* and *vertical axes*, and b and c the *unit points* on those axes.

Now it is a completely general fact about first order theories that if we start with such a theory **T** in a language **L** in which can be proved $\exists x \Phi(x)$ for some condition expressed by a formula Φ of **L**, and if we then add a new constant $\boldsymbol{d}$ to **L** and the new axiom $\Phi(\boldsymbol{d})$ to **T**, the result **T'** will be what is called a *conservative extension* of **T**, meaning that anything statable in **L** (without the new constant) and provable in **T'** (with the new axiom) will be already provable in **T** (without the new axiom). (This is an immediate consequence of the rule of existential elimination found in textbook natural-deduction formulations of first order logic.) Further, by first-order logic, when any conclusion $\Psi(\boldsymbol{d})$ is provable in **T'** the following will be a theorem of **T**:

(7) $\forall x(\Phi(x) \to \Psi(x))$

Similar results hold when adding two, three, or more constants. In particular, if we let **G** be the extension of $\mathbf{G}_0$ obtained by adding the constants $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ and the axiom that the points they denote form a system of benchmarks, then the extension will be conservative, and to that extent harmless. (For it is a theorem of $\mathbf{G}_0$ there do exist systems of benchmarks, and indeed that for any two distinct points there exists a system of benchmarks of which the given points are the first two. Indeed, there exist exactly two such systems.) This technical adjustment will remove the three rather trivial failures of "constructivism" mentioned above. (A) and (B) should henceforth be understood as applying to this conservative extension **G** of $\mathbf{G}_0$, and the terms *proximal* and *distal* henceforth understood relative to the benchmarks denoted by the three new constants.

4 To Express a Construction Theorem

Where we have only points, as in Tarskian formulations, a straightedge and compass construction must be viewed as involving marking new points given previously marked points. (Constructing a line or circle will be a matter of constructing a pair of points coding one.) There are three basic kinds of steps:

(8a) to mark the intersection of the lines coded by x, y and by u, v

(8b) to mark the intersections of the line and circle coded by x, y and u, v

(8c) to mark the intersections of the circles coded by x, y and by u, v

Let now $\mathbf{L}$ be a first-order language and $\mathbf{T}$ a theory in $\mathbf{L}$ and Θ a formula of $\mathbf{L}$ for which it is a theorem of $\mathbf{T}$ that for any xs there exists a unique y such that Θ holds of the xs and y, or in symbols, the following:

(9) $\forall x_1 \forall x_2 \ldots \forall x_n \exists! y \Theta(x_1, x_2, \ldots, x_n, y)$

Then one can add an operator or function symbol $\boldsymbol{\vartheta}$ representing the function that, given any xs as input, gives their y as output. In this situation Θ is called the *defining formula* of $\boldsymbol{\vartheta}$, the relation expressed by $\Theta(x_1, x_2, \ldots, x_n, y)$ is called the *graph relation* of $\boldsymbol{\vartheta}$, while the defining axiom for $\boldsymbol{\vartheta}$, to be added to $\mathbf{T}$ when $\boldsymbol{\vartheta}$ is added to $\mathbf{L}$, is the following:

(10) $\forall x_1 \forall x_2 \ldots \forall x_n \Theta(x_1, x_2, \ldots, x_n, \boldsymbol{\vartheta}(x_1, x_2, \ldots, x_n))$

Because the symbol $\boldsymbol{\vartheta}$ is definable, every formula containing it will have a formula in the original language $\mathbf{L}$ without the new function symbol that can be proved equivalent to it given $\mathbf{T}$ plus the defining axiom.

All this is true quite generally, for arithmetic, for geometry, for anything. In the specific context of geometry, we will want to add in this way to the language $\mathbf{L}$ of $\mathbf{G}$ function symbols corresponding to the basic construction steps (8abc). For this we must first indicate or recall how various auxiliary notions are expressible in the language of $\mathbf{G}$. Here are a few basic ones:

(11a) $|xyz$ or *collinearity*:
"x, y, z lie on a line" or
"y lies on the line coded by x, z":
$\mathbf{B}zxy \vee \mathbf{B}xzy \vee \mathbf{B}xyz$

(11b) $\equiv xyz$ or *equidistance*:
"y lie at the same distance that z does from x" or
"y lies on the circle coded by x,z":
$x \neq y$ & $\mathbf{C}xyxz$

(11c) $\therefore xyz$ or *equilaterality*:
"x, y, z are the vertices of an equilateral triangle":
$\neg|xyz$ & $\mathbf{C}xyyz$ & $\mathbf{C}xzyz$

(11d) $\perp xyz$ or *perpendicularity*:
"the angle yxz is right":
$\neg|xyz$ & $\exists w(\mathbf{B}wxy$ & $\mathbf{C}xwxy$ & $\mathbf{C}zwzy)$ or equivalently
$\neg|xyz$ & $\forall w(Bwxy$ & $\mathbf{C}xwxy \rightarrow \mathbf{C}zwzy)$

(11e) $\oplus xyz$ or *benchmarking*:
"x, y, z form a system of benchmarks"
$\mathbf{C}xyxz$ & $\perp xyz$

(11f) $< xyz$ or *nearness*:
"y lies less far than z does from x":

$\exists w(w \neq z \;\&\; \mathbf{B}zwx \;\&\; \mathbf{C}xyxw)$ or equivalently

$\forall w(\mathbf{B}wzx \to \neg\mathbf{C}xyxw)$

(11g) $\langle tuvxy$ or *proximality*:

"t, u, v constitute a system of benchmarks and
of the two points x and y, the former is proximal
and the latter distal relative to them":

$\oplus tuv \;\&\; [< xyt$ or $(\mathbf{C}xtyt \;\&\; < xyu)$ or $(\mathbf{C}xtyt \;\&\; \mathbf{C}xuyx \;\&\; < xyv)]$

In $\mathbf{G}$, where have distinguished a system of benchmarks denoted $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, we may write simply $\langle xy$ for $\langle \boldsymbol{abc}xy$. The list (11) could be indefinitely extended.

We next define five seven-place functions, one of them, $\boldsymbol{\alpha}$, related to constructions steps of kind (8a), two of them, $\boldsymbol{\beta}$ and $\boldsymbol{\beta}$', related to construction steps of kind (8b), and two of them, $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}$', related to construction steps of kind (8c). The benchmarks will play three roles in what follows.

First, constructions must begin with some data, as Euclid's Proposition 1 starts with there being given a line segment, or equivalently the two distinct points, its endpoints. The benchmarks will serve the starting points for all the constructions with which we will be concerned.

Second, before we can introduce a symbol $\boldsymbol{\vartheta}$for a function in the manner discussed in connection with (9) and (10), the function must be *total*, or defined for all inputs, even in waste cases where we do not care what the output is; and we may conventionally take the first benchmark as the waste-case output. For geometric constructions the definition of waste case for a function

$\boldsymbol{\delta}(t, u, v, w, x, y, z)$

where $\boldsymbol{\delta}$ is any of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$', $\boldsymbol{\gamma}$' will be a disjunction four clauses, the first three the same for all of (8abc):

(12a) t, u, v are not a system of benchmarks

(12b) $w = x$, meaning that the pair w, x fails to code a line or circle

(12c) $y = z$, meaning that the pair y, z fails to code a line or circle

The last clause will be different for each of the three basic construction steps, thus:

(13a) the lines coded by x, y and u, v do not intersect in a point
(they coincide or are parallel)

(13b) the line coded by x, y does not intersect the circle coded by u, v in two points
(the former is disjoint from or tangent to the latter)

(13c) the circles coded by x, y and u, v do not intersect in two points
(they coincide or are disjoint or are tangent)

In any waste case we will set the value of function $\boldsymbol{\delta}$ to be t.

Third, when we need to mark a point of intersection of a line or another circle with a given circle, we have two equally good options that are mirror images of each other, so to speak. Euclid's instructions for producing an equilateral triangle on a given base, as we find them implicit in his Proposition 1, do not tell us how to choose, and this may be said for many others of his construction problems as well: Buridan's ass would be unable to complete the construction given only Euclid's instructions. With benchmarks in the background we can agree that when we are faced with a choice between two points, we should always take the proximal one of that pair.

We can now write down the definitions of symbols for the functions connected with constructions (writing as usual in mathematics "iff" for "if and only if").

(14a) $\boldsymbol{\alpha}(t, u, v, w, x, y, z) = s$ iff we are in a waste case and $s = t$ or s is the point of intersection of the lines coded by w, x and by y, z

(14b) $\boldsymbol{\beta}(t, u, v, w, x, y, z) = s$ iff we are in a waste case and $s = t$ or s is the proximal point of intersection of the line and circle coded by w, x and by y, z

(14c) $\boldsymbol{\gamma}(t, u, v, w, x, y, z) = s$ iff we are in a waste case and $s = t$ or s is the proximal point of intersection of the circles coded by w, x and by y, z

(14d) $\boldsymbol{\beta}$' like $\boldsymbol{\beta}$ but for distal

(14e) $\boldsymbol{\gamma}$' like $\boldsymbol{\gamma}$ but for distal

Here in (14bcde) *proximal* and *distal* are to be understood relative to the benchmarks t, u, v.

Three trivial three-place functions as follows will also be wanted:

(15a) $\mathbf{i}(t, u, v) = t$ (15b) $\mathbf{j}(t, u, v) = u$ (15c) $\mathbf{k}(t, u, v) = v$

Applied to a system of benchmarks these pick out, respectively, the origin and horizontal and vertical unit points.

One further abbreviation will make for conciseness: Let $\S f g_1 \ldots g_n$ denote the m-place function obtained by substituting n given m-place functions in the n-place function f, thus:

(16) $\S f g_1 \ldots g_n(x_1, \ldots, x_m) = f(g_1(x_1, \ldots, x_m), \ldots, g_n(x_1, \ldots, x_m))$

Then a program for constructing a configuration of k points (starting from given benchmarks) can be represented as a k-tuple of complexes of the above-mentioned nine symbols $\boldsymbol{\alpha}$,$\boldsymbol{\beta}$,$\boldsymbol{\gamma}$,$\boldsymbol{\beta}$',$\boldsymbol{\gamma}$',**i**,**j**,**k** and §.

To apply this apparatus to Euclid's Book I, Proposition 1, we will want a complex ψ provably satisfying the following:

(17) $\therefore (\mathbf{i}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}), \mathbf{j}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}), \psi(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}))$

Here **i** and **j** applied to our system of benchmarks will simply give us the first two of them, thus determining a line segment such as is given at the outset in Euclid's construction. The crucial ψ applied to our benchmarks should yield the proximal point of intersection of two circles, the radius of each being the distance between the first two benchmarks, and the centers of the two being the first and the second benchmark. This amounts to $\boldsymbol{\gamma}$ (**a, b, c, a, b, b, a**), which amounts to the composition of the seven-place function $\boldsymbol{\gamma}$ with the seven three-place functions **i, j, k, i, j, j, i**.

Hence the program for the construction of the vertices of an equilateral triangle, given benchmarks a, b, c can be represented by a trio of symbol complexes φ, χ, ψ as follows:

(18) $\varphi = \mathbf{i}$ $\chi = \mathbf{j}$ $\psi = \S\boldsymbol{\gamma abcabba}$

And the existence theorem that there is an equilateral triangle, namely,

(19) $\exists x \exists y \exists z \therefore (x, y, z)$

is implied by the following restatement of (17):

(20) $\therefore (\varphi(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}), \chi(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}), \psi(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}))$

Leaving the benchmarks to be understood we may write this as $\therefore (\varphi, \chi, \psi)$ for short. And what has just been said about the equilateral triangle applies equally to the regular heptadecagon, substituting Gauss's construction for Euclid's. Behind a formalized version of (1) in the style of (19) there stands a formalized version of (3) in the style of (20).

More generally, for any condition expressible in a formula Φ, a statement of type (3), to the effect that an explicitly specified program would produce a configuration of points $x_1, x_2, x_3, \ldots$ satisfying Φ, can be taken to be a formula of the form

(21) $\Phi(\varphi_1, \varphi_2, \varphi_3, \ldots)$

wherein $\varphi_1, \varphi_2, \varphi_3, \ldots$ are symbol complexes of our notation for representing programs, indicating how points $x_1, x_2, x_3, \ldots$ are to be constructed from the benchmarks, applying specified cases of (8abc) in a specified order.

In sum, while a *constructibility* assertion of type (2), to the effect that there exists a program that would produce a given kind of configuration, has to be expressed in the "metalanguage," as the statement that there exist φs for which (21) is a theorem, by contrast we have produced (or at least sketched, relying on results of Tarski and his school) a way, namely (21) itself, of expressing in the "object language" a *construction* assertion of type (3), QEF.

5 There Is a Construction Theorem for Any Pure Existence Theorem

A *purely existential* formula (respectively, *purely universal* formula) is one that, when it is written out in primitive notation, with all defined symbols replaced by their definitions, consists of a string of existential quantifiers $\exists$ (respectively, universal quantifiers $\forall$) in front of a quantifier-free formula. (A bit confusingly, formulations of a theory that have only purely universal formulas as axioms are conventionally called *quantifier-free* formulations. This is because of the custom of omitting the initial universal quantifiers when a purely universal axiom is stated, leaving them to be tacitly understood. Thus for instance in arithmetic the commutative law for addition is expressed as

$$x + y = y + x$$

where what is really meant is the *universal closure* of this formula, namely

$$\forall x \forall y (x + y = y + x).$$

In effect, assertion of the version without the initial string of universal quantifiers is taken to be tantamount to assertion of the version with them.)

A formula provably equivalent in **G** to a formula that is purely existential (respectively, purely universal) is said to be of class Σ (respectively, class Π). A formula of class Δ is any that is of both classes Σ and Π, thus expressing a notion that can be given both a purely existential and a purely universal definition. The function denoted by a symbol $\boldsymbol{\vartheta}$ introduced as above is of class Δ if its defining formula is.

The basic features of these notions belong to general definability theory and include the following closure properties:

(22a) Σ is closed under $\wedge$ and $\vee$ and $\exists$

(22b) Π is closed under $\wedge$ and $\vee$ and $\forall$

(22c) Δ is closed under $\wedge$ and $\vee$ and $\neg$

and all three are closed under substitution of Δ functions. There is nothing specifically geometrical about these facts, nor about the fact that the negations of Σ formulas are Π formulas, and vice versa. Notably, the general proofs about Σ and Δ are very much like those for semi-recursive and recursive in computability theory (compare Boolos et al., 2007, p. 76, Theorem 7.4).

Now obviously (21) implies the existence statement

(23) $\exists x_1, \exists x_2, \exists x_3, \ldots \Phi(x_1, x_2, x_3, \ldots)$

which will be *purely* existential provided Φ is quantifier-free (and will be Σ provided Φ is Σ). This covers the case of the formula stating that the xs are the vertices of a regular polygon of n sides, for any $n \geq 3$. It also covers many other cases, for we will soon see that the class Σ and indeed the class Δ is quite large.

The main result whose proof will be outlined here is that for any purely existential theorem of $\mathbf{G}$ of type (23) there is a theorem of $\mathbf{G}$ of type (21) that implies it. (And note that we have already in effect seen in connection with (7) that if some statement about the distinguished system of benchmarks denoted $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ is a theorem of $\mathbf{G}$, the corresponding generalization about all systems of benchmarks will be a theorem of $\mathbf{G}_0$.)

It is crucial for the proof of this result that the defining formulas Θ for our $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ can be taken to be Δ. This is verified by going through the list of defined notions in the preceding section one by one, and that list could be indefinitely extended. At the beginning, collinearity and equidistance and equilaterality were given quantifier-free definitions (11abc). Then perpendicularity and nearness were each given a Σ and an equivalent Π definition in (11df). The first problematic case comes only with (13a), or the parallelism of the lines coded by x, y and by u, v. Here the obvious definition is Π (the *non*-existence of a common point). The parallel postulate implies a less obvious equivalent definition that is Σ, in terms of the existence of points on the one line and on the other with the segment between them perpendicular to both lines. The other cases are left as exercises to the reader. The end result is that in a case of the kind that interests us, where Φ is Σ, the formulas of type (23) will be Σ, by the closure of that class under substitution of Δ functions and the other closure properties (22abc) of Σ and Δ.

The rest of the proof draws on the beautiful nineteenth-century algebra used to show that the classic problems of giving straightedge and compass constructions for duplicating the cube and trisecting the angle or constructing a regular heptagon are unsolvable (as expounded, for instance, in Papantonopoulou, 2002, chapters 11 and 12; the impossibility of squaring the circle, involving as it does the number pi, requires analytic methods). The *constructible field* is the smallest set of real numbers containing 0 and 1 and closed under the *rational operations* of addition, subtraction, multiplication, and division by non-zero numbers, as well as under extraction of square roots of positive numbers.

Tarski shows that a minimal model $\mathbf{M}_0$ of $\mathbf{G}_0$ can be obtained by taking as "points" pairs constructible numbers and defining the primitive relations just as in the Cartesian model, which is to say, just as in middle school analytic geometry. Square roots are needed to get the circle axiom. This $\mathbf{M}_0$ may be called the *constructible plane* and contrasted with the Cartesian plane in which $\mathbf{B}$ and $\mathbf{C}$ are defined by the same formulas (6ab), but the points are pairs of *arbitrary* real numbers. The points $a = (0, 0)$, $b = (1, 0)$, $c = (0, 1)$ may be taken as benchmarks to get a minimal model $\mathbf{M}$ of $\mathbf{G}$, and a crucial property of **M** is that *every point is straightedge and compass constructible from those benchmarks.*

This result has a long history. Euclid has in Book V a theory (which historians associate with Eudoxus) of ratios and proportions of lengths and other magnitudes, but he does not speak of these ratios as *real numbers* the way we would. Numbers for Euclid are positive integers. That is why it is an anachronism to speak of the Greek discovery of the incommensurability of the side and diagonal of a square as the discovery that $\sqrt{2}$ is an irrational number. In Euclid ratios are not things that can be added and multiplied.

But ratios *were* considered numbers at least as early as Omar Khayyam, *and constructions in Euclid could then be reconstrued as rational operations on ratios, as well as extraction of square roots.* This picture is in the background in Descartes' *Géométrie* and the foreground in Newton's *Universal Arithmetick*, and the theory is treated fully rigorously in Hilbert. It is also behind the results in Burgess (1984) on the possibility of reconstructing or reconstruing analytically-formulated theories of classical physics in a "synthetic" style.

The treatment of the straightedge and compass constructions related to multiplication, division, and square roots has found its way into the more thorough introductory textbooks of abstract algebra, as part of Galois theory. (See Papantonopoulou, 2022, Propositions 11.1.12 through 11.1.14, pp. 345-346.) In the constructible plane, any point $(x, 0)$ on the horizontal axis can be obtained from $(0, 0)$ and $(1, 0)$ by the geometric steps corresponding to the algebraic steps used to obtain x from 0 and 1. Likewise any point $(0, y)$ on the vertical axis is constructible from the benchmarks $(0, 0)$ and $(1, 0)$. And then any point at all (x, y) is constructible from the benchmarks as the intersection of the horizontal line through $(0, y)$ with the vertical line through $(x, 0)$.

Now suppose (23) is a theorem of **G**, with Φ quantifier-free or even just Σ. Then it must be true in the constructible plane, where every point is constructible, and if we take symbol complexes, φs, describing the construction of the pertinent xs, then (21) will be true in the constructible plane. But part of what was meant by calling **M** a "minimal" model is that *any* model of **G** has the constructible plane (or an isomorph or copy thereof) as a submodel. And it is a quite general fact of model theory, not at all tied to geometry specifically, that any Σ statement true in a submodel of a model is true in the model itself. Because (21) is Σ, it will thus be true in every model of **G**. And finally, by the Gödel completeness theorem it follows that it will be a theorem of **G**. And so we have shown (at least in outline) that behind every pure existence theorem there is a construction theorem, QED.

6 Related Issues

Let me take note now of a few questions about potential extensions or refinements of the results discussed so far.

First question. A proof has been outlined for the following "metatheorem": For every pure existence theorem *there exists* a construction theorem implying it, or more long-windedly, for every proof π of a pure existence theorem Φ, *there exists* a pair of proofs (ρ, σ) with ρ being a proof of a construction theorem Ψ, and σ being a proof of the logical implication $\Phi \to \Psi$. The question now arises: is the "metaproof" that has been given or sketched for this result "constructive" in the sense of that there is an effective procedure that applied to any given pertinent π will compute a suitable (ρ, σ)?

Yes, there is, and it goes like this: Nineteenth-century set theory established that given a finite alphabet of symbols, one can effectively list all finite strings τ of symbols from this alphabet in a sequence $\tau_0, \tau_1, \tau_2, \ldots$ indexed by natural numbers. Given a pertinent π, for any τ on the list one can effectively decide or compute whether it is a suitable pair (ρ, σ) for π. So just go down the list until a suitable pair is found. This procedure is "effective" or "computable" in the sense that in any given case it is *possible in principle* (not worrying about the amount of time needed to carry it out or the amount of space needed to record intermediate results), for a digital machine to carry it out.

But it is not *feasible in practice.* The function taking one from π to (ρ, σ) in the manner described is computable or recursive but might take astronomically long to compute. And this raises the question: *Is there any more efficient way to find a construction theorem behind any given existence theorem?* To begin with, could there be a procedure producing not just a recursive but a *primitive* recursive function? Such questions of "complexity theory" virtually always arise when something is proved to be computable or recursive. They are often quite difficult to answer, or anyhow, often go unanswered for a long time after they are originally posed. We need not be especially troubled, therefore, if the question just enunciated turns out to be one that has to be left open for the present and foreseeable future.

Second question. We may ask *how far can results similar or analogous to those established in this note for straightedge and compass construction be obtained for other kinds of geometric construction?* This is obviously not a single question but a series of them for different species of constructibility. Probably the most interesting case, and the only one I will explicitly discuss, is that of so-called *neusis* or *verging* constructions (known to be equivalent to *origami* constructions). Where straightedge and compass permit the duplication of the square and the bisection of an angle and the construction of a regular pentagon, they do not allow for the duplication of the cube or the trisection of an angle or the construction of a regular heptagon. These three things do have neusis or verging constructions.

By definition these are constructions using in addition to the compass a "marked ruler" or "notched straightedge." They are equivalent to constructions involving drawing of conic sections (using a string as mentioned in a note early on here in the case of the ellipse, or in some other way) and finding the intersections of such curves. They are equivalent also to constructions using a curve called *the conchoid of Nicomedes*, for the drawing or producing of which there is a mechanism whose invention is attributed by a commentator (one Eutocius of Ascalon) to Archimedes.

This class of constructions has been analyzed algebraically as thoroughly as have been straightedge and compass constructions. (See Papantonopoulou, 2002, pp. 357ff.) The algebraic counterpart of going beyond straightedge and compass constructions to neusis or verging constructions is in jargon that of going beyond "Euclidean" to "Vietan" fields, or in plainer language, from solving quadratic to solving cubic equations. This last is a problem that itself has long history, in which the outstanding figures are Omar Khayyam in the eleventh or twelfth century, for geometric solutions by intersecting conics, and Gerolama Cardano in the sixteenth century, for algebraic solutions by radicals.

There is much more that could be said, but what is most relevant here is that since the algebra has been so thoroughly analyzed it ought not to be very difficult to extend the methods of the present note to establish analogues of (A) and (B) for the wider class of constructions. Still, since I have not worked through the details, I can for the moment only advance this as a very plausible conjecture rather than claim it as a theorem.

Third question. How widely beyond the case of purely existential theorems can the notions and results of the present note be extended? What has been shown so far is that there is a construction theorem behind any purely existential theorem, that is, any theorem of form (23) above with Φ quantifier-free; and a moment's thought shows that this results extends to the case where Φ consists of a string of existential quantifiers follow by something quantifier-free. The next cases to consider would be those of *universal-existential* theorems, of the type

(24) $\forall x_1, \forall x_2 \ldots \forall x_m \ldots \exists y_1, \exists y_2 \ldots \exists y_n \Phi(x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_n)$

with Φ quantifier-free.

The methods used here do establish *some* results of this kind. For instance, it can be shown not just how to construct an equilateral triangle but how for any line segment to construct an equilateral triangle *having that segment as a side* (which is the actual result in Euclid's Proposition 1). For the construction given above establishes that there is an equilateral triangle having the segment ab joining the first two benchmarks as one side, and we have already seen that what is provable in $\mathbf{G}$ to hold for a, b, c, is provable in $\mathbf{G}_0$ to hold for *any* system of benchmarks, and that for any two distinct points there is a system of benchmarks of which they are the first and the second. The general case is, however, so far as I know open.

The next case to consider would be that of *existential-universal* theorem, of the type

(25) $\exists y_1, \exists y_2 \ldots \exists y_n \forall z_1, \forall z_2 \ldots \forall z_p \Phi(y_1, y_2, \ldots, y_n, z_1, z_2, \ldots, z_p)$

with Φ quantifier-free. But it can be shown that there is *not* always a construction theorem behind such a result. It will be worthwhile to spell out a counterexample, since it will illustrate an important point about the distinction between Euclid-type construction-oriented geometry and Brouwer-type constructivist or intuitionist logic, a theme that thus far has been left in the background here since it was initially enunciated. For the counterexample will depend on the law of excluded middle.

It is known that *proportionality* or equality of ratios of segments, written in traditional notation thus:

$$st : uv :: wx : yz$$

is expressible in the language of $\mathbf{G}$, and indeed is of class Δ. Hence it is expressible that the quintuple $\boldsymbol{u}$of points u_1, u_2, u_3, u_4, u_5 are so related that

(26a) they all lie on the same line and in the order listed

(26b) u_1 and u_5 lie at the same distance from u_2 on opposite sides, so that u_1u_2 and u_1u_5 are in the ratio 1 to 2

(26c) u_1u_2 is to u_1u_3 as u_1u_3 is to u_1u_4 and as u_1u_4 is to u_1u_5, so that u_1u_2 and u_1u_3 are in the ratio 1 to $\sqrt[3]{2}$

Since $\sqrt[3]{2}$ is the number that has to be constructed in the problem of the duplication of the cube, when (26abc) hold for u_1, u_2, u_3, u_4, u_5 let us say the quintuple $\boldsymbol{u}$ is a *cube-duplicator*, and write $\mathbf{Q}\boldsymbol{u}$. Similarly write $\mathbf{P}\boldsymbol{u}$ to express that the quintuple $\boldsymbol{u}$ is *pentagonal* in the sense that u_1, u_2, u_3, u_4, u_5 in that order are the successive vertices of a regular pentagon. Let us also write $\exists\boldsymbol{u}$ to abbreviate $\exists u_1 \exists u_2 \exists u_3 \exists u_4 \exists u_5$. Then the counterexample to be discussed will be statable in this abbreviated notation thus:

(27) $\exists\boldsymbol{u}(\mathbf{Q}\boldsymbol{u} \vee (\mathbf{P}\boldsymbol{u} \mathbin{\&} \neg\exists\boldsymbol{v}\mathbf{Q}\boldsymbol{v}))$

This is logically equivalent to the following existential-universal formula:

$$\exists \boldsymbol{u} \forall \boldsymbol{v}(\mathbf{Q}\boldsymbol{u} \vee (\mathbf{P}\boldsymbol{u} \ \& \ \neg\mathbf{Q}\boldsymbol{v}))$$

It is provable in **G** that $\exists \boldsymbol{u}\mathbf{P}\boldsymbol{u}$. It is neither provable nor disprovable in **G** that $\exists \boldsymbol{u}\mathbf{Q}\boldsymbol{u}$, for there is a cube-duplicator in the full Cartesian plane but not in the restricted constructible plane. To prove (27), note that by excluded middle (in any given model) there either does or does not exist a cube-duplicator. If there does, any cube-duplicator may be taken for the $\boldsymbol{u}$ in (27) and the disjunction will hold because its first disjunct does. If there does not, then the second conjunct of the second disjunct in (27) will hold, and if we take for $\boldsymbol{u}$ any regular pentagon—and we know there will be one—the first conjunct will hold as well, and hence the second disjunct as a whole, and hence the disjunction as a whole. So either way, (27) holds. The argument here is only slightly fancier than that used for (4).

Now consider any quintuple $\boldsymbol{\psi} = (\psi_1, \psi_2, \psi_3, \psi_4, \psi_5)$ of operations compounded out our basic operations $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$, and so on, and apply it to the benchmarks to obtain the quintuple of points

(28) $\psi_1(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}), \psi_2(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}), \psi_3(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}), \psi_4(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}), \psi_5(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$

Part of what was meant by calling the constructible plane a "submodel" of the Cartesian plane is that when applied to the benchmarks or any other elements of the smaller plane, $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$, and the rest *will give the same results whether we are thinking of them as operators on that plane or on the larger plane.* So the notation (28), which may be abbreviated as or $\boldsymbol{\psi}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$, or even just as $\boldsymbol{\psi}$, is unambiguous. A construction theorem implying the existence theorem (27) would have to look like this:

(29) $\mathbf{Q}\boldsymbol{\psi} \vee (\mathbf{P}\boldsymbol{\psi} \ \& \ \neg\exists \boldsymbol{v}\mathbf{Q}\boldsymbol{v})$

But no such thing can be a theorem of **G**, since nothing like (29) can be true in both the constructible and the Cartesian plane. To be true in the smaller plane, where there are no cube-duplicators and hence the first disjunct is false, the second disjunct and in particular its first conjunct, must be true. But that means that $\boldsymbol{\psi}$ is giving us a regular pentagon. By contrast, to be true in the larger plane, where the second conjunct of the second disjunct, and hence the second disjunct as a whole, is false, the first disjunct must be true. But that means that $\boldsymbol{\psi}$ is giving us a cube-duplicator. And nothing is both a regular pentagon and a cube-duplicator.

So while the main results (A) and (B) above hold for *purely* existential theorems as was seen in §3, we have just seen this success does not extend to *arbitrary* existence theorems. One can only conclude that **G** is "constructive" in one sense and "non-constructive" in another sense. But that was true already of Euclid and does *not* in itself demonstrate infidelity of the Tarskian approach to the Euclidean.

7 Not Unrelated Developments

At the suggestion of a reader of an earlier draft of this note, let me before closing describe some previous work on constructivity in geometry by Nancy Moler and Patrick Suppes (1968). It is clearly relevant to the present discussion, even though its results do not directly quite yield (A) and (B) above, while inversely and more emphatically, what is done here certainly does not accomplish the aims of the joint authors' more ambitious project. A certain gap or space exists between the two investigations, which in some respects are even opposite in their perspectives.

Here, in attempting to modernize the Euclidean context of Beere's and Morison's historical discussion of existence theorems versus construction problems, I have simply adopted Tarski's axiomatization, whose primitives are predicates, and whose postulates include items (notably Pasch's axiom) of universal-existential form. The tradition originating with Moler and Suppes, by contrast, seeks to do for geometric theories what has already been done for any number of arithmetic or algebraic theories, namely, to formulate them with only function symbols rather than predicates as primitives—constants are also allowed, since they may be counted as zero-place function symbols—and with all axioms as so-called quantifier-free (really, purely universal) statements. It might not be too inaccurate to say that whereas the interest here has been in finding construction theorems "behind" existence theorems, the interest in the tradition under discussion is more in simply avoiding existence assertions in favor of construction assertions.

Moler and Suppes are aiming to give a different axiomatization of what is in some sense supposed to be the "same" geometric theory as the Tarskian formulation they contrast with it. In the end they establish the agreement of their proposal with earlier models by means of *representation theorems*, which play a large role elsewhere in Suppes' *œuvre*. It is, however, also possible to give a more syntactic definition of the sort of thing that is wanted:

(30a) The function symbols of the new theory are to be definable in terms of the predicates of the old theory.

(30b) The axioms of the new theory are to be deducible from those of the old theory together with the definitions just mentioned of the function symbols.

(30c) The old predicates are to be definable in term of the function symbols of the new theory.

(30d) The axioms of the old theory are to be deducible from those of the new theory together with the definitions just mentioned of the predicates.

What is meant in (30a) is definability in exactly the same sense in which $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ here were defined in terms of $\mathbf{B}$ and $\mathbf{C}$ and $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$. It should be mentioned that it turns out to be needful, in order to avoid trivialization, for Moler and Suppes also to introduce constants $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ denoting three non-collinear points, like the benchmarks in this note. Once this is done there is a very general way, not specifically geometrical, and perhaps not optimal, but always available, to introduce function symbols $\boldsymbol{r}, \boldsymbol{s}, \boldsymbol{t}, \ldots$ for the new theory that will be interdefinable with the predicates $\mathbf{R}$,$\mathbf{S}$,$\mathbf{T}$,... of the old theory.

How this may be done is sufficiently illustrated by the case of a single two-place predicate $\mathbf{R}$. We associate with the predicate a symbol $\mathbf{r}$ for (a version of) the *characteristic* or *indicator* function r for $\mathbf{R}$, which for given x and y as inputs returns as output the item denoted $\boldsymbol{a}$ if $\mathbf{R}(x, y)$ holds, and that denoted $\boldsymbol{b}$ if not. Replacing $\mathbf{R}(x, y)$ by $\boldsymbol{r}(x, y) = \boldsymbol{a}$ will be enough to enable one to accomplish (52cd), and inversely, accomplishing (56ab) can be carried out by the usual methods of eliminating function symbols (as in Boolos et al. 2007, §19.4, pp. 255-258).

hat usual method will involve introducing some existence axioms like (9). Moving in the opposite direction will make such axioms superfluous, but will not by itself eliminate all the *old* existence axioms already present in the original version of the theory. It is, however, as mentioned in passing above, the avowed aim of Moler and Suppes in their geometric context to do just that: eliminate all existential axioms from the new theory, finding an axiomatization using only universal formulas. As they say (p. 143)

> In view of the highly constructive character of Euclidean geometry, it seems natural to strive for a formulation that eliminates all dependence on purely existential axioms...

And there exists what may superficially appear to be an all-purpose method for eliminating existential axioms at the cost of introducing new function symbols, a method of which some readers will have heard, called "reduction to Skolem normal form" or simply "Skolemization" (as in Boolos et al. 2007, §19.2, pp. 247-253). But in fact Skolemization is a red herring, not really relevant to the Moser-Suppes project, or to that of this note. (And readers unfamiliar with it may simply skim or skip the next paragraph here, which explains why.)

In general, Skolemization does not even guarantee that the new functions introduced will be definable in terms of what was there before their introduction, as required by (30a). That aside, it is certainly not something Moler and Suppes wish to rely on, since they continue the passage just quoted with these words:

> ... but not, of course, by use of some wholly logical, non-geometric method of quantifier elimiination.

And Skolemization is the very paradigm of a wholly logical, non-geometric method of quantifier elimination. Rather, Moler and Suppes call for primitive function symbols representing "familiar constructions," of the kind whose use is acknowledged in the tradition of geometric construction problems, and something like this requirement of connection with historically given functions or operations has also been implicitly or tacitly assumed in the approach taken in this note.

Moler and Suppes take as their "familiar" geometric constructions *finding the intersection of lines*, a version of which was the first (8a) of the basic constructions (8) used in §3 above, and *the laying out of segments* as in Euclid, Book I, Proposition 2. Here the aim has been similar, to use only what are recognizably versions of the familiar construction steps (8abc).[3] Moler and Suppes do not concern themselves with finding the intersection of circles as was done in this note with (8bc), because they are working not with the Euclidean geometry that has been the focus here, but with the weaker Pythagorean geometry, which lacks the circle axiom.

The difference between the two geometries parallels the difference in algebra between Euclidean ordered fields, in which every positive element has a square root, and Pythagorean fields, satisfying the weaker condition that every sum of squares has a

[3] That the primitives such as $\boldsymbol{\beta}$ used here are importantly different from the ordinary straightedge-and-compass construction of the intersections of a lines and circles is evident and must be acknowledged. Tibor Beke (personal communication) has in this connection pointed to the difficulty, which I take to be the impossibility, of determining by ordinary staightedge-and-compass constructions which of three collinear points lies between the other two. This *can* be determined if one is able to determine, given any two of the points, which is nearer the third. For if C and B are equidistant from A, then A is the in-between point; and otherwise, if say C is further away than is B from A, then C cannot be the in-between point, and hence A and B must lie on the *same* side of C, in which case whichever is nearer to C is the in-between point. Now by ordinary straightedge-and-compass constructions one can obtain the line on which the three points lie, and on that line the midpoint of the segment AB, and therewith obtain the circle with center on that line passing through both points A and B. With the distinction made by $\boldsymbol{\beta}$ between the *proximal* and *distal* points of intersection of the line in question with the circle in question, relative to a system of benchmarks that may be freely chosen, and in particular so chosen that C is the origin, the proximal point of intersection will be the nearer to and the distal point of intersection the farther from that origin, and the problem can be solved.

square root. The Euclidean as opposed to Pythagorean case was taken up later, by Horst Seeland (1978). By the time we come to the survey of Victor Pambuccian (2008), multiple geometries have come into play (including the stronger geometry of neusis or verging constructions alluded to earlier, and the oldest kind of *non*-Euclidean geometry, the hyperbolic), and the bibliography of this tradition has grown to nearly 150 items. There is a great deal here that might be looked over again from the somewhat different perspective of this note.

Moler and Suppes encounter the difficulty, which they describe as "akin to division by zero," that their functions are not always defined (as the intersection of lines does not exist when the lines are parallel). Indeed, they cite such difficulties as probably the main reason why something like their project was not carried out much earlier in the history of formalized geometry. Here such difficulties were dealt with by conventionally assigning a sort of null value, the origin, to waste cases where this happens. Moler and Suppes take the bolder line of simply allowing some functions in some models to be partial, and generalizing model-theoretic notions such as submodel and isomorphism to apply to this wider-than-usual range of models.

The paper of Moler and Suppes was published in a journal founded by Brouwer after Hilbert had, at a sort of climax of the *Grundlagenstreit*, manœuvred him off the editorial board of the *Mathematische Annalen.* And in a footnote at the bottom of the first page of their work they refer to Brouwer's chief disciple and tell us the following:

> It is a pleasure to dedicate this paper to Professor Heyting on the occasion of his seventieth birthday. In view of his long interest in constructive mathematics and in geometry, we believe the subject of our paper make it particularly appropriate to dedicate it to him.

For there is indeed a long tradition among intuitionists of concern with geometry, from parts of Brouwer's dissertation onward, through work by Heyting partly on problems to which Brouwer directed him, and then on to Heyting's student Dirk van Dalen and beyond. Still. overlap with the Moler-Suppes tradition may be somewhat limited insofar as the latter sticks to classical logic.

A genuine intuitionist would not allow definitions by cases, as repeatedly used here (in particular, in the definitions of $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$) unless the case hypotheses are decidable. And just as in intuitionistic algebra the identity of real numbers is not assumed decidable, so also for the coincidence of planar points in intuitionistic geometry. There exists a large and more recent body of work produced by Michael Beeson, culminating in Beeson (to appear), specifically concerned with geometry that is "constructive" in the double sense of being *both* occupied with straightedge and compass constructions *and* based on Brouwer's and Heyting's intuitionistic or constructivistic logic. But this work inevitably has a rather different flavor from the work discussed or reported here.

References

[1] Beere, J., & Morison, B. A mathematical form of knowing how: The nature of problems in Euclid's geometry. Unpublished manuscript.

[2] Beeson, M. Constructive geometry. To appear. Available at
http://www.michaelbeeson.com/research/papers/ConstructiveGeometryFinalPreprintVersion.pdf

[3] Bernays, P. (1964). On platonism in mathematics. In P. Benacerraf & H. Putnam (Eds.), *Philosophy of mathematics: Selected readings* (pp. 274-286). Englewood Cliffs, NJ: Prentice-Hall.

[4] Blumenthal, O. (1935). Lebensgeschichte. In D. Hilbert, *Gesammelte Abhandlungen, Dritter Band* (pp. 388-429). Springer.

[5] Boolos, G. S., Burgess, J. P., & Jeffrey, R. C. (2007). *Computability and Logic* (5th ed.). Cambridge University Press.

[6] Burgess, J. P. (1984). Synthetic Mechanics. *Journal of Philosophical Logic*, 4, 379-395.

[7] Chihara, C. (1973). *Ontology and the Vicious Circle Principle.* Ithaca, NY: Cornell University Press.

[8] Collingwood, S. D. (Ed.). (1899). *The Lewis Carroll Picture Book.* London: Collins.

[9] Frege, G. (1953). *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number* (J. L. Austin, Trans.; 2nd rev. ed.). New York, NY: Harper & Brothers.

[10] Heath, T. (1968). *The Thirteen bottomks of Euclid's Elements* (2nd ed.). Cambridge: Cambridge University Press.

[11] Hilbert, D. (1902). *The Foundations of Geometry.* LaSalle: Open Court.

[12] Moler, N., & Suppes, P. (1968). Quantifier-free axioms for constructive plane geometry. *Compositio Mathematica*, 20, 143-152.

[13] Netz, R. (1998). Greek Mathematical Diagrams: Their Use and Their Meaning. *For the Learning of Mathematics*, 18, 33-39.

[14] Netz, R. (1999). Proclus' division of the mathematical proposition into parts: How and why was it formulated? *Classical Quarterly*, 49, 282-303.

[15] Pambuccian, V. (2008). Axiomatizing geometric constructions. *Journal of Applied Logic*, 6, 24-46.

[16] Papantonopoulou, A. H. (2002). *Algebra: Pure and Applied.* Upper Saddle River: Prentice-Hall.

[17] Seeland, H. (1978). *Algorithmische Theorien und konstruktive Geometrie.* Stuttgart: Hochschulverlag.

[18] Tarski, A., & Givant, S. (1999). Tarski's system of geometry. *Bulletin of Symbolic Logic*, 2, 175-214.

[19] Weisstein, E. Heptadecagon. Available at
http://mathworld.wolfram.com/Heptadecagon.html

Definiteness in early set theory*

Laura Crosilla & Øystein Linnebo

1 Introduction

Philosophers of mathematics sometimes talk about definiteness and kindred notions. For example, Michael Dummett and others have claimed that the concepts of ordinal number and set are indefinitely extensible, in the sense, roughly, that any definite totality of instances of the concept can be used to define yet another instance, outside of the mentioned totality.[1] It follows that there can be no definite totality of absolutely all ordinals or sets. Others, including both philosophers and mathematicians, react with impatience and incredulity. Furthermore, mainstream mathematicians are today less prone to talk about definiteness.[2] So, one might suspect, the philosophers' notions of definiteness are just detached philosophy and mathematically sterile.[3]

We contend that this reaction is unwarranted. Far from being mathematically sterile, we show that there are philosophically interesting notions of definiteness that lend themselves to precise mathematical explication and ultimately also to some interesting mathematics. In particular, during and after the Cantorian revolution that ushered in modern set theory, various notions of definiteness figured in the work of some of the most prominent mathematicians. Debates about these notions helped shape the emerging theory of sets.

To defend our contention, we discuss some prominent mathematicians' appeals to definiteness in the first half century of Cantorian set theory, roughly between the 1880s and the 1930s. We find it necessary to disentangle two very different forms of definiteness. First, a condition (by which we mean an open formula, possibly with parameters) can be definite in the sense that, given any object, either the condition applies to that object or it does not. We call this *intensional definiteness.* Second, a condition or collection can be definite in the

*Laura Crosilla, Department of Philosophy, University of Florence and Øystein Linnebo, IFIKK, University of Oslo.

[1] See, e.g., (Dummett, 1963, pp. 195–96), (Dummett, 1991, pp. 316–17), and (Dummett, 1993, p. 441), as well as (Shapiro and Wright, 2006) and (Linnebo, 2018) for discussion and analysis.

[2] As we will see in Section 4, Solomon Feferman is an important exception.

[3] See, e.g., the dismissive attitude found in (Boolos, 1998, p. 224) and (Burgess, 2004, p. 205).

sense that, loosely speaking, a totality of its instances or members has been circumscribed. We call this *extensional definiteness*. Whereas intensional definiteness concerns whether an intension applies to objects *considered one by one*, extensional definiteness concerns *the totality of objects* to which the intension applies. Thus, there are two importantly different families of notions of definiteness. We show that notions from both families are invoked repeatedly in the literature during the emergence of set theory.

We also investigate how these two forms of definiteness admit of precise mathematical analyses. A natural explication of intensional definiteness is in terms of bivalence. That is, a condition $\varphi(x)$ is intensionally definite just in case, for any given object a, either $\varphi(a)$ is true or it is false. For example, 'x is transitive' is intensionally definite on the domain of sets, whereas 'x is small' is not. Of course, since we talk about truth and falsity, this explication takes place in a metalanguage. In the context of intuitionistic logic, a closely related explication is available in the object language, namely, that the Law of Excluded Middle (LEM) holds for the condition:

$$\forall x(\varphi(x) \vee \neg\varphi(x))$$

The idea of extensional definiteness leaves more choice. We distinguish two main types of explication. One takes its departure from the Aristotelian idea of potential infinity. According to Aristotle, only finitely many numbers can coexist; the rest are merely potential. As we will see, Cantor liberalized this idea massively so as to allow various huge infinite collections to coexist, while denying that all sets coexist. To be extensionally definite can then be analyzed as (possible) coexistence. Another explication is based on a novel idea due to the great mathematician Hermann Weyl.[4] Loosely speaking, the idea is that extensional definiteness is a matter of proper demarcation. Remarkably, this intuitive idea is given a precise logical explication. A collection is extensionally definite just in case quantification over that collection "behaves classically". As we shall see, the resulting notion is less demanding than the liberalized Aristotelian notion of possible coexistence.

2 Cantor

Recall that a condition is said to be intensionally definite when, for any given object, either the condition applies to the object or it does not. We begin with some appeals to intensional

[4] See our (Crosilla and Linnebo, 2024).

definiteness in the work of Cantor.

> I call a manifold (an aggregate [Inbegriff], a set) of elements, which belong to any conceptual sphere, *well-defined* [wohldefiniert], if on the basis of its definition and in consequence of the logical principle of excluded middle, it must be recognized that it is internally determined whether an arbitrary object of this conceptual sphere belongs to the manifold or not, and also, whether two objects in the set, in spite of formal differences in the manner in which they are given, are equal or not. (Cantor, 1882)[5]

We see that Cantor characterizes a "manifold" as "well-defined" just in case two conditions are met. ('Manifold' is his generic term for any kind of collection.) First, it must be "determined whether an arbitrary object" of the appropriate sort "belongs to the manifold or not". Second, it must be determined whether any two members of the manifold are equal or not. In short, a "manifold" is well-defined just in case it is associated with an intensionally definite membership condition and a criterion of identity for its members.[6]

A similar characterization of a well-defined collection can be found in work by Dedekind dated 1872–74:

> A *set* or *collection* is determined when for every thing it can be judged whether it belongs to the set or not. (Dedekind 1872, quoted in (Ferreirós, 2023, p. 261))

Yet another example from the same period is Frege, whose preferred notion of a collection is that of an extension of a sharply defined concept—or, as we may put it, an intensionally definite concept.[7]

Thus, following Dedekind and Frege, Cantor began with a notion of set that emphasized the provision of an intensionally definite membership condition. This may be slightly more guarded than the so-called naïve conception of sets, which allows any condition to define a set. After all, the three mentioned mathematicians emphasize that the proposed membership condition has to be intensionally definite. Cantor may also assume that the set is contained in a single "conceptual sphere" (see note 6). Regardless, Cantor soon realized that greater care is needed.

[5] Translation by (Tait, 2009, p. 271).

[6] As the quote shows, Cantor also presupposes that the "manifold" belongs to a single "conceptual sphere". He does not, however, explicitly lay that down as a requirement for the manifold to be well-defined, let alone explain how such a requirement is to be understood.

[7] See esp. (Frege, 1893), as well as (Parsons, 2012) for discussion.

An important step in that direction is Cantor's (1895/97, p. 481) famous definition of set:

> By a 'set' we understand every collection into a whole [Zusammenfassung] M of determinate, well-distinguished objects [bestimmten wohlunterschiedenen Objekten] m of our intuition or our thought (which will be called the 'elements' of M). We write this as: $M = \{m\}$. (Cantor 1895, p. 481; translation by (Florio and Linnebo, 2021)).

Here we find a somewhat different characterization of what a set is. We start with some objects m. Indeed, since we may start with a plurality of objects, it is tempting to use the resources of contemporary plural logic and say that we start with one or more objects mm.[8] These objects are then "collected" or gathered into a whole M, which is the set of the objects in question. Using plural logic, we may write this as $M = \{mm\}$. But again, successful definition of a set is not guaranteed. Cantor requires the operation of collecting many objects into a single set to be applied to some "determinate, well-distinguished objects" mm.

How should this requirement be understood? At the very least, the notion of being one of the objects mm to be collected into a single set must be intensionally definite. That is, for any given object x it must be determined whether x is one of mm or not. But the requirement that the operation of collecting into a whole be applied to some "determinate, well-distinguished" objects can also be understood to demand more. To see how, a distinction is useful. A property of some objects is said to be *distributive* just in case the ascription of the property to some objects is equivalent to the ascription of it to each of these objects. By contrast, a property is said to be *collective* or *non-distributive* when it says something about all of the objects together. For example, 'the students are French' involves a distributive predicate, whereas 'the students surround the building' involves a collective one. Returning to the property of being some "determinate, well-distinguished objects", the question is whether this should be understood in a distributive or a collective way.

We submit that the collective reading is more plausible. To have some objects that can be collected into a single set it is not enough that each of the objects is "determinate" and that any two of them are "well-distinguished"; rather, the objects in question have to be

[8] See (Boolos, 1984) for a seminal contribution, as well as (Florio and Linnebo, 2021) for a recent survey and assessment.

collectively determinate, that is, we must have pinned down which objects *they* are. As we will see shortly, this is also the reading that Cantor later says he intended. It is plausible, therefore, to understand Cantor's 1895 characterization of a set as requiring that a set be *extensionally* definite.

In fact, even prior to 1895 Cantor appears to have invoked the idea of extensional definiteness. One example concerns the generative approach of (Cantor, 1883) to transfinite numbers, which are close to what we now call ordinal numbers.[9] His first principle of number generation states that for any number α, we may add one so as to obtain its successor $\alpha+1$. A more interesting second principle states that:

> if any definite [bestimmte] succession of defined integers is put forward of which no greatest exists, a new number is created by means of this second principle of generation, which is thought of as the *limit* of those numbers; that is, it is defined as the next number greater than all of them. (Cantor, 1883, pp. 907–908)

In short, the second principle allows us to form the least upper bound of any "definite succession" of numbers.[10]

Assume, following Cantor, that the natural numbers form a "definite succession". Then the second principle of number generation allows us to define their limit, namely ω. Repeated application of the first principle now yields all numbers of the form $\omega + n$, for n a natural number. Since this too is a "definite succession", we can take their limit, so as to obtain $\omega + \omega$. Clearly, we can generate larger and larger ordinal numbers. This yields what Cantor calls "the extended number sequence" (p. 912).

What is it for some numbers to be a "definite succession" and thus, by the second generating principle, to possess a limit or least upper bound? We submit that this is best understood as some form of extensional definiteness. Unlike the sequence of natural numbers, which is surpassed by their limit, ω, the extended number sequence is "absolutely infinite" (p. 916). As such, it has no limit but (says Cantor, quoting Albrecht von Haller) "lie[s] always ahead of me" (ibid.). This fundamental difference between the natural number sequence and the extended number sequence cannot be explained in terms of intensional definiteness alone. For applied to any given object, the notion of being an extended number is just as clear and well-defined as that of being a natural number: either this object is generated as an (extended) number or it is not. The difference pertains rather to the

[9] See (Tait, 2009) pp. 281-82 for discussion.

[10] Cantor's use of the second principle makes it clear that 'integer', in the quoted passage, is used as synonymous with 'number'. A third principle is also formulated, which need not concern us here.

extended number sequence *in its entirety*, namely, that it is not extensionally definite. Thus, we contend, Cantor's notion of a "definite succession" is best understood as involving some form of *extensional* definiteness.[11]

A second example of extensional definiteness in Cantor prior to 1895 is found in his critical 1885 review of (Frege, 1953). Cantor complains that Frege "overlooks altogether the fact that the 'extension of a concept' is, in general, something quantitatively completely indeterminate" (Ebert and Rossberg, 2009, p. 346).[12] Whether the extension of a concept is "quantitatively determinate" is, we contend, a question about extensional definiteness, not intensional. First, Cantor makes it explicit that the question concerns the *extension* of a concept, not the concept itself. Second, the word "quantitatively" makes it clear that the question is whether all the instances of the concept, *taken together*, can be assigned a number or quantity, not whether it is well-defined what it is for the concept to apply to any given *single* object.

Let us return to our discussion of what it takes for a set to be well-defined. We argued that it is most plausible to interpret Cantor's famous 1895 characterization of a set as requiring that the one or more objects to be collected together into a whole (namely, a set) be collectively well-defined or circumscribed. That is, we find it natural to read Cantor 1895 as requiring that a set be extensionally, not just intensionally, definite. After the set-theoretic paradoxes were discovered and started to be discussed among German mathematicians around 1897, Cantor wrote two letters that clearly and unequivocally make this requirement. These letters also take some important steps towards explaining how he understood this notion of extensional definiteness.

We begin with an 1897 letter to Hilbert.

> I say of a set that it can be thought of as *finished* [...] if it is possible without contradiction (as can be done with finite sets) to think of *all its elements as existing together*, and so to think of the set itself as *a compounded thing for itself*; or (in other words) if it is *possible* to imagine the set as *actually existing* with the totality of its elements. (Ewald, 1996, p. 927)

[11] Cantor thus comes close to the Burali-Forti paradox, which is the paradox of the ordinal associated with the well-ordered sequence of all ordinal numbers. There is a scholarly debate about whether or not Cantor already in 1883 has some understanding of this paradox; cf. (Moore and Garciadiego, 1981), (Ferreirós, 1999, p. 292) and (Tait, 2009, p. 281). We need not here take a stand on this question.

[12] There is a scholarly debate about whether this should be seen as warning Frege of the disaster that struck seventeen years later when he received a letter from Bertrand Russell (Tait, 2009; Ebert and Rossberg, 2009). Cf. also (Dummett, 1994, p. 26) who argued that Cantor was well ahead of Frege in seeing the importance of indefinite extensibility. Once again, we need not take a stand on this matter.

Cantor proceeds to commend the French and Italian words for set ('ensemble' and 'insieme', respectively), which emphasize that the elements of a set must exist together. He continues:

> And so too in the first article of [Cantor 1895, which was quoted above], I define 'set' (meaning thereby only the finite or transfinite) at the very beginning as an 'assembling together' [Zusammenfassung]. But an 'assembling together' is only possible if an '*existing together*' [*Zusammensein*] is *possible.* (928)

We wish to make two observations. First, the property of "existing together" is clearly collective, not distributive. So Cantor in 1897 reads his own 1895 characterization of a set as requiring that a set be extensionally, not just intensionally, definite. Second, the intended notion of extensional definiteness is now becoming clearer. It is a matter of all of the elements of the desired set coexisting so that the set resulting from collecting them together can be regarded as "finished".

An 1899 letter to Dedekind continues this explication of the sense in which each set is extensionally definite.

> [I]t is necessary [. . .] to distinguish two kinds of multiplicities (by this I always mean *definite* multiplicities).
>
> For a multiplicity can be such that the assumption that *all* of its elements 'are together' leads to a contradiction, so that it is impossible to conceive of the multiplicity as a unity, as 'one finished thing'. Such multiplicities I call *absolutely infinite* or *inconsistent multiplicities.* [. . .]
>
> If on the other hand the totality of the elements of a multiplicity can be thought of without contradiction as 'being together', so that they can be gathered together into '*one* thing', I call it a *consistent multiplicity* or a 'set'. (Ewald, 1996, p. 931-932)

Cantor's thought here appears to be that there is an intrinsic difference between multiplicities that form sets and multiplicities that do not, and that this intrinsic difference *explains* why some but not all multiplicities are eligible for set formation. A set is characterized as a "finished" collection, all of whose elements can "exist together" or be imagined as "actually existing". So for a multiplicity to be eligible for set formation, it must be capable of being regarded as "finished", and its elements must be capable of "existing together". A multiplicity that is capable of the sort of "completion" is thus intrinsically suited for set formation, whereas a multiplicity that resist such "completion" is intrinsically unsuited for set formation. For instance, since the multiplicities of everything thinkable and of all ordinals resist "completion", there can be no universal set or set of all ordinals.

Some will complain that the quasi-temporal language of coexistence and being finished has no place in the analysis of an abstract subject matter that exists independently of us. It must be admitted that Cantor's remarks about legitimate set formation [Zusammenfassung] and his associated response to the paradoxes are loose and underdeveloped. We nonetheless find his remarks suggestive.

We would therefore like to propose a way to understand Cantor's remarks about coexistence and being finished on which these remarks have genuine explanatory potential.[13] The idea is to understand Cantor's notion of extensional definiteness as a potentialist notion of *completability.* Let us explain. In stark contrast to Cantor, Aristotle held that only potential infinities are coherent, not actual ones. For example, however many natural number are instantiated—and thus, according to Aristotle, exist—it is possible to instantiate yet more. Writing '$\textsc{Succ}(m, n)$' for the claim that m is directly succeeded by n, he thus endorses:

$$\Box\forall m\Diamond\exists n\, \textsc{Succ}(m, n) \tag{1}$$

However, Aristotle denies that it is possible to complete this process of producing successors of natural numbers:

$$\neg\Diamond\forall m\exists n\, \textsc{Succ}(m, n) \tag{2}$$

Another way to put this is that the notion $\mathbb{N}(x)$ of being a natural number is incompletable, in the sense that it is impossible for some objects to coexist which are all the numbers there ever could be:

$$\neg\Diamond\exists xx\Box\forall y(y \prec xx \leftrightarrow \mathbb{N}(y)) \tag{3}$$

We here treat every plurality as "finished". That is, when we generalize plurally over one or more objects, we assume that the objects in question coexist.[14]

[13] See (Parsons, 1983b), (Linnebo, 2013) and (Studd, 2013) for attempts to realize that explanatory potential.

[14] These are fairly standard assumptions in the contemporary literature on the modal logic of plurals; see (Florio and Linnebo, 2021, ch. 10 and references therein). The contemporary notion of plurality differs, in this regard, from Cantor's less demanding notion of a "multiplicity", which permits multiplicities to be "inconsistent' or "unfinished". As Charles Parsons (1977, Sect. III) observes, Cantor's inconsistent multiplicities are better understood as Fregean concepts than some given objects.

Cantor, of course, is famous for rejecting Aristotle's claims (2) and (3). The natural numbers are completable, he argues, and therefore form a set. Nevertheless, in the quoted letters Cantor can be understood as making claims that are logically analogous to Aristotle's. Let us write '$\textsc{Set}(xx, y)$' for "y is the set obtained by collecting xx into a whole". Cantor can be understood as claiming that the operation of set formation can be applied to any coexistent objects but that it is impossible to complete "the process" of applying the operation:

$$\Box\forall xx\Diamond\exists y\,\textsc{Set}(xx, y) \tag{4}$$

$$\neg\Diamond\forall xx\exists y\,\textsc{Set}(xx, y) \tag{5}$$

These two claims echo (1) and (2), respectively.

More generally, a condition $\varphi(x)$ is completable just in case it is possible for the generation of instances to be "finished", such that all the instances the condition could ever have are available:

$$\Diamond\exists xx\Box\forall y(y \prec xx \leftrightarrow \varphi(y))$$

The "multiplicity" associated with a condition is thus intrinsically suitable to form a set just in case the condition is completable.

Our proposal is that Cantor's use of temporal and modal language need not be dismissed as just colorful language but can be understood as conveying a valuable idea about set formation. Whenever a mathematical operation is iterated, the metaphor of a process is natural. But the operation of set formation has a special feature, namely, that it takes *many objects* as input and outputs their set. This raises the question of what kinds of input this operation might take. Cantor's attempt at an answer is that availability to serve as input to set formation is a matter of coexistence or being finished. Using the metaphor of a process, he thus identifies a new theoretical primitive of *joint availability* for set formation. This new primitive can be explicated using modality and plural logic; and thus explicated, its explanatory potential can be realized (cf. footnote 13).

Summing up, we have seen that Cantor began with a conception of set that emphasizes the intensional definiteness of every well-defined set but gradually shifted to a conception that adds the requirement that a set be extensionally definite. We have also argued that his mature notion of extensional definiteness can be understood as the potentialist one of all the elements of a set "coexisting", or being jointly available for set formation. While this is far from a worked-out account of legitimate set formation, it at least points the direction for such an account.

3 Zermelo

The notion of definiteness plays a key role in Zermelo's foundational work. It features prominently in his well-known axiomatisation of set theory of 1908 and motivates his subsequent foundational reflections in the 1920s and 1930s. A constant reference of Zermelo is Cantor's set theory. Zermelo repeatedly states that Cantor's "original definition" of set in (Cantor, 1895) had been long recognized as insufficient in view of the antinomies of set theory. [15] Throughout his work Zermelo puts forth a number of original proposals that improve Cantor's definition of set.[16] Here we focus on (Zermelo, 1908b) and some ideas presented in (Zermelo, 1929a) and (Zermelo, 1930c).

At the beginning of (Zermelo, 1908b) Zermelo writes:

> [...] the very existence of [set theory] seems to be threatened by certain contradictions, or 'antinomies', that can be derived from its principles [...] and to which no entirely satisfactory solution has yet been found. (Zermelo, 1908b, p. 189)

He claims that in view of "Russell's antinomy" it is no longer admissible "to assign to an arbitrary logically definable notion a 'set', or 'class', as its 'extension' ". This, Zermelo argues, shows that

> Cantor's original [viz. 1895] definition of a 'set' as 'a collection, gathered into a whole, of certain well-distinguished objects of our perception or of our thought' therefore certainly requires some restrictions [...]. (Zermelo, 1908b, p. 189-91)

To overcome these difficulties, Zermelo presents an axiom system intended to be sufficiently restricted to exclude all contradictions, but also sufficiently wide to retain all that is valuable in set theory. Zermelo's key restrictions are introduced in his *schema of separation* (Axiom III) and it is here that definiteness comes into play. Zermelo's separation allows us

[15] See e.g. (Zermelo, 1908b, p. 189-91) and (Zermelo, 1929b, p. 387). Moore (Moore, 1978) argues that Zermelo's axiomatisation was primarily motivated not by the paradoxes but by the desire to reply to widespread criticism of his 1904 well-ordering theorem.

[16] Zermelo's ideas are difficult to interpret as they lack a precise framework and language. We think, however, that Zermelo offers a significant and original perspective on definiteness and for this reason we review its key characteristics. For detailed analysis of Zermelo's thought on definiteness see, for example, (Ebbinghaus, 2003), (Ebbinghaus, 2010) and (Felgner, 2010).

to define a **subset of a given set**, say M, by separating out of M those of its elements that satisfy a "definite" property.

In (Zermelo, 1908b, p. 195) the separation schema reads as follows:[17]

> Whenever the propositional function $\mathfrak{E}(x)$ is definite for all elements of a set M, M possesses a subset $M_{\mathfrak{E}}$ containing as elements precisely those elements x of M for which $\mathfrak{E}(x)$ is true.

Separation is therefore more restrictive than naive comprehension in two respects:

- it defines a subset of *an already given set*, M, by separating out those elements of M that satisfy some property;
- it requires that the property used to define the subset of M is *definite.*

The first restriction is fundamental, as it aims at excluding "contradictory notions such as 'the set of all sets' or 'the set of all ordinal numbers' " (Zermelo, 1908b, p. 195). The second restriction is needed to avoid paradoxes such as Richard's paradox. The idea is that notions such as 'definable by means of a finite number of words' are not definite, so that separation cannot be applied to them. This second restriction is also highly significant, since Zermelo's axiomatisation is not expressed on the basis of a fixed formal language. The restriction is intended to rule out problematic notions such as the notion of definability that figures in Richard paradox. As Zermelo retrospectively observed, "a universally acknowledged 'mathematical logic' on which I could have relied did not exist" at the time (Zermelo, 1929b, p. 359). Therefore in 1908 he appeals to definitenss to excise problematic properties. However, lacking a rigorous description of the language of set theory, Zermelo's notion of definiteness was soon criticised for being too vague. Before looking at this criticism, we need to examine in more detail how Zermelo characterises definiteness and how his notion relates to our distinction between intensional and extensional definiteness.

[17] Zermelo's schema of separation figures also in (Zermelo, 1908a), where Zermelo employs the expression "well-defined property" ("wohldefinierte Eigenschaft"). More precisely in that text the separation axiom reads as follows: "All elements of a set M that have a property E well-defined for every single element are the elements of another set, M_E, a 'subset' of M" (Zermelo, 1908a, p. 121). It is clear that definiteness is here to be read distributively, as applying to each individual element of the set M. As argued in (Felgner, 2010, p. 180), at first Zermelo's terminology reminds us of Cantor's 1882 definition of set in terms of "wohldefinierte Mannigfaltigkeiten", which was quoted at page 3. Similar terminology appears in lecture notes by Zermelo from the winter semester 1900/1901 (with additions from the years 1904-1906). Felgner argues that when later Zermelo employed the term "definit", he borrowed it from Husserl. See also (Moore, 1982, pp. 155-6) for a discussion of earlier attempts by Zermelo to formulate separation.

Zermelo (Zermelo, 1908b, p. 191) starts by presupposing a given domain, B, and a system of *fundamental relations* over B. He then offers the following explanation of definiteness (Zermelo, 1908b, p. 193):

> A question or assertion $\mathfrak{E}$ is said to be "*definite*" if the fundamental relations of the domain, by means of the axioms and the universally valid laws of logic, determine without arbitrariness whether it holds or not. Likewise, a "propositional function" $\mathfrak{E}(x)$ in which the variable term x ranges over all individuals of a class $\mathfrak{K}$, is said to be "*definite*" if it is definite for *each single* individual x of the class $\mathfrak{K}$.

Zermelo's passage suggests that definiteness is determined by three components: (a) the fundamental relations of the domain, (b) the axioms of set theory and (c) the laws of logic. To this we need to add the claim (d) that such a determination ought to be non-arbitrary. But how is this meant to eliminate the threat of paradoxes such as Richard's? Presumably we can reason as follows. Suppose we start from the fundamental relations of set theory. These are of the form $a \in b$ and $a = b$, for a and b elements of the domain B. Given these fundamental relations we do not have a non-arbitrary way of determining, by employing only the laws of logic, whether a property such as 'definable by means of a finite number of words' holds of some number or not. Therefore we cannot apply separation in such cases.

Given this characterisation of definiteness, we may now wonder how it features within our distinction between intensional and extensional definiteness. After stating the axiom of separation, Zermelo writes that the criterion defining a subset of a set M is definite if "for each single element x of M the 'fundamental relations of the domain' must determine whether it holds or not" (Zermelo, 1908b, p. 195). This explanation makes it clear that Zermelo understands definiteness *distributively*, as a definite condition either holds or does not hold of *each individual element* of a given set. In this respect, definiteness looks like a form of *intensional definiteness.*

As we have seen, Zermelo presents his definition of set as an improvement of Cantor's original (1895) definition, which, we argued, can be plausibly read as requiring that a set be extensionally definite. As we find plausible to read Zermelo's definiteness as a form of intensional definiteness, one may wonder how this can result in an improvement with respect to Cantor's definition of set. We have two main observations in this respect. First, there is substantial improvement, of course, in Zermelo's *axiomatic* presentation of set theory, which affords a more rigorous account of sets. While Cantor may require that a set be extensionally definite, his definition does not furnish clear criteria for sethood, which is something Zermelo's axiomatization strives to do. It is true that Zermelo's separation schema runs into difficulties as it is not sufficiently precise, as we will see shortly. But Zermelo's approach made a rigorous formulation of separation urgent, prompting further improvements to the formulation of set theory.

Second, in the separation schema we find a distinctive interplay between intensional definiteness and the restriction of the defining condition to an already given set, M. It is plausible that Zermelo's separation schema is meant to fix the extension of (the property expressed by) a definite condition relative to the given set M, while the latter is thought of as extensionally definite in virtue of the axioms of set theory. The idea seems to be that as we single out a subset of an already given set M by means of a definite condition, we also fully determine the extension of that condition when its range of application is restricted to the elements of the set M.

This seems a reasonable reconstruction of how definiteness is supposed to restrict and improve on Cantor's definition of set. Let us return now to the complaint that Zermelo's notion of definiteness is insufficiently precise to achieve its purpose. Zermelo's characterisation of definiteness is imprecise, as it lacks the required detail that would allow us to decide in each case whether a property is definite or not. For example, Zermelo does not specify the language in which the axioms of set theory are formulated. In addition, an explication of definiteness that refers to the axioms of set theory is problematic, since definiteness is required to make sense of the separation axiom. Indeed, the lack of precision and the resulting vagueness of Zermelo's notion of 'definite property' was criticised by mathematicians such as Weyl, Fraenkel and Skolem (Weyl, 1910; Weyl, 1918; Fraenkel, 1922; Skolem, 1922). They proposed improvements leading to the *first-order* formulation of Zermelo-Fraenkel set theory we are familiar with. We will discuss Weyl's proposal in Section 4.[18]

The reflection on Zermelo's notion of definiteness that followed his axiomatisation of set theory helped shape set theory. The proposed solution to the question of how to define definite properties and eliminate paradoxes such as Richard's developed in a fully rigorous way ideas already implicit in Zermelo's informal explication of definiteness, in particular components (a)–(c). The idea was to make precise the language of set theory in terms of the (now) ordinary first-order language, with the equality and membership predicates as basic.

[18] See (Ebbinghaus, 2003; Ebbinghaus, 2010; Felgner, 2010; Taylor, 2002) for discussions of Zermelo's contemporaries' criticism of his notion of definiteness and for an analysis of Zermelo's different proposals on how to spell out definiteness.

The legitimate (definite) properties are those that are expressible by repeated application of the first-order logical operations to atomic statements of the form $a = b$ and $a \in b$. The thought was that this suffices to rule out problematic definitions such as that of a number definable by means of a finite number of words. As a consequence, the explicit reference to definiteness that characterised Zermelo's 1908's formulation of the separation schema disappeared. There was no further need to distinguish the definite properties from those that are not, as the language, once appropriately regimented, ensures that only definite properties are expressible.

It is natural to wonder, though, whether the resulting explication of definiteness in terms of definability in the first-order language of set theory is satisfactory. Both directions can be challenged: we will now look at Zermelo's challenge to the thought that every definite property is definabile in the first-order language of set theory. In the next section we will discuss Weyl's misgivings about the reverse inclusion.

In a paper from 1929 and in other texts from the 1930s, Zermelo explored new ways of clarifying the notion of definiteness, sketching new intriguing ideas.[19] Zermelo's new reflections on definiteness followed a number of distinct threads without reaching a definitive and clear solution to the difficulties that afflicted his original approach to definiteness. In fact, Ebbinghaus (Ebbinghaus, 2003) argues that Zermelo developed contrasting ideas, differing for the choice of language and the view of the nature of the set-theoretic universe, but did not succeed in bringing them together into a unified approach (Ebbinghaus, 2003, p. 198).

Of particular significance for our investigation is the text "On the concept of definiteness in axiomatics" (Zermelo, 1929a). There Zermelo replies to the criticism levelled against his earlier (Zermelo, 1908b) formulation of definiteness, stating that he had not been understood. He then scrutinizes his critics' proposed solutions. Among them is Fraenkel's approach to definiteness, that Zermelo terms "genetic". This may be considered a variant of the approach proposed by Weyl (1910) and Skolem (1922), which replaces definiteness by first-order definability in the language of set theory. One of Zermelo's complaints is that the genetic approach makes essential use of the natural numbers, in the form of the *finite* repetitions of the logical operations. It therefore presupposes the natural numbers instead of grounding them set-theoretically.[20] Another complaint is that a genetic characterization

[19] See, for example, (Zermelo, 1929a; Zermelo, 1930c; Zermelo, 1930a; Zermelo, 1930b).

[20] See page 16 (below) for a passage by Zermelo in which he considers, but then quickly dismisses, a genetic characterisation of definiteness. The reliance of a genetic characterisation of definiteness on the natural numbers is also an important element of Weyl's reflection.

of definiteness "contradicts [the] purpose and nature of the *axiomatic* method" (Zermelo, 1929a, p. 359). Zermelo proposes instead what he calls an "axiomatic" approach, that is, he gives a series of conditions that are to be satisfied by the collection of definite properties. In this new attempt to pin down the notion of definiteness, Zermelo simultaneously hints at two distinct ideas, one model-theoretic and one syntactic.[21]

The first is presented as follows. Zermelo introduces the notion of *logically closed system*, which is, essentially, the deductive closure of a set of axioms (Zermelo, 1929a, p. 361). He claims that if a logically closed system is consistent,

> "then it must be 'realizable' as well, that is, representable by means of a "model", by means of a *complete matrix* of the '*fundamental relations*' that occur in the axioms or in the system".

Now, each fundamental relation R is '*disjunctive*', in the sense that either R or its negation holds in the model. The same holds also for composite relations, so that "it is uniquely decided in every model by means of the matrix of the fundamental relations whether or not they hold in it." The definite properties are then those that are "decided by means of the fundamental relations in every model". Zermelo (Zermelo, 1929a, p. 361-63) also states that

> "[d]efinite" is thus what is *decided in every* single *model*, but may be decided differently in different models; "decidedness" refers to the individual *model*, whereas "definiteness" itself refers to the *relation* under consideration and to the entire *system.*

This enables Zermelo to claim that non-definite properties are those that are either not uniquely determined by the fundamental relations or "alien to the system". For example, the property "not definable by means of a finite number of words in any European language" is not uniquely determined by the fundamental relations, while the property of being a set "painted in green" is alien to the system. To sum up, the key idea of this model-theoretic characterisation is that definiteness is what is decided in every model of the axioms of set theory on the basis of the fundamental relations of the domain.

It is tempting to see this as a novel development of components (a) - (d) of the 1908 characterisation of definiteness, now focusing on the models of the axioms of set theory. Zermelo briefly mentions also categorical systems of axioms, but does not explain the relation

[21] It should be noted that while Zermelo clearly makes use of second-order quantifiers in his axiomatic characterisation, he does not fully specify the underlying language.

between his notion of definiteness and categorical systems. Categoricity, indeed, becomes a crucial theme in his foundational investigations in the 1930s, when he tries to develop a notion of set as categorical domain, that is, as a domain that can be characterized by a categorical system of axioms. Perhaps one way to round out Zermelo's 1929 proposal is to suggest that a categorical system of axioms would ensure that definite properties are uniquely determined by the fundamental relations in every model *in the same way.* This would ensure *non-arbitrariness*, that is, the satisfaction of requirement (d) that figured in the 1908 characterisation, without been fully accounted for. We will return to Zermelo's model-theoretic considerations shortly, when we discuss (Zermelo, 1930c).

The second characterisation of definiteness Zermelo sketches in (Zermelo, 1929a) is syntactical. Its purpose is to further clarify which properties are definite. Zermelo asks:

> But which propositions and properties are now in fact 'definite'? How can we decide whether a given proposition is 'definite'?

To answer these questions Zermelo first considers a genetic characterisation of definiteness (Zermelo, 1929a, p. 363; his italics):

> *A proposition is called "definite" for a given system if it is constructed from the fundamental relations of the system only by virtue of the logical elementary operations of negation, conjunction and disjunction, as well as quantification, all these operations in arbitrary yet finite repetition and composition.*

He finds a genetic characterisation wanting for the reasons already mentioned, in particular, for its relying on the natural number concept. He therefore presents his "axiomatic approach". As in (Zermelo, 1908b), Zermelo starts by presupposing a given a domain, B, and a system of *fundamental relations* over B. He then gives the following clauses (Zermelo, 1929a, p. 365):

I) all fundamental relations are definite;

II) 1) the logical operations of negation preserves definiteness;

2) conjunction and disjunction preserve definiteness;

3) the first-order quantifiers preserve definiteness;

4) the second-order quantifiers *applied to "definite functors"* preserve definiteness. More precisely, if $F(f)$ is definite for all *definite* "functors" f, then so is the quantified statement $\forall f F(f)$.

Zermelo supplements these clauses with a closure condition, whose aim is to rule out non-definite properties without making use of the natural number concept. Calling P the totality of all definite properties, he states that no subtotality of P satisfies all the postulates I) and II) (Zermelo, 1929a, p. 364).[22]

The fourth clause is particularly important, as it extends the notion of definiteness to the second-order (although it introduces also a puzzling restriction, by requiring that the arguments of the second-order quantifiers be definite). It is here that Zermelo clearly goes beyond the first-order characterisation of definability and considers definite also statements involving the second order quantifiers. He will further explore the fundamental shift to the second-order language (unrestricted) in subsequent work, especially (Zermelo, 1930c).

We have seen that Zermelo sketches both a model-theoretic and a syntactic explication of definiteness. In both cases the fundamental relations play a key role in determining what is and what is not definite. This corresponds to component (a) of definiteness in (Zermelo, 1908b). Since the fundamental relations are presumably determined by the axioms of the theory, condition (b) also plays a role. Furthermore, in both explications the laws of logic, namely component (c), are key to the preservation of definiteness from the fundamental relations to complex ones . Although Zermelo's focus on the decidability of definite properties makes these new explications of definiteness still close to intensional definiteness, there are also new elements to them that point in a different direction. In the model-theoretic case, Zermelo takes a "global" perspective, by requiring intensional definiteness in every model while in the syntactic case he stresses its preservation with respect to the quantifiers.[23] The model-theoretic characterisation gestured at in 1929 was further explored by Zermelo in the 1930s, reaching a more refined form in (Zermelo, 1930c). There, we argue, we find clear elements of extensional definiteness.

In (Zermelo, 1930c) Zermelo presents a second-order set theory.[24] Definiteness is somehow built in and does not figure any more in the separation axiom, similarly to the standard first-order approach to separation in ZFC. Namely, it is the *second-order* language of set theory, alone, that suffices to formulate separation. Zermelo now states that the propositional function used in the separation schema to separate a subset of a given set should be

[22] This new definition was criticised by Skolem (Skolem, 1930), who offered two main criticisms: it requires set theory (to express the closure condition) and it uses a vague notion of second order quantification.

[23] The syntactic characterisation has important similarities, as well as differences, with Weyl's notion of extensional definiteness that is the focus of the next section.

[24] Zermelo's variant of ZF includes second-order formulations of separation and replacement. Zermelo omits infinity, presupposes the axiom of choice, which he considers a "general logical principle" (Zermelo, 1930c, p. 405), and introduces the axiom of foundation. He also assumes urelements.

completely arbitrary.[25]

In this paper Zermelo is concerned with *normal domains*, which are models of his set theory indexed by inaccessible cardinals. Zermelo crucially presupposes the availability of an unbounded series of (strongly) inaccessible cardinals, which ensures that the totality of normal domains is itself open-ended. As each normal domain is followed by a "next" normal domain, it is to be thought of as *closed* from the point of view of the next normal domain. The interaction between normal domains, which are thought of as closed, and their open-ended hierarchy is taken by Zermelo to offer a satisfactory solution to the antinomies of set theory. The hierarchy of normal domains also brings Zermelo to realise that his set theory is non-categorical, although its models satisfy important isomorphism theorems. In (Zermelo, 1930a, p. 437), Zermelo summarises his findings as follows:

> Two normal domains are "isomorphic" if and only if 1) their "bases" (that is, the totality of their urelements) are equivalent to one another and 2) their "characteristics" (that is the upper limits of the occurring alephs) are equal, if, in other words, to each set of one domain there corresponds at least one equivalent one in the other domain. Of two domains with equivalent bases (but different characteristics) one is always isomorphic to a "canonical" development of the other.

A little later, Zermelo mentions a connection between his notion of normal domain and Cantor's "concept of set". He introduces the notion of "closed domain", which, he claims, corresponds to that of set in Cantor's sense. Zermelo claims that a "closed domain" can be reduced to the concept of "categorical system of postulates" and that every normal domain is a closed domain and can therefore be conceived of as a set in a higher normal domain. A more precise characterisation of closed domains is offered in (Zermelo, 1930b, p. 453) as follows:

> A "closed domain" is one which can be determined or ordered by means of a *categorical system of postulates.* It is precisely that which Cantor really meant by his well-known definition of "set".

A closed domain is contrasted by Zermelo with an open domain (Zermelo, 1930b, p. 453):

> An "open domain" is a *well-ordered sequence of domains successively comprising one another* constituted so that *every closed subdomain can always still be*

[25] See (Zermelo, 1930c, p. 403, fn. 2), as well as (Zermelo, 1930b, p. 449).

> *extended in it. [...] Furthermore, the entire open domain can be well-ordered so that all elements of a preceding layer precede all elements of every subsequent one.*

Closed domains are clearly extensionally definite: we are interested in all the elements of the domain, rather than considering them one by one. We may also say that closed domains are complete or "finished".

To summarize, in the new phase of Zermelo's reflection on definiteness beginning with (Zermelo, 1929a), second-order logic and categoricity play a key role. Zermelo sketches both syntactic and model-theoretic presentations of definiteness. The syntactic presentation takes the form of preservation of definiteness through the logical operations, including the second-order quantifiers. The model-theoretic characterisation is given in terms of what is decided within (categorical) models of the axioms of set theory. This subsequently gives rise to a notion of closed domain, which explicates (what we would call) a form of extensional definiteness. It is tempting to think that this notion of definiteness is close to the Cantorian notion of completability. Zermelo, at any rate, seems to have thought that through a crucial use of the axioms of set theory, the notion of closed domain offered a more precise rendering of Cantor's ideas.[26]

4 Weyl

Let us return to the problem of clarifying the notion of a definite property that figures in Zermelo's original 1908 statement of the axiom of Separation. The usual explication—as a property definable in the language of first-order set theory—can be challenged. As we have seen, Zermelo contested one direction, arguing that properties definable in the language of *second-order* set theory can also be definite. We will now see that Weyl comes to reject the reverse direction, that is, to deny that every formula of first-order set theory defines a definite property.

In his Habilitation Vortrag from 1910 Weyl discusses Zermelo's separation schema, writing that:

> According to Zermelo, a definite expression is one whose application or non-application can be determined unequivocally and without arbitrariness on the

[26] See, e.g., (Ebbinghaus, 2003; Taylor, 2002) for detailed analysis of this phase of Zermelo's foundational reflection. See also (Taylor, 2002) for discussion of another proposal by Zermelo, involving infinitary languages.

> basis of the fundamental relation ϵ that holds between the objects of set theory. Here, to my mind, greater precision is necessary because the expression 'determined unequivocally and without arbitrariness' strikes me as too vague. (Weyl, 1910, p. 304; our translation)

He proposes to characterize a definite relation as one obtained from the fundamental relations of set theory, namely membership and equality, by finitely many applications of five principles of definition that he sets forth. These principles have a clear algebraic character but are usually taken to capture the relations that are definable in (what we call) the language of first-order set theory.

Weyl returns to this issue in his pioneering work on predicativity, *Das Kontinuum* (Weyl, 1918), where he mentions that his "investigation began with an examination of Zermelo's axioms for set theory". He now sets out in full detail the principles of definitions in terms of the ordinary logical operations, but crucially he focuses on the case where these operations are applied to the domain of natural numbers. As we will see, the most important change is that Weyl now rejects the idea that quantification over all sets can be assumed to result in a formula that is (intensionally) definite. This rejection springs from a stronger preference for some generative approach to mathematical objects. Indeed, Weyl espouses a view that is closer to Aristotle's austere potentialism than to Cantor's extremely relaxed set-theoretic potentialism. For whereas Cantor accepts a plethora of transfinite sets as completable, Weyl regards every infinity—even that of the natural numbers—as incompletable; for "'inexhaustibility' is essential to the infinite" (Weyl, 1918, p. 23). Weyl's potentialist outlook becomes particularly clear in a later passage, commenting on how to place mathematics on a sound foundation:

> The deepest root of the trouble lies elsewhere: a field of possibilities open into infinity has been mistaken for a closed realm of things existing in themselves. As Brouwer pointed out, this is a fallacy, the Fall and Original Sin of set-theory, even if no paradoxes result from it. (Weyl, 1949, p. 234)

So, for Weyl, every infinite mathematical domain is incompletable.

Weyl proceeds to make an important and highly innovative distinction. Not all infinite (and therefore incompletable) domains have the same character: some are "extensionally determinate", whereas others are not (Crosilla and Linnebo, 2024).[27] Let us follow Weyl

[27] As a terminological convention, we reserve the term 'extensionally determinate' and its cognates for Weyl's particular explication of the broader idea of extensional definiteness.

and begin with a loose and intuitive way of drawing the distinction, which we will gradually make more precise. A concept's being "clearly and unambiguously defined", Weyl contends,

> does not imply that this concept is extensionally determinate, i.e., that it is meaningful to speak of the *existent* objects falling under it as an ideally closed aggregate which is intrinsically determined and demarcated. (Weyl, 1919, p. 109)

Thus, for an infinite domain to be extensionally determinate is, loosely speaking, for it to be properly demarcated.

Weyl's mathematical genius is revealed when he proceeds to give this loose and intuitive idea a precise logical articulation. His first step in this direction reads as follows.

> Suppose P is a property pertinent to the objects falling under a concept C. [...] if the concept C is extensionally determinate, then not only the question "Does a have the property P?" [...] but also the existential question "*Is there* an object falling under C which has the property P?", possesses a sense which is intrinsically clear. (ibid.)

Let us paraphrase. Suppose some property P is well-defined on all objects falling under a concept C. That is, suppose that for every individual instance a of the concept C, the question whether a has P has an "intrinsically clear sense". (In our terminology, we suppose that P is intensionally definite on instances of C.) What is it, then, for the concept C to be extensionally determinate or "properly demarcated"? Weyl proposes that this demarcation would enable us to quantify over all Cs and ask—again with an intrinsically clear sense—whether there is some C that has P.[28] This suggests the following analysis:

> A concept C is *extensionally determinate* just in case: for every property P such that the question 'Pa?' has an intrinsically clear sense whenever a is C, also the quantificational question '$(\exists x : Cx)Px$?' has an intrinsic clear sense.

In short, for a concept C to be extensionally determinate is for quantification over Cs to preserve the property of possessing an intrinsically clear sense.

The next step is to clarify what is it for a statement to "possess a sense which is intrinsically clear". Weyl writes that a question has an intrinsically clear sense when it

[28] In a potentialist setting where objects are successively defined and the principles of definition may themselves be an "open system" to which we can make additions, possession of an "intrinsically clear sense" cannot be taken for granted. See (Weyl, 1918, p. 87), as well as (Linnebo and Shapiro, 2023, pp. 2–4).

"address[es] an existing state of affairs that allows one to answer the question with yes or no" (Weyl, 1921, p. 88). That is, a statement has an intrinsically clear sense when it is subject to the principle of bivalence, which says that the statement is either true or false.

A final step is now very natural—although it is not, as far as we know, explicitly taken by Weyl himself. The idea is to explicate bivalence in the metalanguage with the Law of Excluded Middle holding in the object language. When this step is taken, we arrive at a formal analysis in the object language itself:

> **Extensional determinateness (formal analysis)**
>
> A concept C is *extensionally determinate* iff quantification restricted to C preserves the property of LEM holding, that is, iff, for every property P:

$$(\forall x : Cx)(Px \vee \neg Px) \to (\exists x : Cx)Px \vee \neg(\exists x : Cx)Px$$

We have thus arrived at an idea that Feferman has recently expressed with pleasing succinctness: "What's definite is the domain of classical logic, what's not is that of intuitionistic logic" (Feferman, 2011, 23). To highlight the central role of quantification in this analysis, we prefer to say that a concept that is extensionally determinate in this sense *defines a domain of classical quantification.*

Equipped with this distinction between two kinds of incompletable domains—those that are extensionally determinate and those that are not—we can ask which domains fall on which side. The following passage summarizes Weyl's own view:

> The intuition of iteration assures us that the concept "natural number" is extensionally determinate. [...] However, the universal concept "object" is not extensionally determinate—nor is the concept "property," nor even just "property of natural number". (Weyl, 1919, p. 110)

We see that Weyl is very sparing in what he regards as extensionally determinate. The domain of natural numbers is extensionally determinate, thanks to our "intuition of iteration".[29] But Weyl is unwilling to go much further. Extensional determinateness is lost as soon as we consider properties—or, for that matter, sets—of natural numbers. We will shortly explain why Weyl held this very strict view.

[29] (Hartimo, 2023) suggests that Husserl's notion of "material definiteness" is rather like Weyl's "extensional determinateness", and that the former's "formal definiteness" is like Zermelo's "categorically determined". The idea is that what ensures that a domain is extensionally determinate, for Weyl, is that the domain is "materially" generated from below in some iterative procedure.

First, though, we wish to observe that the view explains why Weyl came to reject the idea that every formula of (what we would now call) first-order set theory is intensionally definite. To begin, since the concept of natural number is extensionally determinate, there is a set of all natural numbers. And this set obviously has all kinds of subsets. A formula of first-order set theory aspires to quantify over all these subsets—and many other sets as well. But according to Weyl, already the collection of sets of natural numbers fails to be extensionally determinate. Thus, a formula of first-order set theory may quantify over a domain that fails to be extensionally determinate. It follows that such a formula may lack an intrinsically clear sense, or, in our terminology, that it may fail to be intensionally definite. Weyl therefore has to reject the standard explication of Zermelo's 1908 notion of a 'definite property' as being too lax. For this explication allows instances of Separation involving conditions that cannot be guaranteed to be intensionally definite.

Why, then, did Weyl reject the prevailing view that there is a properly demarcated domain of sets of natural numbers and perhaps larger domains still? The answer is that Weyl rejects the combinatorial conception of set as applied to infinite domains:

> The notion of an infinite set as a "gathering" brought together by infinitely many individual arbitrary acts of selection, assembled and surveyed as a whole by consciousness, is nonsensical: "inexhaustibility" is essential to the infinite. (Weyl, 1918, p. 23)

Without the combinatorial notion of an arbitrary subset of the natural numbers, we are left without a reason to take the domain of all such sets to be extensionally determinate.

Weyl is not content with these negative claims. Starting in 1918, he develops a positive alternative, namely, a novel *predicative* conception of set, which can legitimately be applied to the domain of natural numbers. Since an infinite set is incompletable (or "inexhaustible"), it needs to be described by means of a rule that "indicates properties which apply to the elements of the set and to no other objects". (Weyl, 1918, p. 20) And these rules need to be carefully specified in a bottom-up manner that avoids any circularity. In essence, when formulating a rule that determines membership in an infinite set, it is permissible to quantify over the natural numbers, since these are independently given; but it is not permissible to quantify over sets of numbers, since these are the very objects we are trying to characterize.

Suppose we follow Weyl and allow only predicative subsets of $\mathbb{N}$. Then there is a strong argument that the collection of sets of naturals is *not* properly demarcated or extensionally determinate. For, if this collection were extensionally determinate, we could use quantification over the collection to define yet further sets of numbers. It follows, therefore, that there is no extensionally determinate collection of absolutely all sets of natural numbers.

Let us examine what this view yields when we explicate extensional determinateness in the way outlined above, that is, using intuitionistic logic. Every atomic predication of natural numbers is *decidable*. For example, we have:

$$x + y = z \vee \neg x + y = z$$

Likewise, since the domain of natural numbers is assumed to be extensionally determinate, quantification over this domain behaves classically. We capture that by means of the so-called principle of *Bounded Omniscience (BOM)*:

$$(\forall x : \mathbb{N}(x))(\varphi(x) \vee \neg\varphi(x)) \to (\exists x : \mathbb{N}(x))\varphi(x) \vee \neg(\exists x : \mathbb{N}(x))\varphi(x)$$

An easy induction on syntactic complexity now shows that LEM holds for every formula of first-order arithmetic. By contrast, we do *not* have Omniscience for sets of naturals. That is, the quantifier '$\forall X \subseteq \mathbb{N}$' behaves intuitionistically, not classically.

5 Three notions of definiteness

Let us take stock. We hope to have made it clear that definiteness isn't just a "philosophers' notion", detached from "real" mathematics. Some leading mathematicians during the first half century of modern set theory thought long and hard about definiteness. They typically started with some intuitive but rather loose ideas. That is, they started in what one may regard as philosophical territory. Through a mixture of philosophical and mathematical analysis, these intuitive ideas were then gradually transformed to precise logical notions.

Moreover, it has transpired that several different notions of definiteness are in play. There are two main families. First, a condition can be *intensionally definite*, in the sense of being well-defined or sharply defined in an instance-by-instance manner. This idea figures in early Cantor as well as in (Zermelo, 1908b) and the ensuing debate. The idea is naturally explicated in terms of bivalence (as in (Weyl, 1921, p. 88))—and thus also, in the context of intuitionistic logic, in terms of LEM (as we have suggested). Then, there is *extensional definiteness*, which is a collective property of a collection of objects, not a distributive one.

We have seen that this notion admits of two distinct explications. One is in terms of the potentialist idea of possible coexistence or completability, which we encountered in Cantor (especially in his letters to Hilbert and Dedekind) but also in Zermelo's later work, such as (Zermelo, 1930c). Weyl proposes a different explication. He starts with the intuitive idea of a domain being properly demarcated, which he explicates as a matter of the domain supporting classical quantification.

The following table summarizes our findings, with intuitive philosophical ideas in italics and sharp logico-mathematical explications in boldface.

ID		**ED**	
sharp, well-defined			*completable*
Cantor			Cantor
↓			↓
Zermelo '08			Zermelo '30
↓			
Weyl	→	*properly demarcated* Weyl	↓
↓		↓	
decidable		**classical quantification**	a plurality

To recall, a concept C is said to be *decidable* just in case $\forall x(Cx \vee \neg Cx)$. Further, C is said to define *a domain of classical quantification* just in case C is decidable and for every property P:

$$(\forall x : Cx)(Px \vee \neg Px) \to (\exists x : Cx)Px \vee \neg(\exists x : Cx)Px$$

It remains only to clarify the precise logico-mathematical cash value of the domain forming a plurality.

Before we do that, though, we would like to make two initial observations about the relation between the two forms of extensional definiteness. First, Weyl's notion of extensional determinateness can be less demanding than Cantor's neo-Aristotelian notion of completability. We have seen that Weyl regards some domains, such as that of the natural numbers, as incompletable yet still extensionally determinate. While such domains have a well-defined extension, they simultaneously have an irreducibly intensional character. The incompletability means that the domain cannot be specified as a plurality of objects but instead requires an intensional specification. Second, Weyl's notion of extensional determinateness is more closely connected with the notion of intensional definiteness than Cantor's neo-Aristotelian notion of completability. For a domain to be extensionally determinate, in Weyl's sense, just is for quantification over the domain to preserve intensional definiteness. This explains the single left-to-right arrow in the above table.

We turn now to the task of providing a sharper description of the logical-mathematical cash value of having a plurality of objects as opposed to "just" an extensionally determinate domain. Part of the answer is found in [an already quoted] passage from Weyl:

> The notion of an infinite set as a "gathering" brought together by infinitely many individual arbitrary acts of selection, assembled and surveyed as a whole by consciousness, is nonsensical: "inexhaustibility" is essential to the infinite. (Weyl, 1918, 23)

Weyl claims that when a domain is incompletable (or "inexhaustible"), the combinatorial conception of an arbitrary subset of the domain is unavailable. He also implies that this is *the reason* why the combinatorial conception is unavailable, which suggests that the availability of the combinatorial conception goes hand in hand with the completability of the domain. We believe this is correct—even if one follows Cantor, as against Weyl (and Aristotle) and regards some infinite domains as completable. The completability of a domain, whether finite or infinite, licences talk about arbitrary subsets of the domain.

We would like to extend this analysis yet further. To do so, we find it useful to recall (Bernays, 1935)'s rightly famous notion of "quasi-combinatorial" reasoning. The idea is introduced in the following passage:

> But analysis is not content with this modest variety of platonism [i.e. the use of classical logic]; it reflects it to a stronger degree with respect to the following notions: set of numbers, sequence of numbers, and function. It abstracts from the possibility of giving definitions of sets, sequences, and functions. These notions are used in a "quasi-combinatorial" sense, by which I mean: in the sense of an analogy of the infinite to the finite. (p. 259)

To reason quasi-combinatorially is thus to treat an infinite domain as if it were finite. This "analogy of the infinite to the finite" has several aspects. First, quantification over the domain can be understood as infinite conjunctions or disjunctions of instances. Second, we argue as if we can form arbitrary sets of objects from the domain. Thus, in particular, impredicative comprehension is permissible for sets of objects from the domain. Relatedly, the Axiom of Choice will be acceptable. For any family of non-empty and non-overlapping sets from the domain, there is a choice set containing precisely one member of each of the mentioned sets.

We can now state our proposal. *The logical-mathematical cash value of a domain being completable, or specifiable as a plurality, is that quasi-combinatorial reasoning about the domain is licensed.*[30] Plugging this analysis into the above table and compressing the chronology, we obtain the following truncated table:

ID	**ED**	
sharp, well-defined	*properly demarcated*	*completable*
Cantor, Zermelo, Weyl	Weyl	Cantor, Zermelo
decidable	**classical quantification**	**quasi-combinatorial**

We observed above that a concept can define a domain of classical quantification without defining a quasi-combinatorial domain. We now contend that there is an inclusion in the reverse direction; that is, every quasi-combinatorial domain is also a domain of classical quantification. When a domain can be completed as a plurality of objects, say *dd*, then for any generalization over the domain, there is a well-defined plurality of instances concerned with objects from the domain. This enables us to understand quantification over the domain as a (perhaps infinite) conjunction or disjunction of instances. And this, in turn, ensures that the domain can be taken to support classical quantification. Thus, the Cantorian notion of extensional definiteness entails Weyl's. It follows that Cantor's notion is strictly more demanding than Weyl's.

6 A classification of views in the foundations of mathematics

We wish to close by using the two forms of extensional definiteness—namely, defining a domain of classical quantification or even a quasi-combinatorial domain—to describe a way to classify various views in the foundations of mathematics. A key question to ask of any view in the foundations of mathematics, we propose, is how large a domain its proponents are willing to regard as extensionally definite in either of these two senses.

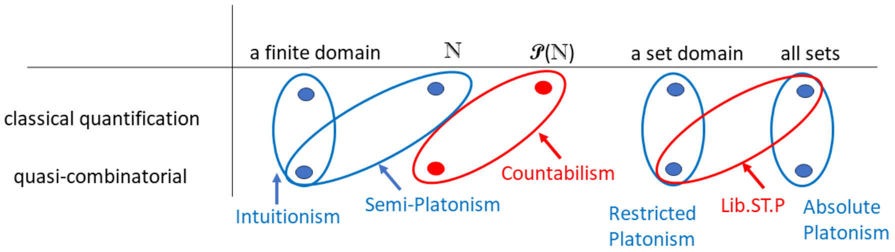
[30] See (Florio and Linnebo, 2021, Chs. 10 and 12) for an elaboration and defense of this proposal. As the authors acknowledge (pp. 288-89), this explication means that plural logic has rich and strong mathematical content. Indeed, as we have seen, Weyl (1918, p. 23) finds the idea of an infinite plurality, in this sense, "nonsensical".

The most important types of domain to consider are: finite domains, the domain of all natural numbers, the domain of all sets of natural numbers, any domain based on a Cantorian transfinite set, and the domain of all Cantorian sets. Thus, there are five types of domain, each of which can be classified as having either a weak or a strong form of extensional definiteness or neither of the two. We propose to measure the theoretical commitments of a view in the foundations of mathematics by asking how strong a form of extensional definiteness the view is willing to ascribe to how large a domain. Bernays (1935) famously regards these theoretical commitments as steps towards platonism in the foundations of mathematics.

To illustrate how our proposed classification works, let us apply it to four important positions in the foundations of mathematics identified by Bernays (1935), whose associated theoretical commitments he orders as follows:[31]

intuitionism < "semi-platonism" < "restricted platonism" < "absolute platonism"

First, intuitionism accepts only finite domains as extensionally definite in either respect. Next, "semi-platonism" regards the collection of natural numbers as a domain of classical quantification but not as a quasi-combinatorial domain (pp. 263 and 268). This is predicativism in the sense of (Weyl, 1918). Then, "restricted platonism", which is Bernays' own preferred view, regards any Cantorian transfinite set as completable (and *a fortiori* also a classical domain) (p. 261). Lastly, there is "absolute platonism", which regards the entire universe of sets as completable (pp. 261, 267, and 269). According to Bernays, absolute platonism is shown incoherent by the set-theoretic paradoxes, while "restricted platonism is not touched at all by the antinomies" (p. 261; see also p. 269). We thus obtain the following classification:



The classification can be extended beyond Bernays' four positions as well. Let us briefly mention two examples. First, we believe there is a coherent and interesting position that

[31] Although (Bernays, 1935) does not use the word 'definite' or any cognate thereof, the four positions he describes are concerned with which domains have the two forms of extensional definiteness that we have identified.

regards the countable infinity of natural numbers as completable, and the collection of all sets of natural numbers as a classical domain, but that refuses to go further with either notion of definiteness. This position is a form of *countabilism*.[32] Second, it is interesting to ask whether there is a coherent position that regards the collection of all transfinite sets as a domain of classical quantification but not as a quasi-combinatorial domain. In the literature on potentialism, this view has had several defenders, sometimes under the name of "liberal set-theoretic potentialism".[33] Others, though, have questioned the coherence of the view, arguing that in the end it fares no better than Bernays' "absolute platonism".

We submit that the two notions of extensional definiteness that emerges from our investigation enable us to identify and understand some of the most important fault lines in the philosophy and foundations of mathematics. Thus, far from being detached philosophy, these notions are mathematically illuminating.[34]

References

Benacerraf, P. and Putnam, H., editors (1983). *Philosophy of Mathematics: Selected Readings*, Cambridge. Cambridge University Press. Second edition.

Bernays, P. (1935). On platonism in mathematics. Reprinted in (Benacerraf and Putnam, 1983).

Boolos, G. (1984). To be is to be a value of a variable (or to be some values of some variables). *Journal of Philosophy*, 81(8):430–449.

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

Builes, D. and Wilson, J. M. (2022). In defense of countabilism. *Philosophical Studies*, 179(7):2199–2236.

Burgess, J. P. (2004). *E Pluribus Unum*: Plural logic and set theory. *Philosophia Mathematica*, 12(3):193–221.

Cantor, G. (1882). Über unendliche, lineare Punktmannigfaltigkeiten, 3. *Mathematische Annalen*, 20:113–121.

Cantor, G. (1883). *Grundlagen einer allgemeinen Mannigfaltigkeitslehre.* B.G. Teubner, Leipzig. Repr. and trans. in (Ewald, 1996).

Cantor, G. (1885). Review of G. Frege, *Grundlagen der Arithmetik*. *Deutscher Literaturzeitung*, 6:728–729.

Cantor, G. (1895). Beiträge zur Begründung der transfiniten Mengenlehre I. *Mathematische Annalen*, 46:481–512.

Crosilla, L. and Linnebo, Ø. (2024). Weyl and two kinds of potential domains. *Noûs*, 58(2):409–430.

Dummett, M. (1963). The philosophical significance of Gödel's theorem. In *Truth and Other Enigmas (1978)*, pages 186–214. Duckworth.

[32] See, e.g. (Scambler, 2021), (Builes and Wilson, 2022), and (Rathjen, 2016).

[33] See, e.g. (Linnebo, 2013), (Studd, 2013), and (Linnebo and Shapiro, 2019).

[34] Thanks to two anonymous referees and audiences in Oslo, Princeton, Paris, Milan and Prague for comments and discussion. Both authors gratefully acknowledge support from the Research Council of Norway (grant no. 314435).

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

Dummett, M. (1993). What is mathematics about? In *The Seas of Language*, pages 429–445. Oxford University Press.

Dummett, M. (1994). Chairman's Address: Basic Law V. *Proceedings of the Aristotelian Society*, 94:243–51.

Ebbinghaus, H.-D. (2003). Zermelo: definiteness and the universe of definable sets. *History and Philosophy of Logic*, 24(3):197–219.

Ebbinghaus, H.-D. (2010). Introductory note to 1908b. In Ebbinghaus, H.-D., Fraser, C. G., and Kanamori, A., editors, *Ernest Zermelo: Collected Works. Gesammelte Werke. Vol I*, pages 29–47. Springer-Verlag, Berlin.

Ebert, P. A. and Rossberg, M. (2009). Cantor on Frege's Foundations of Arithmetic: Cantor's 1885 Review of Frege's Die Grundlagen der Arithmetik. *History and Philosophy of Logic*, 30(4):341–348.

Ewald, W. (1996). *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, volume 2. Oxford University Press, Oxford.

Feferman, S. (2011). Is the continuum hypothesis a definite mathematical problem? Unpublished manuscript.

Felgner, U. (2010). Introductory note to 1908b. In Ebbinghaus, H.-D., Fraser, C. G., and Kanamori, A., editors, *Ernest Zermelo: Collected Works. Gesammelte Werke. Vol I*, pages 29–47. Springer-Verlag, Berlin.

Ferreirós, J. (1999). *Labyrinth of Thought: A History of Set Theory and Its Role in Modern Mathematics.* Birkhäuser, Basel.

Ferreirós, J. (2023). The role of syllogistic logic in early set theory. In Verburgt, L. M. and Cosci, M., editors, *Aristotle's Syllogism and the Creation of Modern Logic*, chapter 13, pages 247–68. Bloomsbury Academic, London.

Florio, S. and Linnebo, Ø. (2021). *The Many and the One: A Philosophical Study of Plural Logic.* Oxford University Press, Oxford.

Fraenkel, A. (1922). Über den Begriff 'definit' und die Unabhängigkeit des Auswahlsaxioms. *Sitzungsberichte der Preussischen Akademie der Wissenschaften, Physik-math. Klasse*, pages 253–257. Translated in (van Heijenoort, 1967, pp. 284–289).

Frege, G. (1893). *Grundgesetze der Arithmetik I.* Georg Olms Verlag, Hildesheim.

Frege, G. (1953). *Foundations of Arithmetic.* Blackwell, Oxford. Transl. by J.L. Austin.

Hartimo, M. (2023). Husserlian Features in Weyl's Das Kontinuum – Husserl and Weyl as Philosophers of Mathematical Practice. Unpublished Manuscript.

Linnebo, Ø. (2013). The potential hierarchy of sets. *Review of Symbolic Logic*, 6(2):205–228.

Linnebo, Ø. (2018). Dummett on indefinite extensibility. *Philosophical Issues*, 28(1):196–220.

Linnebo, Ø. and Shapiro, S. (2019). Actual and potential infinity. *Noûs*, 53(1):160–191.

Linnebo, Ø. and Shapiro, S. (2023). Predicativism as a form of potentialism. *Review of Symbolic Logic*, 16(1):1–32.

Mancosu, P. (1998). *From Brouwer to Hilbert: The Debate on the Foundations of Mathematics in the 1920s.* Oxford University Press.

Moore, G. H. (1978). The origins of Zermelo's axiomatization of set theory. *J Philos Logic*, 7:307–329.

Moore, G. H. (1982). *Zermelo's axiom of choice: Its Origins, Development, and Influence*, volume 8 of *Studies in the History of Mathematics and Physical Sciences.* Springer-Verlag, New York.

Moore, G. H. and Garciadiego, A. (1981). Burali-Forti's Paradox: A Reappraisal of its Origins. *Historia Mathematica*, 8:319–50.

Parsons, C. (1977). What is the iterative conception of set? In Butts, R. and Hintikka, J., editors, *Logic, Foundations of Mathematics, and Computability Theory*, pages 335–367. Reidel, Dordrecht. Reprinted in (Benacerraf and Putnam, 1983) and (Parsons, 1983a).

Parsons, C. (1983a). *Mathematics in Philosophy.* Cornell University Press, Ithaca, NY.

Parsons, C. (1983b). Sets and modality. In *Mathematics in Philosophy*, pages 298–341. Cornell University Press, Ithaca, NY.

Parsons, C. (2012). Some remarks on Frege's conception of extension. In *From Kant to Husserl: Selected Essays.* Harvard University Press, Cambridge, MA.

Rathjen, M. (2016). Indefiniteness in Semi-Intuitionistic Set Theories: On a Conjecture of Feferman. *Journal of Symbolic Logic*, 81(2):742–754.

Scambler, C. (2021). Can all things be counted? *Journal of Philosophical Logic*, 50(5):1079–1106.

Shapiro, S. and Wright, C. (2006). All things indefinitely extensible. In Rayo, A. and Uzquiano, G., editors, *Absolute Generality*, pages 255–304. Oxford University Press, Oxford.

Skolem, T. (1922). Some Remarks on Axiomatized Set Theory. In (van Heijenoort, 1967).

Skolem, T. (1930). Einige Bemerkungen zu der Abhandlung von E. Zermelo: "Über die Definitheit in der Axiomatik". *Fundamenta Mathematicae*, 15:337–341.

Studd, J. (2013). The iterative conception of set: A (bi-)modal axiomatisation. *Journal of Philosophical Logic*, 42(5):697–725.

Tait, W. (2009). Cantor's Grundlagen and the Paradoxes of Set Theory. In Sher, G. and Tieszen, R., editors, *Between Logic and Intuition: Essays in Honor of Charles Parsons*, pages 269–290. Cambridge Univerity Press, Cambridge.

Taylor, R. G. (2002). Zermelo's Cantorian Theory of Systems of Infinitely Long Propositions. *The Bulletin of Symbolic Logic*, 8(4):478–515.

van Heijenoort, J., editor (1967). *From Frege to Gödel*, Cambridge, MA. Harvard University Press.

Weyl, H. (1910). Über die Definitionen der mathematischen Grundbegriffe. *Mathematisch-naturwissenschaftliche Blätter*, 7:93–95 and 109–113. Reprinted in (Weyl, 1968).

Weyl, H. (1918). *Das Kontinuum.* Verlag von Veit & Comp, Leipzig. Translated as *The Continuum* by S. Pollard and T. Bole, Dover, 1994.

Weyl, H. (1919). Der circulus vitiosus in der heutigen Begründung der Analysis. *Jahresbericht der Deutschen Mathematikervereinigung.* English translation in (Weyl, 1918).

Weyl, H. (1921). Über die neue Grundlagenkrise der Mathematik. *Mathematische Zeitschrift*, 10(1–2):39–79. English translation in (Mancosu, 1998).

Weyl, H. (1949). *Philosophy of Mathematics and Natural Science.* Princeton University Press.

Weyl, H. (1968). *Gesammelte Abhandlungen, volume I–IV.* Springer Verlag, Berlin.

Zermelo, E. (1908a). Neuer Beweis für die Möglichkei einer Wohlordnung. *Mathematische Annalen*, 65:107–28. Translated and reprinted in (Zermelo, 2010), pp. 140-159. Page numbers refer to the English translation.

Zermelo, E. (1908b). Untersuchungen über die Grundlagen der Mengenlehre I. *Mathematische Annalen*, 65:261–281. Translated and reprinted in (Zermelo, 2010), pp. 160-229. Page numbers refer to the English translation.

Zermelo, E. (1929a). Über den Begriff der Definitheit in der Axiomatik. *Fundamenta Mathematicae*, 14:339–344. Translated and reprinted in (Zermelo, 2010), pp. 358-367. Page numbers refer to the English translation.

Zermelo, E. (1929b). Vortrags-Themata für Warschau 1929. Translated and reprinted as *Lecture topics for Warsaw 1929* in (Zermelo, 2010), pp. 375-389. Page numbers refer to the English translation.

Zermelo, E. (1930a). Bericht an die Notgemeinschaft der Deutschen Wissenschaft über meine Forschungen betreffend die Grundlagen der Mathematik. Translated and reprinted as *Report to the Emergency Association of German Science about my reasearch concerning the foundations of mathematics* in (Zermelo, 2010), pp. 434-443. Page numbers refer to the English translation.

Zermelo, E. (1930b). Über das mengentheoretische Modell. Translated and reprinted as *On the set-theoretic model* in (Zermelo, 2010), pp. 446-453. Page numbers refer to the English translation.

Zermelo, E. (1930c). Über Grenzzahlen und Mengenbereiche. *Fundamenta Mathematicae*, 16:29–47. Translated and reprinted in (Zermelo, 2010), pp. 401–431. Page numbers refer to the English translation.

Zermelo, E. (2010). *Ernst Zermelo: Collected works. Gesammelte Werke. Volume I: Set theory, miscellania. Mengenlehre, Varia*. Springer-Verlag, Berlin.

Categoricity-like properties in the first order realm

Ali Enayat[*] Mateusz Łełyk[†]

July 22, 2024

Abstract

By classical results of Dedekind and Zermelo, second order logic imposes categoricity features on Peano Arithmetic and Zermelo-Fraenkel set theory. However, we have known since Skolem's anti-categoricity theorems that the first order formulations of Peano Arithmetic and Zermelo-Fraenkel set theory (i.e., PA and ZF) are not categorical. Here we investigate various categoricity-like properties (including tightness, solidity, and internal categoricity) that are exhibited by a distinguished class of first order theories that include PA and ZF, with the aim of understanding what is special about canonical foundational first order theories.

Contents

[*] Department of Philosophy, Linguistics, and Theory of Science, University of Gothenburg, Sweden (e-mail: ali.enayat@gu.se). Research supported by the National Science Centre, Poland (NCN), grant number 2019/34/A/HS1/00399.

[†] Faculty of Philosophy, University of Warsaw, Poland (e-mail: mlelyk@uw.edu.pl). Research supported by the National Science Centre, Poland (NCN), grant number 2022/46/E/HS1/00452.

1 Introduction

By classical results of Dedekind and Zermelo, second order logic imposes categoricity features on Peano Arithmetic and Zermelo-Fraenkel set theory. More explicitly:

Dedekind Categoricity Theorem (1888). *There is a sentence σ in second order logic of the form $\forall X \varphi(X)$, where $\varphi(X)$ only has first order quantifiers, such that σ holds in a structure $\mathcal{M}$ iff $\mathcal{M} \cong (\mathbb{N}, S, 0)$, where S is the successor function.*

Zermelo Quasi-categoricity Theorem (1930). *There is a sentence θ in second order logic of the form $\forall X \psi(X)$, where $\psi(X)$ only has first order quantifiers, such that θ holds in a structure $\mathcal{M}$ iff $\mathcal{M} \cong (V_\kappa, \in)$, where κ is a strongly inaccessible cardinal.*

The above categoricity results have captured the imagination of several generations of philosophers in relation to the debate about the determinacy of the truth-value of arithmetical and set-theoretical statements. The debate is mediated by Skolem's ingenious theorems that indicate that categoricity fails dramatically for the first order formulation PA of Peano Arithmetic, and for the first order formulation ZF of Zermelo-Fraenkel set theory.[1]

Skolem Anti-categoricity Theorems.

(a) (1934) There is a structure $\mathcal{M}$ such that $\mathsf{Th}(\mathbb{N}, +, \cdot) = \mathsf{Th}(\mathcal{M})$, but $\mathcal{M}$ contains an 'infinite' element. Thus $\mathcal{M} \ncong (\mathbb{N}, +, \cdot)$.

(b) (1922) Every structure in a countable language has a countable elementary submodel. Thus, for any strongly inaccessible cardinal κ, there is a (countable) structure $\mathcal{M} \ncong (V_\kappa, \in)$ with $\mathsf{Th}(\mathcal{M}) = \mathsf{Th}(V_\kappa, \in)$.

Here we bring together two distinct research threads in the foundation of mathematics that can be viewed as attempting to regain the 'lost' categoricity of PA and ZF by formulating 'categoricity-like' features that demonstrably hold for PA, ZF, and certain other canonical foundational theories.

One of these threads can be traced back to a potent result of Visser concerning PA that appears in his substantial paper [58] that builds a category-theoretic framework for the study of (relative) interpretability theory. Visser's result inspired Enayat [12] to introduce the notion of solidity of an arbitrary first order theory, which allows Visser's result to be expressed as: PA *is a solid theory*. Solidity is defined model-theoretically (see Definition 7), but the completeness theorem of first order logic makes it clear that it implies a remarkable purely syntactic condition dubbed *tightness*: a theory T is tight if there is no pair of distinct deductively closed extensions of T (in the same language as T) that are bi-interpretable (for the definition of bi-interpretability, see Definition 1). Thus Visser's aforementioned theorem on the solidity of PA implies that if T_1 and T_2 are two consistent extensions of T in the same language as PA such that for some arithmetical sentence φ, we have:

$$T_1 \vdash \varphi \text{ and } T_2 \nvdash \varphi,$$

then T_1 and T_2 are not bi-interpretable. The solidity of certain canonical foundational theories, including ZF, was established in [12] (see Theorem 14 of this paper). Enayat's work was further extended by Freire and Hamkins [21] in the context of ZF and its fragments; their work shows that Z (Zermelo set theory) and ZF$^-$ (ZF without the powerset axiom) fail to be tight (and therefore are not solid). More recently, Freire and Williams [22] investigated tightness in the context of fragments of Kelley-Morse theory of classes and their arithmetical counterparts; their work shows that the commonly studied proper subtheories of the aforementioned theories fail to be tight.

The other research thread explored in this paper has a more complicated history, but for the purposes of this introduction it can be described as germinating in the introduction of the concept of internal categoricity of PA in full second order logic by Hellman and Parsons, which was followed by Väänänen's introduction of the notion of internal categoricity of Peano Arithmetic and Zermelo-Fraenkel set theory in the context of Henkin models of second order logic in [53] and in joint work with Wang [56], and later in the context of first order logic, as in [54] and [55]. Internal categoricity has been substantially explored and debated in the philosophical literature, as witnessed by Button and Walsh's monograph [5], the recent monograph of Maddy and Väänänen [36], and in the recent work of Fischer and Zicchetti [20].

Both threads explore categoricity-like properties of first order foundational theories, but they differ in an important respect: the former thread is 'extensional' in the sense that its objects of study are first order theories viewed as a set of sentences, whereas the latter thread studies the 'intensional' *schematic representations* of PA and ZF. As we shall see in Theorem 52 of Section 4, this distinction is fundamental since despite the internal categoricity of the 'usual' axiomatizations of PA and ZF, there are other schematic representations of PA and ZF that fail to be internally categorical.

[1] Both results are seminal: the former led to the study of nonstandard models of arithmetic, and the latter is what is now commonly known as an instance of the Löwenheim-Skolem Theorem.

The paper is planned as follows. Section 2 is devoted to preliminary matters and a review of the pertinent results in the literature. Our novel technical results are presented in Sections 3 and 4. Section 3 presents a number of 'negative' results that demonstrate the failure of tightnness/soldity of commonly studied subtheories of PA and ZF, thus probing the optimality of the solidity proofs established in [12]. One of these negative results shows if T_0 is a subtheory of any of the known canonical solid theories T (such as $T = \mathsf{PA}$ or $T = \mathsf{ZF}$) that is axiomatized by a collection of sentences of bounded quantifier-complexity, then T_0 fails to be tight. Section 4 studies internal categoricity and the generalizations of the categoricity-like properties studied in [12] to the context of schematic representations. This allows us to delineate the relationship between the aforementioned threads that explore categoricity-like properties of first order foundational theories. For example, we introduce a straightforward generalization of the concept of solidity (dubbed e-solidity) that is applicable to schematic presentations of theories and holds for the usual schematic axiomatization of PA (but not for ZF), and show that e-solidity implies internal categoricity. In Section 5 we reflect on the philosophical ramifications of the technical results of the paper; and Section 6 describes what remains to be done.

Acknowledgements. We have benefitted from conversations (some via email) with several colleagues concerning the topics explored in this paper, including (in reverse alphabetical order of last names) Matteo Zicchetti, Kameryn Williams, Albert Visser, Jouko Väänänen, Stephen Simpson, Philipp Schlicht, Michael Rathjen, Fedor Pakhomov, Toby Meadows, David Marker, Roman Kossak, Leszek Kołodziejczyk, Joel Hamkins, Piotr Gruza, Alfredo Freire, Thomas Forster, Vincenzo Dimonte, Tim Button, and Lev Beklemishev. A special note of thanks goes to Daniel Isaacson and other participants of the Oxford Philosophy of Mathematics Seminar for their probing questions and perceptive comments following a seminar presentation in January 2024 by the first-named-author. Finally, hats off to the two anonymous referees whose reports were instrumental in the preparation of the final draft.

2 Preliminaries

2.1 Basic Definitions

Definition 1. We view a first order theory as a set of sentences, thus in the setting of this paper a theory need not be deductively closed. Suppose U and V are first order theories, formulated in relational languages[2] $\mathcal{L}_U$ and $\mathcal{L}_V$ (respectively); and let $\mathsf{Form}_{\mathcal{L}_U}$ and $\mathsf{Form}_{\mathcal{L}_V}$ be the set of first order $\mathcal{L}_U$-formulae and $\mathcal{L}_V$-formulae (respectively).

(a) We say that $\mathcal{I}$ *is an interpretation of* U *in* V, written $U \trianglelefteq^{\mathcal{I}} V$, if $\mathcal{I}$ specifies a *translation function*

$$\sigma : \mathsf{Form}_{\mathcal{L}_U} \to \mathsf{Form}_{\mathcal{L}_V}$$

such that for each $\varphi \in \mathcal{L}_U$,

$$U \vdash \varphi \Rightarrow V \vdash \sigma(\varphi),$$

and the translation $\varphi^\sigma := \sigma(\varphi)$ of φ satisfies the following three conditions:

(1) There is a designated $\mathcal{L}_V$-formula $\delta(x_1, ..., x_k)$ referred to as the *domain formula* (and k is referred to as the *dimension* of the interpretation).

(2) There is a designated mapping $P \mapsto F_P$ that translates each n-ary $\mathcal{L}_U$-predicate P into some kn-ary $\mathcal{L}_V$-formula F_P (including the case when P is the equality relation).

(3) The translation function σ commutes with propositional connectives, and is subject to:

$$(\forall x \varphi)^\sigma = \forall x_1 \cdots x_n \, (\delta(x_1, \cdots, x_n) \to \varphi^\sigma).$$

The translation σ is called *direct* iff it is unrelativized, i.e.

$$\delta(x_1, \ldots, x_k) = (x_1 = x_1) \wedge \ldots \wedge (x_k = x_k),$$

and σ translates equality to equality on k-tuples.

Note that each translation $\sigma : \mathsf{Form}_{\mathcal{L}_U} \to \mathsf{Form}_{\mathcal{L}_V}$ gives rise to a uniform transformation of $\mathcal{L}_V$-structures into $\mathcal{L}_U$-structures. That is, given any $\mathcal{L}_V$-structure $\mathcal{M}$ we obtain an $\mathcal{L}_U$-structure $\sigma(\mathcal{M})$ whose domain is the set defined in $\mathcal{M}^k$ by δ and each predicate P (including the equality predicate) is interpreted as the set defined in $\mathcal{M}$ by F_P. Moreover, an interpretation $\mathcal{I}$ based on σ such that $U \trianglelefteq^{\mathcal{I}} V$ gives rise to an *internal model construction that* ***uniformly*** *builds a model* $\mathcal{M}^{\mathcal{I}} \models U$ *for any* $\mathcal{M} \models V$, where $\mathcal{M}^{\mathcal{I}} := \sigma(\mathcal{M})$.[3]

[2] This assumption is only for ease of exposition.

[3] We often write $\mathcal{M}^{\mathcal{I}}$ instead of $\sigma(\mathcal{M})$ to emphasize that $\sigma(\mathcal{M}) \models U$ in the context of $U \trianglelefteq^{\mathcal{I}} V$. Note that translations relate formulae, and interpretations relate theories (and their models).

(**b**) U is *interpretable* in V, written $U \trianglelefteq V$, if $U \trianglelefteq^{\mathcal{I}} V$ for some interpretation $\mathcal{I}$. U and V are *mutually interpretable* when $U \trianglelefteq V$ and $V \trianglelefteq U$.

(**c**) We indicate the universe of each structure with the corresponding Roman letter, e.g., the universes of structures $\mathcal{M}$, $\mathcal{N}$, and $\mathcal{M}^*$ are respectively M, N, and M^*. Given an $\mathcal{L}$-structure $\mathcal{M}$ and $X \subseteq M^n$ (where n is a positive integer), we say that X is (*parametrically*) $\mathcal{M}$*-definable* iff there is an n-ary formula $\varphi(x_1, \ldots, x_n)$ in the language $\mathcal{L}_M$ (respectively $\mathcal{L}_M$, where $\mathcal{L}_M$ is the result of augmenting $\mathcal{L}$ with constant symbols c_m for each $m \in M$), such that $X = \varphi^{\mathcal{M}}$, where $\varphi^{\mathcal{M}} = \{(m_1, \cdot\cdot\cdot, m_k) \in M^n : (\mathcal{M}, m)_{m \in M} \models \varphi(c_{m_1}, \cdot\cdot\cdot, c_{m_k})\}$. We stress that in this paper $\mathcal{M}$-definability means "definability in $\mathcal{M}$ without parameters".

(**d**) Suppose $\mathcal{N}$ is an $\mathcal{L}_U$-structure and $\mathcal{M}$ is an $\mathcal{L}_V$-structure. We say that $\mathcal{M}$ (*parametrically*) *interprets* $\mathcal{N}$, written $\mathcal{M} \trianglerighteq \mathcal{N}$ ($\mathcal{M} \trianglerighteq_{\text{par}} \mathcal{N}$), iff there is a (parametric) translation σ of $\mathcal{L}_U$-formulae to $\mathcal{L}_V$-formulae such that $\mathcal{N} = \sigma(\mathcal{M})$. Unravelling the definition, this means that the universe of discourse of $\mathcal{N}$, as well as all the $\mathcal{N}$-interpretations of $\mathcal{L}_U$-predicates are (parametrically) $\mathcal{M}$-definable.[4] Note that both $\trianglelefteq$ and $\trianglelefteq_{\text{par}}$ are transitive relations.

(**e**) Suppose $\mathcal{M}$ is structure that (parametrically) interprets the structures $\mathcal{N}_0$ and $\mathcal{N}_1$. Let δ_0 and δ_1 be the domain formulae, and $\sim_0$, $\sim_1$ be translations of the equality relation in the respective translations. A (parametrically) $\mathcal{M}$-definable isomorphism between $\mathcal{N}_0$ and $\mathcal{N}_1$ is a (parametrically) $\mathcal{M}$-definable relation $R \subseteq \delta_0^{\mathcal{M}} \times \delta_1^{\mathcal{M}}$ such that the following hold:

1. The domain and the codomain of R are $\delta_0^{\mathcal{M}}$ and $\delta_1^{\mathcal{M}}$, respectively.
2. If $x_0 R x_1$ and $y_0 \sim_0 x_0$ and $y_1 \sim_1 x_1$, then $y_0 R y_1$.
3. The function $F : N_0 \to N_1$, defined: $F([x_0]_{\sim_0}) = [x_1]_{\sim_1}$ iff $x_0 R x_1$, is an isomorphism between $\mathcal{N}_0$ and $\mathcal{N}_1$.

(**f**) A structure $\mathcal{M}$ is *a* (*parametric*) *retract* of a structure $\mathcal{N}$ if there is an isomorphic copy $\mathcal{M}^*$ of $\mathcal{M}$ such that $\mathcal{M} \trianglerighteq \mathcal{N} \trianglerighteq \mathcal{M}^*$ ($\mathcal{M} \trianglerighteq_{\text{par}} \mathcal{N} \trianglerighteq_{\text{par}} \mathcal{M}^*$), and moreover there is a (parametrically) $\mathcal{M}$-definable isomorphism between $\mathcal{M}$ and $\mathcal{M}^*$.

(**g**) U is a *retract* of V iff there are interpretations $\mathcal{I}$ and $\mathcal{J}$ with $U \trianglelefteq^{\mathcal{I}} V$, and $V \trianglelefteq^{\mathcal{J}} U$ a binary U-formula F such that F is, U-verifiably, an isomorphism between id_U (the identity interpretation on U) and $\mathcal{I} \circ \mathcal{J}$ (where $\circ$ is the composition operation on interpretations). In model-theoretic terms, this translates to the requirement that every model of U is a retract of some model of V in *uniform and parameter-free* manner, i.e., that the following holds for every $\mathcal{M} \models U$:

$$F^{\mathcal{M}} : \mathcal{M} \xrightarrow{\cong} \mathcal{M}^* := \mathcal{M}^{\mathcal{I} \circ \mathcal{J}}.$$

(**h**) U and V are *bi-interpretable* iff there are interpretations $\mathcal{I}$ and $\mathcal{J}$ as above that witness that U is a retract of V, and additionally, there is a V-formula G, such that G is, V-verifiably, an isomorphism between id_V and $\mathcal{J} \circ \mathcal{I}$, where id_V is the identity interpretation. In particular, if U and V are bi-interpretable, then given $\mathcal{M} \models U$ and $\mathcal{N} \models V$, we have

$$F^{\mathcal{M}} : \mathcal{M} \xrightarrow{\cong} \mathcal{M}^* := \mathcal{M}^{\mathcal{I} \circ \mathcal{J}} \text{ and } G^{\mathcal{N}} : \mathcal{N} \xrightarrow{\cong} \mathcal{N}^* := \mathcal{N}^{\mathcal{J} \circ \mathcal{I}}.$$

(**i**) Suppose T is a theory formulated in a language $\mathcal{L}$, and T^+ is a theory formulated in a language $\mathcal{L}^+ \supseteq \mathcal{L}$, and assume without loss of generality that both $\mathcal{L}$ and $\mathcal{L}^+$ are relational languages. T^+ is said to be a *definitional extension* of T if for each n-ary relation $R \in \mathcal{L}^+ \setminus \mathcal{L}$ there is an n-ary $\mathcal{L}$-formula δ_R such that T^+ is logically equivalent to the $\mathcal{L}^+$-theory obtained by augmenting T with axioms of the form

$$\forall x_1 \cdots \forall x_n \, [R(x_1, \cdot\cdot\cdot, x_n) \leftrightarrow \delta_R(x_1, \cdot\cdot\cdot, x_n)].$$

With the above definition in mind, two theories T_1 and T_2 that are formulated in disjoint languages are said to be *definitionally equivalent* if they have a common definitional extension, i.e., there is a theory T such that T is a definitional extension of both T_1 and T_2. Definitional equivalence is also commonly referred to as *synonymy*, see [35] for more detail.

[4] In the case of translations that are not equality-preserving, $\mathcal{N}$ is the quotient structure of a definable subset of M^n under an $\mathcal{M}$-definable equivalence relation, and the predicates and functions on $\mathcal{N}$ are treated accordingly.

Remark 2. As shown by Visser in Theorems 4.2 and 4.12 of his paper [58], two theories U and V are definitionally equivalent iff they satisfy a stronger form of bi-interpretability, namely, in the definition of bi-interpretability, the condition that F is U-verifiably an isomorphism between id_U and $\mathcal{I} \circ \mathcal{J}$ is strengthened to the stronger condition that $F(x) = x$, and similarly, the condition that G is V-verifiably an isomorphism between id_V and $\mathcal{J} \circ \mathcal{I}$ is strengthened to the condition that $G(x) = x$. In model theoretic terms this translates to:

$$\mathcal{M}^{\mathcal{I}\circ\mathcal{J}} = \mathcal{M} \text{ for all } \mathcal{M} \models U, \text{ and } \mathcal{N}^{\mathcal{J}\circ\mathcal{I}} = \mathcal{N} \text{ for all } \mathcal{N} \models V.$$

Thus definitional equivalence is a stronger form of bi-interpretation; however, by a result of Friedman and Visser [23], in many cases definitional equivalence is implied by bi-interpretability, namely, when the two theories involved are sequential[5], and the bi-interpretability between them is witnessed by a pair of one-dimensional, identity preserving interpretations. See also Theorem 16 below.

2.2 Examples

In what follows, $\mathbb{Q}$ is the set of rational numbers, ω is the set of finite ordinals (i.e., natural numbers), and $\mathsf{Th}(\mathcal{M})$ is the first order theory of the structure $\mathcal{M}$.

Theorem 3. *(J. Robinson [46])* $\mathsf{Th}(\mathbb{Q}, +, \cdot)$ *and* $\mathsf{Th}(\omega, +, \cdot)$ *are bi-interpretable.*[6]

Theorem 4. *Let* $\mathsf{ZF}_{\mathsf{fin}}$ *be the result of replacing the axiom of infinity in the usual axiomatization of* ZF *by its negation, and let* TC *denote the statement that every set has a transitive closure.*

(*a*) (*Ackermann [1], Mycielski [42], Kaye-Wong [31]*) PA *and* $\mathsf{ZF}_{\mathsf{fin}} + \mathsf{TC}$ *are definitionally equivalent. Moreover,* PA *is a retract of* $\mathsf{ZF}_{\mathsf{fin}}$.

(*b*) (*Enayat-Schmerl-Visser [18]*) $\mathsf{ZF}_{\mathsf{fin}}$ *is not a retract of* PA.

Theorem 5 below arose from the work of Mostowski (based on earlier ideas going back to of Gödel).[7] In what follows,

$$\mathsf{ZF}^- := \mathsf{ZF}^{\mathsf{Sep+Coll}} \setminus \{\mathsf{Powerset}\},$$

where $\mathsf{ZF}^{\mathsf{Sep+Coll}}$ is the result of substituting the Replacement scheme in the usual axiomatization of ZF with the schemes of Separation and Collection.[8] Also, in part (b) $\mathsf{Inacc}(\kappa)$ expresses "κ is a strongly inaccessible cardinal greater than $\aleph_0$".

Theorem 5. *(Mostowski)*

(*a*) $\mathsf{Z}_2 + \Pi^1_\infty$-$\mathsf{AC}$ *is bi-interpretable with* $\mathsf{ZF}^- + \forall x\ |x| \leq \aleph_0$.

(*b*) $\mathsf{KM} + \Pi^1_\infty$-$\mathsf{AC}$ *is bi-interpretable with* $\mathsf{ZF}^- + \exists\kappa(\mathsf{Inacc}(\kappa) \wedge \forall x\ |x| \leq \kappa)$.

Remark 6. As shown in Theorem 16 bi-interpretability cannot be strengthened to definitional equivalence in Theorem 5. Note that Friedman and Visser [23] exhibited a pair of ad hoc theories to show that bi-interpretability does not imply definitional equivalence; thus Theorem 16 points to a counterexample involving canonical theories.

[5] At first approximation, a theory is sequential if it supports a modicum of coding machinery to handle finite sequences of all objects in the domain of discourse. Sequentiality is a modest demand for theories of arithmetic and set theory; however, by a theorem of Visser [60], (Robinson's) Q is not sequential. However, by a theorem of Jeřábek [28] the 'algebraic' fragment PA^- of PA is already sequential; for more on this theory, see the paragraph preceding Definition 30.

[6] As pointed out by Friedman and Visser, the main result of [23] can be used to show that bi-interpretability can be improved to definitional equivalence here.

[7] For further bibliographical references for part (a) of Theorem 5, see Notes for section VII.3 of Simpson's encyclopedic exposition [51], and for part (b) see Chapter 2 of Williams' doctoral dissertation [62].

[8] Indeed this is how ZF is sometimes axiomatized, as in Chang and Keisler's textbook on model theory[7], in contrast to set theory textbooks by Kunen [34] and Jech [27] that use the replacement scheme. Recall that the instances of the Separation scheme consist of universal generalizations of formula of the form

$$\forall b \exists a \forall x \, (x \in a \leftrightarrow \varphi(x)),$$

where the parameters of φ are suppressed, and instances of the Collection scheme consist of universal generalizations of formulae of the form

$$(\forall x \in a \, \exists y \, \psi(x, y)) \to (\exists b \, \forall x \in a \, \exists y \in b \, \psi(x, y)),$$

where the parameters of ψ are suppressed. It is well-known that ZF and $\mathsf{ZF}^{\mathsf{Sep+Coll}}$ axiomatize the same theory; but in the absence of the powerset axiom, the latter theory is stronger; as demonstrated in [24]. By a classical theorem of Scott [49], the same discrepancy between ZF and $\mathsf{ZF}^{\mathsf{Sep+Coll}}$ arises in the absence of the extensionality axiom.

2.3 Categoricity-like properties

The notions of solidity, neatness, and tightness encapsulated in the following definition were formulated in [12], but the notion of a minimalist theory in (a) below is new.

Definition 7. Suppose T is a first-order theory.

- Each of the following definitions is in the form of an implication. Note that in parts (a) and (b), the parameters in the conclusion of each of the implications should be understood to be a subset of the parameters used in the premises. [9]

(a) T is *minimalist* (*maximalist*) iff for every model $\mathcal{M} \models T$ and every model $\mathcal{N} \models T$ such that $\mathcal{M} \trianglerighteq_{\text{par}} \mathcal{N}$, there is a unique parametrically $\mathcal{M}$-definable embedding of $\mathcal{M}$ into $\mathcal{N}$ (resp. embedding of $\mathcal{N}$ into $\mathcal{M}$).

(b) T is *solid* iff the following holds for all models $\mathcal{M}$, $\mathcal{M}^*$, and $\mathcal{N}$ of T:

If $\mathcal{M} \trianglerighteq_{\text{par}} \mathcal{N} \trianglerighteq_{\text{par}} \mathcal{M}^*$ and there is a parametrically $\mathcal{M}$-definable isomorphism $i_0 : \mathcal{M} \to \mathcal{M}^*$, then there is a parametrically $\mathcal{M}$-definable isomorphism $i : \mathcal{M} \to \mathcal{N}$.

In other words, T is solid iff for every two models $\mathcal{M}$ and $\mathcal{N}$ of T, if $\mathcal{M}$ is a parametric retract of $\mathcal{N}$, then $\mathcal{M}$ and $\mathcal{N}$ are isomorphic via a parametrically $\mathcal{M}$-definable isomorphism.

(c) T is *neat* iff for any two deductively closed extensions U and V of T (both of which are formulated in the language of T), if U is a retract of V, then $V \subseteq U$.

(d) T is *tight* iff for any two deductively closed extensions U and V of T (both of which are formulated in the language of T), U and V are bi-interpretable iff $U = V$.

Remark 8. The following can be readily verified on the basis of the relevant definitions.

(a) If T is minimalist (maximalist), then T is solid.

(b) A solid theory is neat, and a neat theory is tight.

(c) Solidity, neatness and tightness are invariant under bi-interpretations.

Remark 9. Suppose we define a theory T to be *strongly solid* if, in the definition of solidity, the conclusion that there is a (parametrically) $\mathcal{M}$-definable isomorphism between $\mathcal{M}$ and $\mathcal{N}$, is changed to the existence of a (parametrically) $\mathcal{N}$-definable isomorphism between $\mathcal{N}$ and $\mathcal{M}^*$. It is easy to see that strong solidity implies solidity. Surprisingly, strong solidity is implied by solidity. We owe this observation to Leszek Kołodziejczyk. To see why, suppose that i is an isomorphism between $\mathcal{M}$ and $\mathcal{N}$, and j is an isomorphism between $\mathcal{M}$ and $\mathcal{M}^*$. Also let $h(p)$, where p denotes the parameters, be the translation that yields an interpretation of $\mathcal{M}^*$ in $\mathcal{N}$. Then $\mathcal{M}$ sees $h(i^{-1}(p))$ and sees an isomorphism between itself and the model given by $h(i^{-1}(p))$ (the isomorphism is $i^{-1} \circ j$). Hence by applying i to $\mathcal{M}$, $\mathcal{N}$ sees its isomorphism with the model given by $h(p)$, and hence with $\mathcal{M}^*$.

3 Categoricity of first-order theories

3.1 Positive results

In this section we present known results about categoricity-like properties. We begin with the following result of Albert Visser that inspired the first-named-author's paper [12]. Visser's result follows from Corollaries 9.4 and 9.6 of his paper [58]. For a streamlined proof, see [12].

Theorem 10. (*Visser*) PA *is solid.*

Another way of seeing that PA is solid is via part (a) of Remark 8 since PA can be readily shown to be a minimalist theory, as indicated in the proposition below.

Proposition 11. PA *is minimalist.*

[9] This restriction on parameters was only implicit in [12]; it is made explicit in light of two considerations: (1) the solidity proofs established in [12] that inspired the definition abide by this restriction; and (2) The proof of Theorem 74 of this paper that establishes that internally categoricity is implied by e-solidity does not go through for the 'liberal' notion of solidity that imposes no restriction on the behaviour of the parameters (indeed we have an example that shows that the aforementioned implication provably fails if the restriction stipulated here on parameters is violated).

Proof. Suppose $\mathcal{M}$ and $\mathcal{N}$ are models of PA and $\mathcal{M} \trianglerighteq_{\mathrm{par}} \mathcal{N}$. A recursion within $\mathcal{M}$ can be used to construct a parametrically definable embedding f of $\mathcal{M}$ onto an initial segment of $\mathcal{N}$, where $f(0^{\mathcal{M}}) = 0^{\mathcal{N}}$, and $f((x+1)^{\mathcal{M}}) = (f(x)+1)^{\mathcal{N}}$. Then, by using induction in $\mathcal{M}$, one can readily show that if g is a parametrically definable embedding of $\mathcal{M}$ onto an initial segment of $\mathcal{N}$, then $f(m) = g(m)$ for all m in $\mathcal{M}$.[10] □

In light of part (a) of Theorem 4 and part (b) of Remark 8, Theorem 10 yields the following corollary.

Corollary 12. $\mathsf{ZF_{fin}} + \mathsf{TC}$ *is solid.*

Remark 13. It is easy to see that PA is not a maximalist theory, since for example a nonstandard model of PA is interpretable in the standard model of PA, but of course the latter cannot be embedded in the former. It is also noteworthy that ZF is neither minimalist nor maximalist. To see the former, note that if $\mathcal{M} \models \mathsf{ZF} + \exists\kappa \mathsf{Inacc}(\kappa)$, then $\mathcal{M}$ can interpret ZF via its definable submodel $(V_\kappa, \in)^{\mathcal{M}}$, however there can be no definable embedding of $\mathcal{M}$ into $(V_\kappa, \in)^{\mathcal{M}}$.[11] To see that ZF is not a maximalist theory, let $\mathcal{M}$ be a well-founded model of ZFC, and let $\mathcal{N}$ be an internal ultrapower of $\mathcal{M}$ modulo a nonprincipal ultrafilter over $\mathcal{P}(\omega)^{\mathcal{M}}$. Note that $\mathcal{N}$ is interpretable in $\mathcal{M}$ and $\mathcal{N}$ is ill-founded, and clearly there can be no embedding of $\mathcal{N}$ into $\mathcal{M}$.

The list of examples of solid theories was extended in the first-named-author's paper [12] to other well-known foundational theories, as indicated in the next theorem. In what follows Z_n is n-th order arithmetic, and KM_n is n-th order Kelley-Morse theory of classes, where $\mathsf{Z}_1 := \mathsf{PA}$, and $\mathsf{KM}_1 := \mathsf{ZF}$. Also, Z_ω is the theory of types (with full comprehension) whose level-zero objects form a model of PA (equivalently $\mathsf{ZF_{fin}} + \mathsf{TC}$), and KM_ω is the theory of types whose level-zero objects form a model of ZF are also solid theories.

Theorem 14. *Each of the theories in the family $\mathcal{S}$ of theories below is solid:*

$$\mathcal{S} = \{\mathsf{Z}_n : 1 \leq n \in \omega\} \cup \{\mathsf{KM}_n : 1 \leq n \in \omega\} \cup \{\mathsf{Z}_\omega, \mathsf{KM}_\omega\}.$$

As noted in [12], the following corollary follows from putting part (c) of Remark 8 together with Theorem 14 and the bi-interpretability of the theories in (a) and (b) of Theorem 5 with appropriate extension of Z_2 and KM (respectively). Recall from the paragraph preceding Theorem 5 that ZF^- is the result of removing the Powerset axiom from $\mathsf{ZF}^{\mathsf{Sep+Coll}}$.

Theorem 15. *The following theories are solid:*

(*a*) $\mathsf{ZF}^- + \forall x \ |x| \leq \aleph_0$.

(*b*) $\mathsf{ZF}^- + \exists\kappa(\mathrm{Inacc}(\kappa) \wedge \ \forall x \ |x| \leq \kappa)$.

Theorem 16. *Bi-interpretability cannot be improved to definitional equivalence in parts* (*a*) *and* (*b*) *of Theorem 5* (*assuming the consistency of* ZF *plus an inaccessible cardinal for part* (*a*), *and the consistency of* ZF *with two inaccessible cardinals for part* (*b*)).

Proof. We first deal with part (a). Note that by a classical argument going back to Mostowski (which inspired Theorem 5), within ZFC the standard model $(\mathcal{P}(\omega), \omega, +, \cdot, \in)$ of $\mathsf{Z}_2 + \Pi^1_\infty\text{-}\mathsf{AC}$ is bi-interpretable with the standard model $(H(\omega_1), \in)$ of $\mathsf{ZF}^- + \forall x \ |x| \leq \aleph_0$. Thus, thanks to Theorem 5 and part (a) of Theorem 15 any model of $\mathsf{ZF}^- + \forall x \ |x| \leq \aleph_0$ that is bi-interpretable with $(\mathcal{P}(\omega), \omega, +, \cdot, \in)$ must be isomorphic to $(H(\omega_1), \in)$.

On the other hand, by a classical theorem, due to Solovay [52, Theorem 2], assuming the existence of an inaccessible cardinal, there is a model $\mathcal{M}$ of ZFC (obtained by set forcing) in which CH holds (i.e., $2^{\aleph_0} = \aleph_1$) and all projective sets of reals are Lebesgue measurable. Note that (1) the submodel $H(\omega_1)^{\mathcal{M}}$ of hereditarily countable sets of such an $\mathcal{M}$ satisfies $\mathsf{ZF}^- + \forall x \ |x| \leq \aleph_0$, and the standard model $(\mathcal{P}(\omega), \omega, +, \cdot, \in)^{\mathcal{M}}$ of Z_2 of $\mathcal{M}$ satisfies $\Pi^1_\infty\text{-}\mathsf{AC}$. Within $\mathcal{M}$ evidently $H(\omega_1)$ has a parameter-free definable subset of order-type ω_1, but $(\mathcal{P}(\omega), \omega, +, \cdot, \in)$ does not have a parametrically definable subset of order-type ω_1. More specifically, recall that the projective sets are precisely the subsets of $\mathcal{P}(\omega)$ that are parametrically definable in $(\mathcal{P}(\omega), \omega, +, \cdot, \in)$; and under CH, Fubini's theorem (in measure theory) implies that any linear order on $\mathcal{P}(\omega)$ that is of order type ω_1 fails to be measurable when viewed as a subset of the cartesian product of $\mathcal{P}(\omega)$ with itself as first observed by Sierpiñski. Here $\mathcal{P}(\omega)$ is identified with the coin-tossing product measure space 2^ω. In light of the first paragraph of the proof, this shows that the definitional equivalence of the theories $\mathsf{Z}_2 + \Pi^1_\infty\text{-}\mathsf{AC}$ and $\mathsf{ZF}^- + \forall x \ |x| \leq \aleph_0$ within $\mathcal{M}$ implies that, as viewed in $\mathcal{M}$, there is a projective nonmeasurable subset of the cartesian product of $\mathcal{P}(\omega)$ with itself, contradiction.

[10] Note that the proof only uses the fact that $\mathcal{N} \models \mathsf{Q}$ (where Q is Robinson Arithmetic).

[11] One can bypass the appeal to the existence of an inaccessible cardinal in $\mathcal{M}$ with a more refined argument that takes advantage of the reflection theorem in $\mathcal{M}$.

A similar argument can be carried out for part (b), using a generalization of Solovay's proof (which requires the existence of at least two inaccessible cardinals) in which ω is replaced by an inaccessible cardinal, and "projective" is replaced by "parametrically definable in the natural model of KM associated with V_κ in which classes are interpreted as members of $V_{\kappa+1}$". For a proof outline of this generalization, see Schultzenberg's MathOverflow answer [48].

Note that the statement "U and V are definitionally equivalent", where both U and V are arithmetically definable theories (let alone computable theories) is an *arithmetical sentence*, and therefore its truth-value does not change in the passage to a Boolean-valued extension of the universe obtained by forcing; thus, we have shown (reasoning in ZFC plus "there is an inaccessible cardinal") that the two theories asserted to be bi-interpretable in part (a) of Theorem 5 not only fail to be definitionally equivalent in some model of ZFC, but are outright not definitionally equivalent assuming the existence of an inaccessible cardinal. A similar comment applies to part (b) of Theorem 5, assuming the existence of two inaccessible cardinals. □

Remark 17. In private communication, Vincenzo Dimonte has pointed out that the above generalization of Solovay's theorem (for the proof of part (b) of Theorem 16) can be derived from Theorem 2.19 and Lemma 2.21 of Schlicht' paper [47](that pertain to the same model considered by Schultzenberg). Moreover, one of the referees has pointed out the following two points. Firstly, Schultzenberg's aforementioned proof in [48] contains a small gap that can be filled by consulting Kanamori's exposition ([29], page 141) of Solovay's proof that the perfect set property holds. Secondly, the assumption for part (a) of Theorem 16 can probably be reduced to the consistency of ZF since by forcing with the partial order $\mathrm{Col}(\omega, <\mathrm{Ord})$ one obtains a model that sufficiently behaves like the Solovay model (and one can probably use the same idea to reduce the assumption in part (b) of Theorem 16 to the consistency of ZF plus an inaccessible).

The following adds another prominent theory to the list of solid theories.

Theorem 18. $\mathsf{ZF} \setminus \{\mathsf{Infinity}\} + \mathsf{TC}$ *is solid.*

Proof. As shown by Visser (see Corollaries 9.6 and 9.8 of [58]), if U is a consistent extension of PA and V is a consistent extension of ZF, then U is not a retract of V, and V is not a retract of U. An inspection of Visser's proofs of the aforementioned results allows us to conclude the following stronger statements:

(1) No model of PA is a parametric retract of a model of ZF.

(2) No model of ZF is a parametric retract of a model of PA.

Since TC is a theorem of ZF, it is routine to verify that $\mathsf{ZF} \setminus \{\mathsf{Infinity}\} + \mathsf{TC}$ is a solid theory with the help of (1) and (2) together with the bi-interpretability of $\mathsf{ZF_{fin}} + \mathsf{TC}$ and PA (part (a) of Theorem 4), the solidity of PA and ZF (Theorem 14), and part (c) of Remark 8. □

3.2 Reflecting on the positive results

Let $\mathcal{S}$ be the list of theories whose solidity is asserted in Theorem 14. The following question is motivated by the fact that none of the theories $T \in \mathcal{S}$ is finitely axiomatizable. This is because each theory $T \in \mathcal{S}$ is *inductive*, i.e., T proves $\mathsf{Q}^{\mathcal{N}} + \mathsf{Ind}^{\mathcal{N}}(\mathcal{L}_T\ )$; here $\mathcal{N}$ is a designated interpretation $\mathcal{N}$ of arithmetic in T, Q is Robinson Arithmetic, and $\mathsf{Ind}^{N}(\mathcal{L}_T)$ is the scheme of induction over natural numbers for $\mathcal{L}_T$-formulae (in which the parameters are allowed to vary over the domain of discourse and are not limited to the 'numbers' of $\mathcal{N}$). Recall that by a classical theorem of Montague [41], for a sequential[12] theory T, the T-provability of $\mathsf{Q}^{\mathcal{N}} + \mathsf{Ind}^{\mathcal{N}}(\mathcal{L}_T\ )$ implies that T is not finitely axiomatizable.

Question A. *Is there a consistent sequential finitely axiomatized solid theory*?

In Theorem 39 and 38 of this paper we give a partial negative answer to Question A by showing that finitely axiomatized subtheories of the theories PA, $\mathsf{Z_2}$, ZF, and KM fail to be tight (and therefore they are not solid). Indeed our method can be used to show that no finitely axiomatized subtheory of any theory T in the list $\mathcal{S}$ of Theorem 14 is tight.

[12] At first approximation, a theory is sequential if it supports a modicum of coding machinery to handle finite sequences of all objects in the domain of discourse. Sequentiality is a modest demand for theories of arithmetic and set theory; however, by a theorem of Visser [60], (Robinson's) Q is not sequential. There are many equivalent definitions of sequentiality; the original definition due to Pudlák is as follows: A theory T is sequential if there is a formula $N(x)$, together with appropriate formulae providing interpretations of equality, and the operations of successor, addition, and multiplication for elements satisfying $N(x)$ such that T proves the translations of the axioms of Q (Robinson's arithmetic) when relativized to $N(x)$; and additionally, there is a formula $\beta(x, i, w)$ (whose intended meaning is that x is the i-th element of a sequence w) such that T proves that every sequence can be extended by any given element of the domain of discourse. For more detail and references see Visser's [59].

An inspection of the proofs of the different cases of Theorem 14 presented in [12] makes it clear that the proofs of solidity of each of the $T \in \mathcal{S}$ uses the 'full power' of T. This prompts the next question.

Question B. *Is there an example T of one of the theories whose solidity is established in Theorem 14, and some solid deductively closed proper subtheory of T?*

Remark 19. Very recent joint work (in progress) of Piotr Gruza, Leszek Kołodziejczyk, and the second-named-author of this paper shows that Question B has a positive answer for $T = \mathsf{PA}$; indeed their work shows that the example in this case can be even required to contain any prescribed $\mathsf{I}\Sigma_n$. Also, note that by putting Corollary 12 and part (a) of Theorem 18 we obtain a positive answer to Question B for the subtheory $\mathsf{ZF} \setminus \{\mathsf{Infinity}\} + \mathsf{TC}$ of $T = \mathsf{ZF}$.

In the remaining subsections of this section, we first review published results concerning the failure of solidity/tightness established in [12], [21], and [22] in Subsection 3.3, and then we present new results concerning the failure of solidity/tightness in Subsection 3.4.

3.3 Known negative results

As pointed out in [12] inductive[13] sequential theories need not be tight since the theory $\mathsf{PA}(G)$, with no extra axioms for G, is not tight (where G is a fresh predicate). Here $\mathsf{PA}(G)$ is the result of augmenting PA with the scheme of induction for all formulae in the language $\mathcal{L}_{\mathsf{PA}} \cup \{G\}$. In the same paper, the proofs of the following results were outlined:

(1) $\mathsf{ZF_{fin}}$ is not tight (the proof is based on a construction from [18]). This makes it clear that $\mathsf{ZF} \setminus \{\mathsf{Infinity}\}$ is not tight, since tightness is inherited by theory extensions (in the same language).

(2) $\mathsf{ZF} \setminus \{\mathsf{Foundation}\}$ and $\mathsf{ZF} \setminus \{\mathsf{Extensionality}\}$ is not tight (the proofs employs classical techniques for building models in which Extensionality or Foundation fails).

The list of proper subtheories of ZF that fail to be tight/solid was further extended by Freire and Hamkins [21], who showed that Z (Zermelo set theory) and $\mathsf{ZF} \setminus \{\mathsf{Powerset}\}$ also fail to be tight (even when ZF is formulated with the schemes of Separation and Collection).

More recently, Freire and Williams [22] established the failure of tightness of well-known fragments of Z_2 and KM. Their work shows that for each $n \in \omega$, the fragments of Z_2 and KM in which the comprehension schema is limited to Π^1_n-formulae fails to be tight even when the full scheme of induction (for the case of fragments of Z_2) or the full scheme of $\in$-induction (for the case of fragments of KM) are included. In particular, their work shows that the subsystems Π^1_n-CA of Z_2 in which the comprehension scheme is limited to Π^1_n-formulae (but the full induction scheme is kept) fail to be tight.[14]

3.4 New negative results

Definition 20. $\mathbb{N}$ is the standard model of PA, i.e., $(\omega, +, \cdot)$, and given a model of arithmetic $\mathcal{M}$,

$$\mathsf{Th}_{\Pi_n}(\mathcal{M}) = \{\varphi \in \Pi_n : \mathcal{M} \models \varphi\}.$$

Also $\mathsf{PA}_{\Pi_n} = \{\varphi \in \Pi_n : \mathsf{PA} \vdash \varphi\}$.[15] Here Π_n refers to the usual hierarchy of arithmetical formulae as in [30] and [25].

Definition 21. For $\mathcal{M} \models \mathsf{PA}$ and $n \in \omega$, $K_n(\mathcal{M})$ is the submodel of $\mathcal{M}$ whose universe consists of elements of $\mathcal{M}$ that are definable in $\mathcal{M}$ by a Σ_n-formula.

The following result is classical; it is due independently to the joint work of Kirby and Paris [44] and to Lessan.[16] An exposition can be found in Kaye's monograph [30].

In what follows the notation $\mathcal{A} \prec_{\Pi_n} \mathcal{B}$ means that Π_n-formulae are absolute in the passage between the arithmetical structures $\mathcal{A}$ and $\mathcal{B}$; $\mathsf{I}\Sigma_n$ is the fragment of PA in which the induction scheme is limited to Σ_n-formulae, and $\mathsf{B}\Sigma_n$ is the result of restricting the collection scheme to Σ_n-formulae.

[13] Inductive theories are defined in the paragraph preceding Question A.

[14] Independently and earlier, the first-named-author of this paper demonstrated the failure of tightness of the subsystem ACA (i.e., $\mathsf{ACA_0}$ with the full scheme of induction) using a forcing argument similar to the one used by Freire and Williams. The proof was presented in an online seminar talk [13].

[15] It is easy to see that $\mathrm{I}\Sigma_n$ is a proper subtheory of $\mathsf{PA}_{\Pi_{n+2}}$. The inclusion follows from the fact that each axiom of $\mathrm{I}\Sigma_n$ is of complexity at most Π_{n+2}; for the inequality, use Gödel's second incompleteness theorem and the fact that for all $n \in \omega$, $\mathsf{Con}_{\mathrm{I}\Sigma_n} \in \mathsf{PA}_{\Pi_1}$ (thanks to Mostowski's Reflection Theorem, which asserts that PA and all of its extensions prove the formal consistency of each of their finitely axiomatized subtheories).

[16] The result appears in Lessan's doctoral dissertation (Manchester, 1978) that was reprinted in [6].

Theorem 22. (*Kirby-Paris, Lessan*) *Suppose $n \in \omega$, $n \geq 1$, and $\mathcal{M}$ is a nonstandard model of* PA*, then:*

(*a*) $K_n(\mathcal{M}) \prec_{\Pi_n} \mathcal{M}$*, hence* $K_n(\mathcal{M}) \models \mathrm{Th}_{\Pi_{n+1}}(\mathcal{M})$.

(*b*) *If $K_n(\mathcal{M})$ is nonstandard, then* $K_n(\mathcal{M}) \models \mathsf{PA}_{\Pi_{n+1}} + \mathsf{I}\Sigma_{n-1} + \neg\mathsf{B}\Sigma_n$.

The following result also appears in Lessan's doctoral dissertation.

Theorem 23. (*Lessan*) *Suppose $n \in \omega$, $n \geq 1$, $\mathcal{M} \models$ PA, and $K_n(\mathcal{M})$ is nonstandard for some model $\mathcal{M} \models$ PA. Then the standard cut ω is first order definable in $K_n(\mathcal{M})$.*

Proof. (Outline) The key idea is that the complement of the standard cut can be defined in $K_n(\mathcal{M})$ via the formula that expresses "every element is definable by a Σ_n-formula whose code is below x". Recall that there is a Σ_k-definable satisfaction class of complexity Σ_k within models of $\mathrm{I}\Delta_0 + \mathsf{Exp}$ for each nonzero $k \in \omega$ (see [25]). In particular since $K_n(\mathcal{M}) \models \mathsf{PA}_{\Pi_2}$ for $n \geq 1$, $K_n(\mathcal{M})$ carries a Σ_n-definable satisfaction class of complexity Σ_k for each nonzero $k \in \omega$. □

Theorem 24. *For each $n \in \omega$ the following statements hold:*

(*a*) *$\mathbb{N}$ is bi-interpretable with a nonstandard model of the form $K_{n+1}(\mathcal{M})$, where $\mathcal{M} \models$ PA, and* $K_{n+1}(\mathcal{M}) \models \mathrm{Th}_{\Pi_n}(\mathbb{N}) + \neg\mathsf{B}\Sigma_{n+1}$.

(*b*) $\mathrm{Th}(\mathbb{N})$ *is bi-interpretable with a complete extension of* $\mathsf{PA}_{\Pi_{n+2}} + \mathrm{Th}_{\Pi_n}(\mathbb{N}) + \neg\mathsf{B}\Sigma_{n+1}$.

Proof. To establish (a), we note that for a prescribed $n \in \omega$, $\mathrm{Th}_{\Pi_n}(\mathbb{N})$ is arithmetically definable, thanks to the existence of a definable Π_n-satisfaction class, i.e., an arithmetically definable predicate that satisfies Tarski's recursive clauses for a satisfaction predicate for all Π_n-formulae. Consider the arithmetically definable theories T_n and T_n^+ defined by:

$$T_n := \mathsf{PA} + \mathrm{Th}_{\Pi_n}(\mathbb{N}), \text{ and } T_n^+ := T_n + \neg\mathsf{Con}(T_n).$$

Recall that Gödel's second incompleteness theorem, when phrased for arbitrary consistent sound arithmetical theories U, states that $U + \neg\mathsf{Con}(U)$ is consistent (see, e.g., Chao and Seraji's [8]). Here U is sound means that $\mathbb{N} \models U$. Therefore $\mathbb{N} \models \mathsf{Con}(T_n^+)$. Thanks to the arithmetized completeness theorem, this makes it clear that there is an arithmetical model $\mathcal{M} \models T_n^+$; thus $\mathbb{N} \trianglerighteq \mathcal{M}$, so we can fix a translation σ_1 such that $\mathcal{M} = \sigma_1(\mathbb{N})$.

Let $c \in M$ be the least proof of inconsistency of T_n in $\mathcal{M}$. Since True_{Π_n} is Π_n-definable (for $n \geq 1$), the predicate $P(x)$ expressing "x codes a proof of contradiction from $\mathsf{PA} + \mathsf{True}_{\Pi_n}$" is Π_n, and therefore the predicate $Q(x) := P(x) \wedge \forall y < x\, \neg P(x)$ is of the form $\Pi_n \wedge \Sigma_n$; hence c is Σ_{n+1}-definable in $\mathcal{M}$. This shows that $K_{n+1}(\mathcal{M})$ is nonstandard. Thanks to the availability of a truth-definition for Σ_{n+1}-formulae in $\mathcal{M}$, $K_{n+1}(\mathcal{M})$ is interpretable in $\mathcal{M}$, thus there is a translation σ_2 such that $K_{n+1}(\mathcal{M}) = \sigma_2(\mathcal{M})$. This makes it clear that by composing σ_1 and σ_2 we obtain a translation σ such that $K_{n+1}(\mathcal{M}) = \sigma(\mathbb{N})$.

By Theorem 22,

$$K_{n+1}(\mathcal{M}) \models \mathsf{PA}_{\Pi_{n+2}} + \neg\mathsf{B}\Sigma_{n+1},$$

and by Theorem 23, $\mathbb{N}$ is definable in $K_{n+1}(\mathcal{M})$. This makes it clear that $K_{n+1}(\mathcal{M}) \trianglerighteq \mathbb{N}$. It is routine now to verify that $\mathbb{N}$ is a retract of $K_{n+1}(\mathcal{M})$. Note that this shows that $\mathrm{Th}_{\Pi_n}(\mathbb{N})$ is not solid since $\mathbb{N}$ and $K_{n+1}(\mathcal{M})$ are non-isomorphic. Indeed, $K_{n+1}(\mathcal{M})$ is also a retract of $\mathbb{N}$ (with the same interpretations at work) since $K_{n+1}(\mathcal{M})$ has a truth-definition for Σ_{n+1}-formulae, so it can define an isomorphism between itself, and the model that $\mathbb{N}$ cooks up via the translation σ (recall that $K_{n+1}(\mathcal{M}) = \sigma(\mathbb{N})$). This concludes the proof of (a).

The proof of (b) is based on the observation that the proof of (a) works *uniformly* for all models of for $U = \mathrm{Th}(\mathbb{N})$ and $V = \mathrm{Th}(K_{n+1}(\mathcal{M}))$.

□

Part(b) of Theorem 24 immediately implies the following result.

Theorem 25. *The following statements hold for each $n \in \omega$.*

(*a*) $\mathsf{Th}_{\Pi_n}(\mathbb{N})$ *is not tight.*

(*b*) PA_{Π_n} *is not tight.*

The failure of tightness (and, a fortiori, the failure of solidity) for bounded fragments of PA (i.e., those axiomatized by a collection of sentences of bounded quantifier complexity) naturally prompts one to investigate the possibility of solidity/tightness of commonly studied fragments of PA that are unbounded. One such fragment is presented in the following definition.

Definition 26. Given a recursively enumerable theory T extending $\mathrm{I}\Delta_0 + \mathsf{Exp}$, let Ref_T consist of the collection of sentences of the form

$$\mathsf{Prov}_T(\ulcorner\varphi\urcorner) \to \varphi,$$

where φ ranges over arithmetical sentences, and $\mathsf{Prov}_T(x)$ is the arithmetical sentence that expresses "φ is provable in T ".

Proposition 27. (*Feferman* [19]). $T + \mathsf{Th}_{\Pi_1}(\mathbb{N}) \vdash \mathsf{Ref}_T$.[17]

Proof. It suffices to show that if $\mathcal{M} \models T + \mathsf{Th}_{\Pi_1}(\mathbb{N})$, then $\mathcal{M} \models \mathsf{Prov}_T(\ulcorner\varphi\urcorner) \to \varphi$ for any given arithmetical sentence φ. Suppose not, then:

(1) $\mathcal{M} \models \mathsf{Prov}_T(\ulcorner\varphi\urcorner)$, and

(2) $\mathcal{M} \models \neg\varphi$.

The key observation is that since $\mathsf{Prov}_T(\ulcorner\varphi\urcorner)$ is a Σ_1-statement, $\mathbb{N} \models \mathsf{Prov}_T(\ulcorner\varphi\urcorner)$; otherwise $\mathbb{N} \models \neg\mathsf{Prov}_T(\ulcorner\varphi\urcorner)$, which by the assumption that $\mathcal{M} \models \mathsf{Th}_{\Pi_1}(\mathbb{N})$, contradicts (1). But if $\mathbb{N} \models \mathsf{Prov}_T(\ulcorner\varphi\urcorner)$, then φ is provable in T, which contradicts (2). □

Theorem 28. $\mathsf{PA}_{\Pi_n} + \mathsf{Ref}_{\mathsf{PA}_{\Pi_n}}$ *is not tight for each* $n \in \omega$.

Proof. This is readily established by putting Theorem 25 together with Proposition 27 (for $T = \mathsf{PA}_{\Pi_n}$). □

Remark 29. It is important to bear in mind that the stronger scheme REF_T whose instances are *uniform versions* of the instances of Ref_T, i.e., implications of the form

$$\forall x\ \mathsf{Prov}_T(\ulcorner\varphi(\dot{x})\urcorner) \to \varphi(x)),$$

behaves very differently, since by a classical result of Kreisel and Levy, [33], for $T = \mathrm{I}\Delta_0 + \mathsf{Exp}$ the deductive closure of $T + \mathsf{REF}_T$ coincides with the deductive closure of PA.

Another unbounded subtheory of PA that we shall consider is $\mathsf{PA}^- + \mathsf{Collection}$. The following definition describes an important family of models of the well-known fragment PA^- of PA, whose axioms describe the non-negative substructure of discretely ordered rings (with no instance of the induction scheme, hence the minus superscript), as in Chapter 2 of Kaye's text [30] on models of PA. As we shall see, $\mathsf{PA}^- + \mathsf{Collection}$ fails to be solid.

Definition 30. Let $\mathbb{Z}$ be the ring of integers and $(I, <_I)$ be a linear order.

(a) $\mathbb{Z}[X_i : i \in I]$ is the ring of polynomials with coefficients in $\mathbb{Z}$, whose indeterminates come from the I-indexed collection $\{X_i : i \in I\}$ of indeterminates.

(b) Each nonconstant $p \in \mathbb{Z}\,[X_i : i \in I]$ can be written as:

$$p(X_{i_1}, X_{i_2}, \cdots, X_{i_n}) \text{ where } i_1 <_I i_2 <_I \cdots <_I i_n,$$

for some finite subset $\{i_1, i_2, \cdots, i_n\}$ of I, with the understanding that $\{i_1, i_2, \cdots, i_n\}$ is the least (in the sense of inclusion) subset of $A \subseteq I$ such that $p \in \mathbb{Z}[X_i : i \in A]$. With this notation in mind, we refer to $\{i_1, i_2, \cdots, i_n\}$ as the *support* of p.

(c) The ordering on $\mathbb{Z}\,[X_i : i \in I]$ is defined by first declaring $p(X_{i_1}, X_{i_2}, \cdots, X_{i_n}) > 0$, provided the coefficient of X_{i_n} is positive; and then given p and q in $\mathbb{Z}\,[X_i : i \in I]$, we define $p > q$ iff $p - q > 0$.

The following proposition is well-known and easily verified.

Proposition 31. *For every linear order* $(I, <_I)$ *the substructure* $\mathbb{Z}[X_i : i \in I]^{\geq 0}$ *of non-negative elements of* $\mathbb{Z}[X_i : i \in I]$ *is a model of* PA^-.

[17] In Feferman's formulation, T was specifically chosen as PA, but the same reasoning handles the general case. Also, as pointed by Lev Beklemishev (in private conversation) Feferman noticed this fact while thinking on Turing's (false) hope that iterated local reflection can prove all true Π_2-sentences.

Remark 32. The set ω of standard natural numbers (equivalently: constant non-negative polynomials) is defined in $\mathbb{Z}[X_i : i \in I]^{\geq 0}$ as the longest initial segment of elements with parity, i.e., for all $p \in \mathbb{Z}[X_i : i \in I]^{\geq 0}$, $p \in \omega$ iff $\forall q \leq p \, \exists r \, [(q = 2r) \vee (q = 2r+1)]$.

In the proposition below, $\mathbb{Q}$ is the set of rational numbers and $\mathbb{Q}^-$ is the set of negative rational numbers, both equipped with their natural ordering.

Proposition 33. $\mathbb{Z}\,[X_i : i \in \mathbb{Q}]^{\geq 0}$ *elementarily end extends* $\mathbb{Z}\,[X_i : i \in \mathbb{Q}^-]^{\geq 0}$.

Proof. It is routine to verify that $\mathbb{Z}\,[X_i : i \in \mathbb{Q}^-]^{\geq 0}$ is end extended by $\mathbb{Z}\,[X_i : i \in \mathbb{Q}]^{\geq 0}$, so we focus on verifying elementarity. Suppose $\varphi(p_1, \cdots, p_k)$ is an arithmetical sentence all of whose parameters $p_1, \cdots, p_k$ are from $\mathbb{Z}[X_i : i \in \mathbb{Q}^-]^{\geq 0}$. Since the support of each polynomial is finite, there is an initial segment $\mathbb{S}$ of $\mathbb{Q}^-$ such that:

(1) $\mathbb{S}$ has no last element;

(2) $\mathbb{Q}^- \backslash \mathbb{S}$ has no first element; and

(3) For $1 \leq i \leq k$, the support of p_i is a subset of $\mathbb{S}$.

Note that (1) implies that $\mathbb{S}$ is a DLO (dense linear order without end points), and (2) implies that both $\mathbb{Q}^- \backslash \mathbb{S}$ and $\mathbb{Q} \backslash \mathbb{S}$ are also DLOs. Thanks to $\aleph_0$-categoricity of DLOs, this shows that there is an order isomorphism $f : \mathbb{Q}^- \to \mathbb{Q}$ such that $f(s) = s$ for each $s \in \mathbb{S}$. This isomorphism naturally induces an isomorphism F with:

$$F : \mathbb{Z}\left[X_i : i \in \mathbb{Q}^-\right]^{\geq 0} \to \mathbb{Z}\left[X_i : i \in \mathbb{Q}\right]^{\geq 0},$$

such that F fixes all elements in $\mathbb{Z}\,[X_i : i \in \mathbb{S}]^{\geq 0}$. This makes it clear that $\varphi(p_1, \cdots, p_k)$ holds in $\mathbb{Z}\,[X_i : i \in \mathbb{Q}^-]^{\geq 0}$ iff $\varphi(p_1, \cdots, p_k)$ holds in $\mathbb{Z}\,[X_i : i \in \mathbb{Q}]^{\geq 0}$, since by an elementary fact of model theory if g is an isomorphism between two $\mathcal{L}$-structures $\mathcal{M}$ and $\mathcal{N}$, then for any n-ary $\mathcal{L}$-formula $\varphi(x_1, \cdots, x_k)$, and any k-tuple $(a_1, \cdots, a_k)$ from $\mathcal{M}$, $\mathcal{M} \models \varphi(a_1, \cdots, a_k)$ iff $\mathcal{N} \models \varphi(g(a_1), \cdots, g(a_k))$. □

In light of the well-known fact that the Collection scheme holds in linearly ordered structures that possess an elementary end extension[18], the following is an immediate consequence of Proposition 33.

Corollary 34. $\mathbb{Z}[X_i : i \in \mathbb{Q}]^{\geq 0} \models \mathsf{Collection}$.

Theorem 35. $\mathsf{PA}^- + \mathsf{Collection}$ *is not solid.*[19]

Proof. By Proposition 31 and Corollary 34 $\mathbb{Z}[X_i : i \in \mathbb{Q}]^{\geq 0}$ is a model of $\mathsf{PA}^- + \mathsf{Collection}$, and of course so is $\mathbb{N}$, and the two structures are nonisomorphic, so it suffices to verify that there are $\mathcal{I}$ and $\mathcal{J}$ such that:

$$\mathbb{N} \trianglerighteq^{\mathcal{I}} \mathbb{Z}[X_i : i \in \mathbb{Q}]^{\geq 0} \trianglerighteq^{\mathcal{J}} \mathbb{N},$$

and there is an $\mathbb{N}$-definable isomorphism $i : \mathbb{N} \to \mathbb{N}^{\mathcal{J} \circ \mathcal{I}}$. The description of $\mathbb{Z}[X_i : i \in \mathbb{Q}]^{\geq 0}$, and the fact that there is an arithmetically definable ordering $<_{\mathbb{Q}}$ on ω such that $(\omega, <_{\mathbb{Q}}) \cong \mathbb{Q}$ yields $\mathcal{I}$ such that $\mathbb{N} \trianglerighteq^{\mathcal{I}} \mathbb{Z}[X_i : i \in \mathbb{Q}]^{\geq 0}$. On the other hand, the definability of ω in $\mathbb{Z}[X_i : i \in \mathbb{Q}]^{\geq 0}$ (as indicated in Remark 32) provides $\mathcal{J}$ such that $\mathbb{Z}[X_i : i \in \mathbb{Q}]^{\geq 0} \trianglerighteq^{\mathcal{J}} \mathbb{N}$, and there is an $\mathbb{N}$-definable isomorphism $i : \mathbb{N} \to \mathbb{N}^{\mathcal{J} \circ \mathcal{I}}$. □

Remark 36. We conjecture that the method of proof of Theorem can be adapted to show the failure of solidity of the much stronger theory $\mathsf{IOpen} + \mathsf{Collection}$, where IOpen is the fragment of PA in which the induction scheme is limited to *open formulae* (i.e., quantifier-free formulae). Note that $\mathsf{IOpen} + \mathsf{Collection}$ is a proper subtheory of PA. Our conjecture is based on the following two well-known facts:

(1) There is a countable nonstandard model $\mathcal{M}_0$ of IOpen that has no nonstandard primes; this implies, thanks to Bertrand's 'postulate', that the standard cut is definable in $\mathcal{M}_0$ by the Δ_0 formula $\delta(x)$:= "There is a prime number between x and $2x$". This result is due to Shepherdson [50].

(2) As noted by Marker [38] every model of IOpen has an end extension to a model of IOpen; the basic ingredients of Marker's construction are found in Shepherdson's paper [50]; the end extension is built using the technology of Puiseux series/polynomials.

[18] Indeed, by a classical result of Keisler, for countable models, the converse also holds, i.e., if a countable linearly ordered model $\mathcal{M}$ (in a countable language) satisfies the collection scheme, then $\mathcal{M}$ has an elementary end extension.

[19] Perhaps the proof of this theorem can be fine-tuned so as to demonstrate the failure of tightness of $\mathsf{PA}^- + \mathsf{Collection}$, but we have not yet been able to formulate such a fine-tuning.

More specifically, since the axioms of IOpen are of the form $\forall\exists$, they are preserved under unions of chains, so (2) can be used to build a continuous end-extension chain of countable model $\langle \mathcal{M}_\alpha : \alpha < \omega_1 \rangle$, where $\mathcal{M}_0$ is as in (1). This makes it clear that the union $\mathcal{M}_{\omega_1}$ of the chain satisfies $\mathsf{IOpen} + \mathsf{Coll}$ and end extends $\mathcal{M}_0$. It is evident that $\mathcal{M}_{\omega_1}$ is a model of IOpen in which the standard cut is definable, thus $\mathcal{M}_{\omega_1}$ is not a model of PA. By a variation of this argument, we can build an arithmetically definable model $\mathcal{M}(I)$ of IOpen for any arithmetically definable linear order $(I, <_I)$ such that $\mathcal{M}(\mathbb{Q}^-)$ is elementarily end extended by $\mathcal{M}(\mathbb{Q})$, and thus $\mathcal{M}(\mathbb{Q})$ satisfies the collection scheme. Then with an adaptation of the proof strategy of Theorem 35 one can verify the failure of solidity of $\mathsf{IOpen} + \mathsf{Collection}$. A detailed description of $\mathcal{M}(\mathbb{Q})$ is presented in [16].

Remark 37. In contrast with Theorem 35 and Remark 36, $\mathrm{I}\Delta_0 + \mathsf{Collection}$ is a solid theory since it is well-known (see [30]) that it is deductively equivalent to PA, which is solid by Theorem 10.

We close this subsection by showing that Theorem 25 can be extended to other theories whose solidity was asserted in Theorem 14.

Theorem 38. *No finitely axiomatizable subtheory of T is tight, where T is any of the theories in the list $\mathcal{S}$ of Theorem 14.*

Below in Theorem 39 we present a stronger version of the above result for $T = \mathsf{ZF}$; our method of proof together with Theorem 5 make it clear that similar results can be given for each of the theories T whose solidity was asserted in Theorem 14. What allows us to extend the argument is that the analogue of Theorem 5 holds for Z_α and KM_α for each $\alpha \leq \omega$, as shown in [37].

Our method of proof is similar to the proof of Theorem 24. In what follows we use the notation Δ_0, Σ_n, and Π_n in the context of the *Levy* hierarchy of set-theoretical formulae; and we use ZF_{Π_n} to denote $\{\varphi \in \Pi_n : \mathsf{ZF} \vdash \varphi\}$.

Theorem 39. *If ZF is consistent, then for each $n \in \omega$ the following hold:*

(*a*) ZF_{Π_n} *is not solid.*

(*b*) ZF_{Π_n} *is not tight.*

Proof. Assuming the consistency of ZF, by Gödel's second incompleteness theorem, and Gödel's relative consistency proof of $\mathsf{ZF} + \mathsf{V} = \mathsf{L}$, $\mathsf{ZF} + \mathsf{V} = \mathsf{L} + \neg\mathsf{Con}_{\mathsf{ZF}}$ is consistent, and therefore by the arithmetized completeness theorem there is a countable model $\mathcal{M}$ such that:

(1) $\mathcal{M} \models \mathsf{ZF} + \mathsf{V} = \mathsf{L} + \neg\mathsf{Con}_{\mathsf{ZF}}$, and

(2) The elementary diagram of $\mathcal{M}$ is definable in $\mathbb{N}$; in particular, $\mathbb{N} \trianglerighteq \mathcal{M}$.

As in the arithmetical case, given $n \in \omega$, let $K_n(\mathcal{M})$ be the submodel of $\mathcal{M}$ whose universe consists of elements of $\mathcal{M}$ that are definable in $\mathcal{M}$ by a Σ_n-formula.

The lemma below is the analogue of Theorem 22. Notice that there is a 'lag' in relation to the arithmetical case in the amount of theory-inheritance of $K_n(\mathcal{M})$ from the parent-structure $\mathcal{M}$; this is due to the higher complexity of the global well-ordering at work in set theory (under the assumption that $\mathsf{V} = \mathsf{L}$), as we shall see. Also notice that in part (b) there is no nonstandardness assumption. In part (b) of Lemma 40, $\mathsf{Coll}(\Sigma_n)$ is the collection scheme for Σ_n-formulae of set theory, whose instances are of the form:

$$(\forall x \in a\ \exists y\ \psi(x, y)) \to (\exists b\ \forall x \in a\ \exists y \in b\ \psi(x, y)),$$

where ψ is a Σ_n-formula whose parameters are suppressed.

Lemma 40. *Suppose $\mathcal{M} \models \mathsf{ZF} + \mathsf{V} = \mathsf{L}$, and $n \in \omega$ with $n \geq 2$.*

(*a*) $K_n(\mathcal{M}) \prec_{\Pi_{n-1}} \mathcal{M}$, *hence* $K_n(\mathcal{M}) \models \mathrm{Th}_{\Pi_n}(\mathcal{M})$.

(*b*) $K_n(\mathcal{M}) \models \mathsf{ZF}_{\Pi_n} + \neg\mathsf{Coll}(\Sigma_{n+1})$.

Proof of Lemma 40. The proof of part (a) is similar to the proof of part (a) of Theorem 22 except for one important difference: in the arithmetical case, the well-ordering $<$ of the universe has a Δ_0-graph, but within $\mathsf{ZF} + \mathsf{V} = \mathsf{L}$ the canonical well-order $<_\mathsf{L}$ of the universe has a Σ_1-graph (see Lemma 13.19 of [27]). We present the proof of (a) for $n = 2$, the general case is handled similarly.

By the Tarski-Vaught test, to show that $K_2(\mathcal{M}) \prec_{\Pi_1} \mathcal{M}$ it suffices to verify that if for some $a \in K_2(\mathcal{M})$ we have $\mathcal{M} \models \exists x \sigma(x, a)$ for some Σ_1-formula $\sigma(x, y)$, then there is some $b \in K_2(\mathcal{M})$ such that $\mathcal{M} \models \sigma(b, a)$. Towards this goal, assume that for some $a \in K_2(\mathcal{M})$ we have:

$$\mathcal{M} \models \exists x\ \exists v\ \delta(x, v, a),$$

where $\delta(x, v, a)$ is Δ_0. Thanks to the availability of a Δ_0-definable ordered pairing function $\langle\cdot,\cdot\rangle$ we can re-write the above as:

$$\mathcal{M} \models \exists w\, \delta((w)_0, (w)_1, a),$$

where $w = \langle (w)_0, (w)_1\rangle$. In the next step we minimize the w in (3) using the Σ_1-definable global well-ordering $<_{\mathsf{L}}$ by choosing c in $\mathcal{M}$ such that:

$$\mathcal{M} \models \delta((c)_0, (c)_1, a) \wedge \forall z\, (z <_{\mathsf{L}} c \to \neg\delta((c)_0, (c)_1, a))\,.$$

An inspection of the above makes it clear c has a Π_1-definition $\pi(y, a)$ (with parameter a) in $\mathcal{M}$. Since the element a is Σ_2-definable in $\mathcal{M}$ by assumption, let $\psi(y)$ be a Σ_2-formula defining a in $\mathcal{M}$. It suffices to show that by eliminating a from $\pi(y, a)$ with the help of $\psi(y)$ we reach a Σ_2-definition of c in $\mathcal{M}$. Towards this aim, consider the formula:

$$\theta(x) := \exists y(\pi(x, y) \wedge \psi(y)).$$

Clearly c is the unique element in $\mathcal{M}$ that satisfies $\theta(x)$. The formula $\theta(x)$ is of the form $\exists(\Sigma_2 \vee \Sigma_1)$, so it can be put in the form $\exists(\Sigma_2)$, which makes it clear that θ is Σ_2. This concludes the proof of part (a).

We now turn to the proof of (b). It is clear from part (a) that $K_n(\mathcal{M}) \models \mathsf{ZF}_{\Pi_{n-1}}$ since $\mathcal{M}$ is assumed to be a model of ZF. So we focus on the failure of Σ_n-Collection in $K_n(\mathcal{M}) \models \mathsf{ZF}_{\Pi_{n-1}}$. But first we need to address the question of definability of Σ_k-satisfaction predicate within *fragments* of ZF (it has been known since Levy's pioneering work that within full ZF there is a Σ_k-satisfaction class that is Σ_k-definable for each $k \geq 1$).

As pointed out in Theorem 2.9 of [17] (the statement of which involves KP + the axiom of infinity, but the axiom of infinity is not used in the proof) within KP there is a Σ_k-satisfaction class that is Σ_k-definable for each $k \geq 1$.[20] Here KP is Kripke-Platek set theory; a frugal theory of sets whose axioms do not include the axiom of infinity. In our formulation (following recent practice, led by Mathias [39]) the scheme of foundation is limited to Π_1-formulae (equivalently: the scheme of $\in$-induction for Σ_1-formulae). Thus in contrast to Barwise's KP in [2], which includes the full scheme of foundation, our version of KP is *finitely axiomatizable*; and a straightforward calculation shows that it is axiomatizable by a Π_2-sentence.

With the above preliminaries in place, for $n \geq 2$, $K_n(\mathcal{M}) \models \mathsf{KP}$, thanks to Π_2-axiomatizability of KP and part (a) of Lemma 40. Therefore there is a Σ_n-satisfaction predicate in $K_n(\mathcal{M})$ for Σ_n-formulae. Consider the function f that maps (the code of) each Σ_n-formula in $K_n(\mathcal{M})$ to 0, if there is no element that satisfies $\sigma(x)$, and otherwise to the least $<_{\mathsf{L}}$-element satisfying $\sigma(x)$. Thanks to the availability of a Σ_n-definable satisfaction predicate in $K_n(\mathcal{M})$ for Σ_n-formula, the graph of f is readily seen to be defined by a Σ_{n+1}-formula. Note that the domain of f is a set in $K_n(\mathcal{M})$ but its range is the whole of $K_n(\mathcal{M})$. Thus $\mathsf{Coll}(\Sigma_{n+1})$-fails in $K_n(\mathcal{M})$. This concludes the proof of part (b) Lemma 40. □

Lemma 41. *The standard cut ω is definable in $K_n(\mathcal{M})$ for each $n \in \omega$ such that $n \geq 2$.*

Proof. Recall that $K_n(\mathcal{M})$ is not ω-standard since it satisfies $\neg\mathsf{Con}_{\mathsf{ZF}}$. So, similar to the proof of Theorem 23, i is a nonstandard member of $\omega^{K_n(\mathcal{M})}$ iff every element of $K_n(\mathcal{M})$ is definable by a Σ_n-formula below i. This is first order expressible, thanks to the definability of a Σ_n-satisfaction predicate in $K_n(\mathcal{M})$. □

By Lemma 40 given any $2 \leq n < k$, both in ω, $\mathbb{N}$ is bi-interpretable with a model of $\mathsf{ZF}_{\Pi_n} + \neg\mathsf{Coll}(\Sigma_{n+1})$, and also $\mathbb{N}$ is bi-interpretable with a model of ZF_{Π_k}. Since $\mathsf{Coll}(\Sigma_n)$ is of complexity at most Π_{n+3}, this shows that by choosing $k \geq n+3$, $\mathsf{Coll}(\Sigma_n) \in \mathsf{ZF}_{\Pi_k}$, which in turn makes it clear that ZF_{Π_n} is not solid.

To see that ZF_{Π_n} is not tight it suffices to note that the above argument works uniformly; thus there are distinct deductively closed extension of ZF_{Π_n} that are bi-interpretable. This concludes the proof of Theorem 39. □

4 Categoricity of schemes

In this section we investigate various categoricity-like notions within the framework of first order logic that pertain specifically to *scheme-templates* (as opposed to arbitrary first order theories). These notions include adaptations of the notions introduced in Definition 7, but there is one which has no counterpart in Definition 7, namely, internal categoricity. Internal categoricity has been widely present in philosophical discussions in the context PA and ZF but to our knowledge it has not been previously probed in the context of general first-order schematic theories. In Subsection 4.1 we focus on internal categoricity, and in Subsection 4.2 we investigate various generalizations of the notions introduced in Definition 7.

[20] The existence of definable partial satisfaction classes in KP follows from two facts: (1) KP can prove that every set is contained in a transitive set; and (2) KP can define the satisfaction predicate for all of its internal set structures. The proofs of both of these facts can be found in Barwise's monograph [2]; the proofs therein make it clear that only Π_1-Foundation is invoked.

Definition 42. Assume $\mathcal{L}$ is a finite language. In what follows all languages are allowed to be n-sorted for some $n \in \omega$.

(**a**) An $\mathcal{L}$-*template* (for a scheme) is given by an $\mathcal{L}$-sentence $\tau(P)$ formulated in the language obtained by augmenting $\mathcal{L}$ with an n-ary predicate $P(x_1, ..., x_n)$ for some $n \in \omega$, where $x_1, \ldots x_n$ belong to the same sort.[21] Given a language $\mathcal{L}^+ \supseteq \mathcal{L}$, an $\mathcal{L}^+$-sentence ψ is said to be *an instance of* τ if ψ is of the form $\forall v \ \tau[\varphi(x_1, .., x_n, v)/P]$, where $\tau[\varphi(x_1, .., x_n, v)/P]$ is the result of substituting all subformulae of the form $P(t_1, ..., t_n)$, where each t_i is a term, with $\varphi(t_1, t_2, \cdots, t_n, v)$ (and re-naming bound variables of φ to avoid unintended clashes).

By calling τ an $\mathcal{L}$-template we shall mean that $\mathcal{L}$ is the smallest language (w.r.t. to the inclusion) which augmented with P contains τ.

(**b**) Suppose $\mathcal{L}^+ \supseteq \mathcal{L}$. We use $\tau[\mathcal{L}^+]$ to denote the collection of all $\mathcal{L}^+$-formulae that are instances of τ. We say that an T is *axiomatized by a scheme* if $T = \tau[\mathcal{L}^+]$ for some scheme τ and some $\mathcal{L}^+$.

(**c**) For familiar schematic theories such as PA and ZF the notation $\mathsf{PA}(\mathcal{L}^+)$ and $\mathsf{ZF}(\mathcal{L}^+)$ is often used in the literature to respectively denote $\tau_{\mathsf{Ind}}[\mathcal{L}^+]$ and $\tau_{\mathsf{Repl}}[\mathcal{L}^+]$.[22] Here τ_{Ind} is the template that is conjunction of two sentences, the first conjunct of which in turn is the conjunction of the finitely many axioms of (Robinson's) Q, and the second conjunct of which is the sentence expressing "If P includes 0 and is closed under successors, then P is everything". Similarly, τ_{Repl} is the template that is conjunction of two sentences, the first of which asserts that Emptyset, Pairs, Union, Infinity, and Powerset hold, and the second conjunct of which is the sentence expressing "If P is a class function whose domain is a set, then its range is also a set".

Remark 43. Let $\mathcal{L}_{\mathsf{Arith}}$ denote the language of arithmetic. As shown in [15], the complexity of the collection of (Gödel numbers) of $\mathcal{L}_{\mathsf{Arith}}$-templates τ such that $\tau[\mathcal{L}_{\mathsf{Arith}}] = \mathsf{PA}$ is a complete Π_2-set in the arithmetical hierarchy, and in particular it is not recursively enumerable.

4.1 Internal Categoricity

The notion of internal categoricity was initially explored in the context of second order logic (see the recent monograph [36] by Maddy and Väänänen for references). The definition below of internal categoricity generalizes Väänänen's formulation, who focused on the internal categoricity of the first order formulations of Peano Arithmetic and Zermelo-Fraenkel set theory, as summarized in [55]. More specifically, the notion of internal categoricity encapsulated in the definition below coincides with the one formulated by Väänänen in the setting of the usual schematic formulations of PA and ZF, so we recommend the readers unfamiliar with the notion of internal categoricity to consult [55] for the motivation of the definition below.

Definition 44. Suppose $\tau(P)$ is an $\mathcal{L}$-template.

(**a**) Let $\mathcal{L}^{\text{red}}$ and $\mathcal{L}^{\text{blue}}$ be two fresh disjoint copies of the language $\mathcal{L}$ and R and B be two fresh unary predicates. Let σ_r and σ_b be the trivial direct translations of $\mathcal{L}$ into $\mathcal{L}^{\text{red}}$ and $\mathcal{L}^{\text{blue}}$ respectively and let σ_r^{rel} (σ_b^{rel}) be the extension of σ_r (σ_b, resp.) which relativize the quantifiers to R (B, resp.). The result of applying σ_r (σ_b, σ_r^{rel}, σ_b^{rel}) to τ will be denoted τ^{red} (τ^{blue}, $\tau^{\text{red}}_{\text{rel}}$, $\tau^{\text{blue}}_{\text{rel}}$, respectively). We assume that each of the translations acts trivially on P.

(**b**) Let $\mathcal{L}^{\text{duo}} = \mathcal{L}^{\text{red}} \cup \mathcal{L}^{\text{blue}}$ and let

$$\tau^{\text{duo}} = \tau^{\text{red}}(P) \wedge \tau^{\text{blue}}(P).$$

$\tau(P)$ is said to be *internally categorical* if there is an $\mathcal{L}^{\text{duo}}$-binary formula $\alpha(x, y)$ such that, provably in $\tau^{\text{duo}}[\mathcal{L}^{\text{duo}}]$, α defines an isomorphism between the $\mathcal{L}^{\text{red}}$-reduct and the $\mathcal{L}^{\text{blue}}$-reduct of the universe, i.e., for all models $\mathcal{M}$ of $\tau^{\text{duo}}[\mathcal{L}^{\text{duo}}]$, α defines an isomorphism between $\mathcal{M}^{\text{red}}$ (the $\mathcal{L}^{\text{red}}$-reduct of $\mathcal{M}$) and $\mathcal{M}^{\text{blue}}$ (the $\mathcal{L}^{\text{blue}}$-reduct of $\mathcal{M}$).[23]

(**c**) Let $\mathcal{L}^{\text{duo+}} = \mathcal{L} \cup \mathcal{L}^{\text{blue}} \cup \mathcal{L}^{\text{red}} \cup \{R, B\}$ and let

$$\tau^{\text{duo+}} := \tau[\mathcal{L}] \cup \tau^{\text{red}}_{\text{rel}}[\mathcal{L}^{\text{duo+}}] \cup \tau^{\text{blue}}_{\text{rel}}[\mathcal{L}^{\text{duo+}}].$$

Thus in contrast with $\mathcal{L}^{\text{duo}}$-structures $\mathcal{M}$, the domain of discourse of the 'colored' submodels are allowed to be proper subsets of the domain of discourse of an $\mathcal{L}^{\text{duo+}}$ structure $\mathcal{M}$. τ is said to be *strongly internally categorical* iff there is an $\mathcal{L}^{\text{duo+}}$-binary formula $\alpha(x, y)$ such that, provably in $T^{\text{duo+}}$, α defines an isomorphism between the $\mathcal{L}^{\text{red}}$ model with the universe given by R (called *the red model*) and $\mathcal{L}^{\text{blue}}$ model with the universe given by B (*the blue model*). From the model-theoretic perspective, this is equivalent to α defining an isomorphism between $\sigma_r^{\text{rel}}(\mathcal{M})$ (the red model) and $\sigma_b^{\text{rel}}(\mathcal{M})$ (the blue model) for an arbitrary $\mathcal{L}^{\text{duo+}}$ model $\mathcal{M} \models T^{\text{duo+}}$.

[21] The theories we study all have access to a pairing function and therefore by elementary coding we can limit ourselves to a single unary predicate P, but for the purposes of exposition we are not insisting on P being unary. Note that notion of a schematic axiomatization presented here is not affected in our context where a pairing function is available if the template τ is allowed to use finitely many predicate symbols $P_1, \cdots, P_k$ of various finite arities.

[22] The notation PA^* is used in the model theory of arithmetic for $\mathsf{PA}(\mathcal{L})$, where $\mathcal{L}$ is clear from the context, or unimportant.

[23] By Compactness, this is equivalent to: For all models $\mathcal{M}$ of $\tau^{\text{duo}}[\mathcal{L}^{\text{duo}}]$, there is an $\mathcal{M}$-definable isomorphism between $\mathcal{M}^{\text{red}}$ and $\mathcal{M}^{\text{blue}}$.

Example 45. Let τ be τ_{Repl}, $\mathcal{L}$ be the language of set theory, and κ_1, κ_2, and κ_3 be strongly inaccessible cardinals, with $\kappa_1 > \kappa_2 > \kappa_3$. Then we can expand $(V_{\kappa_1}, \in)$ into a model $\tau^{\text{duo}+}$ by interpreting R as V_{κ_2}, B as V_{κ_3}, and interpreting $\in$, $\in^{\text{red}}$, and $\in^{\text{blue}}$ as the membership relation in the real world.

The following proposition motivates our choice of the terminology "strong internal categoricity":

Proposition 46. *If a scheme template τ is strongly internally categorical, then τ is internally categorical.*

Proof. Assume that τ is strongly internally categorical and let α be a $\mathcal{L}^{\text{duo}+}$ binary formula which witness the isomorphism. Let α_r result from α by

- substituting each occurrence of $R(t)$ and $B(t)$ with $t = t$ and
- colouring each symbol of $\mathcal{L}$ into red (i.e. applying σ_r to the formula obtained in the first step).

We claim that α_r provably in $\tau^{\text{duo}}[\mathcal{L}^{\text{duo}}]$ defines an isomorphism between the red and the blue part of the universe. Indeed, take any model $\mathcal{M} \models \tau^{\text{duo}}[\mathcal{L}^{\text{duo}}]$. By interpreting R and B as the whole universe and interpreting $\mathcal{L}$ symbols as their red counterparts, one can see that $\mathcal{M}$ definably expands to a model $\mathcal{M}^+ \models T^{\mathsf{duo}+}$. Hence α defines in $\mathcal{M}^+$ an isomorphism between the red and the blue model. By definitions, α_r defines an isomorphism between $\mathcal{M}^{\text{red}}$ and $\mathcal{M}^{\text{blue}}$. □

The converse of the above proposition need not hold; this was implicitly observed in [36]. Part (a) of the theorem below appears without proof in [36] as Theorem 9, p. 32, where the result is stated in a different notation.

Theorem 47. (*Väänänen*) *Let τ_{Ind} and τ_{Repl} be the scheme templates axiomatizing* PA *and* ZF (*respectively*) *as in part* (*c*) *of Definition 42.*

(*a*) *τ_{Ind} is strongly internally categorical.*

(*b*) *τ_{Repl} is internally categorical, but not strongly internally categorical* (*assuming the consistency of* $\mathsf{ZF} + \exists\kappa\mathsf{Inacc}(\kappa)$).

Proof. To see that (a) holds, suppose $\mathcal{M}$ is a model of PA, and $\mathcal{N}_1$ and $\mathcal{N}_2$ are models of PA such that for $i \in \{1,2\}$ the domain of discourse N_i of $\mathcal{N}_i$ is a subset of the domain of discourse M of $\mathcal{M}$ (and therefore the graphs of addition and multiplication for $\mathcal{N}_i$ are subsets of M^2). Furthermore, assume:

(∇) For each subset D of M that is parametrically definable in $(\mathcal{M}, \mathcal{N}_1, \mathcal{N}_2)$, and for each $i \in \{1,2\}$, $(\mathcal{N}_i, D \cap N_i) \models \mathsf{PA}(X)$, where X is interpreted by $D \cap N_i$.

Let $\varphi(x, v_1, v_2)$ be the formula that expresses that $v_1 \in N_1$, $v_2 \in N_2$ and x codes an isomorphism between the initial segment of $\mathcal{N}_1$ determined by v_1 and the initial segment of $\mathcal{N}_2$ determined by v_2 (when addition and multiplication are viewed as ternary relations).

Let

$$I_1 = \{a \in N_1 : (\mathcal{M}, \mathcal{N}_1, \mathcal{N}_2) \models \exists x \exists v_2\, \varphi(x, a, v_2)\},$$

and

$$I_2 = \{b \in N_2 : (\mathcal{M}, \mathcal{N}_1, \mathcal{N}_2) \models \exists x \exists v_1\, \varphi(x, v_1, b)\}.$$

Using the assumptions on $\mathcal{N}_1$ and $\mathcal{N}_2$ it is easy to see that I_1 is an initial segment of N_1 that contains $0^{\mathcal{N}_1}$ and is closed under the successor relation of $\mathcal{N}_1$, and therefore by (∇) $I_1 = N_1$; and I_2 is an initial segment of N_2 that contains $0^{\mathcal{N}_2}$ and is closed under the successor relation of $\mathcal{N}_2$ and therefore $I_2 = N_2$. It can also be readily shown using (∇) that for every a in $\mathcal{N}_1$ there is a unique b in $\mathcal{N}_2$ such that $(\mathcal{M}, \mathcal{N}_1, \mathcal{N}_2) \models \exists x \varphi(x, a, b)$. This makes it clear that there is an $(\mathcal{M}, \mathcal{N}_1, \mathcal{N}_2)$-definable isomorphism between $\mathcal{N}_1$ and $\mathcal{N}_2$.[24]

For (b), the proof that τ_{Repl} is internally categorical appears in [54] (note the proof is stated for the schematic axiomatization of ZFC, but the axiom of choice can be dispensed with). Let $\mathcal{M}$ be a model of ZF and let $\kappa \in M$ be a strongly inaccessible cardinal (we do not assume that $\mathcal{M}$ is well-founded). We expand $\mathcal{M}$ to a model of $\mathcal{M}^+$ by interpreting

- $\in, \in^{\text{red}}, \in^{\text{blue}}$ as $\in^{\mathcal{M}}$,
- R as the whole universe and
- B as $(V_\kappa)^{\mathcal{M}}$.

[24] It is worth pointing out that the proof makes it clear that the assumption $\mathcal{M} \models \mathsf{PA}$ can be substantially reduced to the assumption that the theory of $\mathcal{M}$ is sequential. Thus in light of the sequentiality of PA^- (established in [28]) it is sufficient to assume that $\mathcal{M} \models \mathsf{PA}^-$.

Since in $\mathcal{M}^+$ the red model is just the whole universe, then clearly, the red part satisfies the replacement scheme. That for every parametrically definable subset D of $\mathcal{M}$ the expansion $(V_\kappa^{\mathcal{M}}, \in^{\mathcal{M}}, D \cap V_\kappa^{\mathcal{M}})$ of the blue model satisfies the replacement scheme in the extended language follows from basic properties of strongly inaccessible cardinals. □

Remark 48. The direct proof of tightness of ZF (as presented in [21]) is similar to proof of internal categoricity of τ_{Repl} (as presented in [54]), and they can be viewed as refinements of Zermelo's quasi-categoricity theorem for ZF. However, the proof of solidity of ZF in [12] involves a step that is not present in the proofs of tightness/internal categoricity of ZF, namely, an appeal to Tarski's undefinability of truth theorem.

Remark 49. Hamkins and Freire [21] showed that two distinguished schematic subtheories of ZF fail to be internally categorical, namely Zermelo set theory (axiomatized by the separation scheme), and ZF$^-$ (axiomatized by the schemes of separation and collection). In the same paper, Hamkins and Freire showed that internal categoricity of ZF implies the tightness of ZF using a soft argument (an easier similar soft argument gives the corresponding result for PA); see Theorem 16 of [21], and the paragraph following it. The proof of the implication employs a class version of the Schröder-Bernstein theorem, plus the fact that ZF has a property known in model theory as *eliminating imaginaries* (cf. Section 4.4 of [26]), i.e., if E is a definable equivalence relation on the universe of discourse, then there is a definable function f that maps distinct E-equivalence classes to distinct objects (in ZF such a function can be readily described by what is commonly known as the 'Scott trick', which takes advantage of the stratification of the universe into the well-ordered family of sets of the form V_α, as α ranges over the ordinals; see Exercise 2 of Section 4.4 of [26]). In light of the fact that the class version of Schröder-Bernstein is provable in all sequential theories as shown by Friedman and Visser [23], the proof strategy of Hamkins and Freire can be used to show the following general proposition.

Proposition 50. *For sequential theories T that eliminate imaginaries, the internal categoricity of T implies the tightness of T.*

Remark 51. There is an interesting conceptual difference between internal categoricity and the notions from the previous section (tightness, solidity), namely: the formulation of the latter ones do not depend on T being a schematic theory, but internal categoricity applies primarily to schemes. More explicitly: given a scheme τ the theory $\tau^{\mathrm{duo}}[\mathcal{L}^{\mathrm{duo}}]$ is not simply the union of $\tau^{\mathrm{red}}[\mathcal{L}^{\mathrm{red}}]$ and $\tau^{\mathrm{blue}}[\mathcal{L}^{\mathrm{blue}}]$. Additionally, there seems to be no "natural" definition of such a doubling of a theory T, if T is not readily presented through a scheme. A way to bypass this issue would be to show that actually for any two schemes τ and τ' and a language $\mathcal{L}$, if $\tau[\mathcal{L}]$ and $\tau'[\mathcal{L}]$ are deductively equivalent (axiomatize the same theory), then τ is internally categorical iff τ' is internally categorical. The next theorem shows that, actually, the exact opposite of the above claim is true: every sufficiently interesting theory of arithmetic is axiomatizable via a scheme that is *not* internally categorical. In particular, even though the induction scheme template is internally categorical, there is a scheme template that can be used to axiomatize the same first-order theory (i.e. Peano Arithmetic) that is not internally categorical. Moreover, we cannot escape this conclusion by looking at "sufficiently strong" theories.

Theorem 52. *Let T be any consistent recursively enumerable extension of* PA (*in the same language*) *or of* ZF (*in the same language*). *Then there is a schematic axiomatization of T that is not internally categorical.*

Proof. We shall first explain the proof when T is an extension of PA. Fix T and let σ be a formula that strongly represents a primitive recursive axiomatization of T (such a σ exists by Craig's trick). Let $\mathsf{Sat}(P, x)$ be the usual formula that expresses that P is a satisfaction predicate for formulae of logical depth at most x. More explicitly, $\mathsf{Sat}(P, x)$ asserts:

(1) Members of P are ordered pairs $\langle \varphi, \alpha \rangle$, where φ is a formula of depth[25] at most x, and α is an assignment for φ; and

(2) P satisfies Tarski's clauses for a satisfaction predicate for all formulae of depth at most x.

Let $\mathsf{EA} = \mathsf{I}\Delta_0 + \mathsf{Exp}$; by a theorem of Wilkie, EA is finitely axiomatizable (see [25] for an exposition). Consider the following scheme template $\tau_{\mathsf{Vaught},\sigma}(P)$:

$$\mathsf{EA} + \forall x\big(\mathsf{Sat}(P, x) \rightarrow \forall \varphi\big(\mathsf{pdepth}(\varphi) \leq x \wedge \sigma(\varphi) \rightarrow P(\varphi, \varnothing)\big)\big).$$

In the above $\mathsf{pdepth}(\varphi) \leq x$ expresses that the pure depth[26] φ is at most x and $\varnothing$ is the empty assignment. As shown in Theorem 12 of [15] $\tau_{\mathsf{Vaught},\sigma}$ axiomatizes T; the proof therein is a streamlined version of Vaught's original proof in [57].

[25] The depth of a formula φ is the length of the longest chain in the formation tree of φ (which allows for arbitrary atomic formulae as leaves).

[26] The pure depth of a formula takes into account also the complexity of terms occurring in it. More precisely: the pure depth of an atomic formula is the maximal complexity of terms occurring in it. The pure depth of compound formulae is then defined recursively in the standard way. See [15] for details.

To see that $\tau_{\mathsf{Vaught},\sigma}$ is not internally categorical take any two non-isomorphic nonstandard models $\mathcal{M}$ and $\mathcal{N}$ of T such that neither $\mathcal{M}$ nor $\mathcal{N}$ is recursively saturated. For example, using the assumptions on T, we can invoke the first and second incompleteness theorems to get hold of two distinct completions T_1 and T_2 of $T+\neg\mathsf{Con}_T$. Note that T_1 and T_2 have the property that all of their models are nonstandard. We can choose $\mathcal{M}$ to be a pointwise definable model of T_1, and choose $\mathcal{N}$ to be a pointwise definable model of T_2. Since $\mathcal{M}$ and $\mathcal{N}$ are both countable, without loss of generality assume that the universes of $\mathcal{M}$ and $\mathcal{N}$ are the same. Let $\mathcal{K}$ be the model with the same universe as that of $\mathcal{M}$ and which carries the interpretations from both $\mathcal{M}$ and $\mathcal{K}$; thus the red model of $\mathcal{K}$ is isomorphic to $\mathcal{M}$, and the blue model of $\mathcal{K}$ is isomorphic to $\mathcal{N}$. As previously, we shall refer to symbols of T as interpreted in $\mathcal{M}$ ($\mathcal{N}$) as the *red* (*blue*) ones. We claim that $\mathcal{K} \models \tau^{\mathrm{duo}}[\mathcal{L}^{\mathrm{duo}}]$. To see this is it is enough to argue that for any formula $\Psi(x,y) \in \mathcal{L}^{\mathrm{duo}}$ we have:

$$\mathcal{K} \models \forall x\big(\mathsf{Sat}(\Psi,x) \to \forall\varphi\big(\mathsf{pdepth}(\varphi) \leq x \wedge \sigma(\varphi) \to \Psi(\varphi,\varnothing)\big)\big),$$

where all basic symbols which appear outside of Ψ belong to $\mathcal{L}^{\mathrm{red}}$ (the same argument applies in the case of $\mathcal{L}^{\mathrm{blue}}$).

Fix Ψ and work in $\mathcal{K}$. Fix x, and assume that $\mathsf{Sat}(\Psi,x)$. We claim that x is a standard number (according to $\leq^{\mathrm{red}}$). If not, then the satisfaction class S defined by Ψ in $\mathcal{K}$ is a partial nonstandard satisfaction class on $\mathcal{M}$, and therefore by Lachlan's theorem (see 15.5 of [30]) $\mathcal{M}$ must be recursively saturated, contradicting our choice of $\mathcal{M}$. So take an arbitrary φ of pure depth at most x. It follows that φ corresponds to a standard formula (of $\mathcal{L}^{\mathrm{red}}$). Assuming additionally that $\sigma(\varphi)$ holds, we conclude that φ is an axiom of T^{red}. So φ must hold in $\mathcal{M}$. Consequently $\Psi(\varphi,\varnothing)$ holds because Ψ satisfies the Tarski T-scheme for sentences of $\mathcal{L}^{\mathrm{red}}$.

The above proof strategy can be modified to work for extensions T of ZF. More specifically, the definition of the template $\tau_{\mathsf{Vaught},\sigma}(P)$ is modified by replacing EA with KP.[27] Since it is well-known that Lachlan's theorem also holds for ω-nonstandard models of ZF, to complete the proof we only need to explain how to get to hold of two countable nonisomorphic ω-nonstandard models of T that are not recursively saturated. As in the arithmetical case we can use the first and second incompleteness theorems to get hold of two distinct completions T_1 and T_2 of $T+\neg\mathsf{Con}_T$. Note that T_1 and T_2 have the property that all of their models are ω-nonstandard. Then we can choose $\mathcal{M}$ to be a Paris model of T_1 and $\mathcal{N}$ to be a Paris model of T_2 (A Paris model is a model all of whose ordinals are pointwise definable, therefore a Paris model is not recursively saturated). Since every consistent extension of ZF has a Paris model by a classical theorem of Paris (see [11]) this concludes the proof. □

Remark 53. Since, as shown in [28], the canonical theory of (the non-negative parts) of the discretely ordered rings, i.e. PA^- (see [30]), is sequential, the above proof works also for an arguably more natural axiomatization of PA which results from $\tau_{\mathsf{Vaught},\sigma}(P)$ by replacing EA with PA^-.

Proposition below complements Theorem 52.

Proposition 54. *Every recursively enumerable extension T of* PA *or of* ZF *can be axiomatized by a scheme whose template is internally categorical.*

Proof. Use Craig's trick to get hold of σ that strongly represents a primitive recursive axiomatization of T. Thanks to Vaught's Theorem [57] and Theorem 47, if T is an extension of PA, then the desired template is $\tau_{\mathsf{Vaught},\sigma}(P) \wedge \tau_{\mathsf{Ind}}(P)$; and if T is an extension of ZF, then the desired template is $\tau_{\mathsf{Vaught},\sigma}(P) \wedge \tau_{\mathsf{Repl}}(P)$. □

At first sight Theorem 52 and Proposition 54 might seem to diminish the importance of the notion of internal categoricity, because there is a strong intuition that a foundationally interesting property of theories shouldn't depend on how the theory is presented. However, in what follows we experiment with a different option: we check whether schemes can be viewed as a self-standing and, in a sense, primitive object of foundational analysis. In the rest of this section we scrutinize mathematical benefits from taking this point of view: we show that this perspective enables us to generalize and make explicit some patterns that can be observed in the hierarchy of solid/tight theories.

Furthermore, there are natural adaptations of the notion of solidity to the context of schemes, which bring to light interesting distinctions between various natural foundational schemes, such as the scheme of induction and the replacement scheme. We shall discuss the philosophical consequences of taking this perspective in the next section.

Definition 55. The scheme template $\tau^{\mathsf{Arith}}_{\mathsf{Coll}}$ is the *arithmetical sentence* that is conjunction of two sentences; the first conjunct asserts that EA holds, and the second conjunct is the following sentence:

$$\forall x\big(\forall y < x\ \exists z\ P(y,z) \to \exists b\ \forall y < x\ \exists z < b\ P(y,z)\big).$$

Similarly, The scheme template $\tau^{\mathsf{Set}}_{\mathsf{Coll}}$ is the *set-theoretical sentence* that is conjunction of two sentences; the first conjunct asserts that KP holds, and the second conjunct is the following sentence:

$$\forall x\big(\forall y{\in}x\ \exists z\ P(y,z) \to \exists b\ \forall y \in x\ \exists z \in b\ P(y,z)\big).$$

[27] Recall that KP is Kripke-Platek set theory with Π_1-Foundation (as in [39]); its well-known that KP is finitely axiomatizable.

Theorem 56. *The scheme templates $\tau^{\mathsf{Arith}}_{\mathsf{Coll}}$ and $\tau^{\mathsf{Set}}_{\mathsf{Coll}}$ are not internally categorical, even though they respectively axiomatize* PA *and* ZF*.*

Proof. We already commented in Remark 37 PA is axiomatized by all arithmetical instances of τ_{Coll}. It is also well-known that ZF is axiomatized by the set-theoretical instances of $\tau^{\mathsf{Set}}_{\mathsf{Coll}}$ (since an induction on the complexity of formulae shows that the full separation scheme is provable in the resulting theory, thanks to the availability of Δ_0-separation and full collection).

To demonstrate the failure of internal categoricity for $\tau^{\mathsf{Arith}}_{\mathsf{Coll}}$, we resort to the fact that there are (many) non-isomorphic ω_1-like models of PA (e.g., see p.237 of [32]). Recall that a model of PA is said to be ω_1-like if it is uncountable but every proper initial segment of it is countable.

Let $\mathcal{M}$ and $\mathcal{N}$ be any two non-isomorphic ω_1-like models of PA, let $\mathcal{K}$ be the disjoint union $\mathcal{M}, \mathcal{N}$ (as in the proof of Theorem 52). Obviously $\mathcal{K} \models \mathsf{EA}^{\mathsf{red}} \wedge \mathsf{EA}^{\mathsf{blue}}$. Moreover, since each component of $\mathcal{K}$ is ω_1-like, for every $\varphi(x,y) \in \mathcal{L}^{\mathsf{duo}}$,

$$\mathcal{K} \models \tau^{\mathsf{Arith}}_{\mathsf{Coll}}[\varphi(x,y)/P].$$

The set-theoretical case is handled similarly to the arithmetical case; thanks to the well-known fact that there are (many) nonisomorphic ω_1-like models of ZF; see [10]. Here, a model of ZF is said to be ω_1-like if it is uncountable but whose every proper rank-initial segment is countable. □

Remark 57. The Vaught schemes template $\tau_{\mathsf{Vaught},\sigma}(P)$ (introduced in the proof of Theorem 52, where σ is a formula that strongly represents a primitive recursive axiomatization of PA), and the Collection scheme template $\tau^{\mathsf{Arith}}_{\mathsf{Coll}}(P)$ (introduced in Definition 55) do not give rise to Π^1_1-statements $\forall X \tau(X)$ that characterize the standard model of arithmetic $\mathbb{N}$. Clearly the standard model of arithmetic satisfies the Π^1_1-statements corresponding to $\tau_{\mathsf{Vaught},\sigma}$ and to $\tau^{\mathsf{Arith}}_{\mathsf{Coll}}$. However, the Π^1_1-statement corresponding to $\tau_{\mathsf{Vaught},\sigma}$ is also satisfied in any countable model of PA that is not recursively saturated (by the proof of Theorem 52), and the Π^1_1-statement corresponding to $\tau^{\mathsf{Arith}}_{\mathsf{Coll}}$ is satisfied in any ω_1-like model of PA.

Remark 58. It is easy to see that if $\tau(P)$ is a template in the language of arithmetic that is internally categorical and the scheme it generates is true in $\mathbb{N}$, then the Π^1_1-statement $\forall X \tau(X)$ characterizes $\mathbb{N}$ in full second order logic. However, as shown in recent (to be published) joint work of the second-named-author with Piotr Gruza, there is a template $\tau(P)$ in the language of arithmetic that is not internally categorical but which has the property that the Π^1_1-statement $\forall X \tau(X)$ characterizes $\mathbb{N}$ in full second order logic.

Next, we show that internal categoricity is preserved upwards w.r.t. to the ordering determined by provability of Π^1_1 sentences in two-sorted logic.

Definition 59. For a theory T in an n-sorted language $\mathcal{L}$, let $\mathsf{PC}(T)$ be the extension of T in the $(n+1)$-sorted language by the predicative comprehension axioms, that is all sentences of the form

$$\exists X^{n+1} \forall X^n \big(X^n \in_{n+1} X^{n+1} \equiv \varphi(X^n, \overline{Y})\big),$$

where $\varphi(X^n, \overline{Y})$ is an $\mathcal{L}$-formula (see [43]). The upper index of a variable denotes its sort.

Theorem 60. *Let τ and σ be two $\mathcal{L}$-scheme templates that $\tau[\mathcal{L}]$ and $\sigma[\mathcal{L}]$ axiomatize the same theory T. Suppose further that* $\mathsf{PC}(T) \vdash \forall X \tau[X] \to \forall X \sigma[X]$*. Then if σ is internally categorical, then so is τ.*

Proof. Take any $\mathcal{M} \models \tau^{\mathrm{red}}[\mathcal{L}^{\mathrm{duo}}] \wedge \tau^{\mathrm{blue}}[\mathcal{L}^{\mathrm{duo}}]$. Let $\mathfrak{X}$ consists of $\mathcal{L}^{\mathrm{duo}}$-definable subsets of $\mathcal{M}$. Then, it follows that for $\mathsf{col} \in \{\mathrm{red}, \mathrm{blue}\}$,

$$(\mathcal{M}, \mathfrak{X}) \models \mathsf{PC}(T^{\mathsf{col}}) \wedge \forall X \tau^{\mathsf{col}}[X].$$

In particular, by our assumption $(\mathcal{M}, \mathfrak{X}) \models \mathsf{PC}(T^{\mathsf{col}}) \wedge \forall X \sigma^{\mathsf{col}}[X]$. It follows that

$$\mathcal{M} \models \sigma^{\mathsf{col}}[\mathcal{L}^{\mathsf{duo}}],$$

hence, by the internal categoricity of σ, it follows that there is a definable isomorphism between the red and blue parts of $\mathcal{M}$. □

4.2 Generalizations of solidity

In this subsection we adapt the categoricity-like properties of theories introduced in Subsection 2.3 to the context of schemes. As we shall see in Theorem 74, these generalizations can be viewed as providing a (short) hierarchy of 'categoricity grades' of scheme templates.

Definition 61. Below all the languages are arbitrary n-sorted languages. In the paper we shall deal only with translations which are sort-preserving, i.e. types of all the variables are preserved under the translation. Moreover we shall deal only with the translations which have the same dimension on every sort. Last but not least, we assume that the isomorphisms between sorted structures need to preserve sorts.

(**a**) Let $m \leq n$. A (k-dimensional) translation σ between an m-sorted language $\mathcal{L}_1$ and an n-sorted $\mathcal{L}_2$ is given by

- a tuple of formulae $(\delta_1(\bar{x}^1), \ldots, \delta_m(\bar{x}^m))$, where each $\delta_i(\bar{x}^i)$ is a formula of the i-th order logic and all $\bar{x}^i$ are variables of the i-th sort.
- a mapping $P \mapsto A_P$, where P is an l-ary predicate of $\mathcal{L}_1$ and A_P is a $k \cdot l$-ary formula

As in the previous section we allow the equality to be redefined and our interpretations to be multidimensional. A translation σ is *direct* iff it is direct with respect to all sorts.

(**b**) Let $\tau(P)$ be an $\mathcal{L}_1$-template, where P is a predicate of arity l and σ be any k-dimensional translation $\mathcal{L}_1 \to \mathcal{L}_2$. Let R be any fresh predicate of arity kl. Let $\sigma[P/R]$ be a translation $\mathcal{L}_1 \cup \{P\} \to \mathcal{L}_2 \cup \{R\}$ which translates P to R and behaves like σ otherwise. By τ^σ we denote the following scheme template:

$$(\text{"}R \text{ is a } \approx\text{-invariant } l\text{-ary relation on } k\text{-tuples"}) \longrightarrow \tau^{\sigma[P/R]}.$$

In the above $\approx$ is the translation of equality according to σ and $\tau^{\sigma[P/R]}$ is the translation of τ according to $\sigma[P/R]$.

(**c**) We say that an $\mathcal{L}_2$-structure $\mathcal{M}$ *accommodates* a scheme template $\tau(P)$ iff there is a translation σ of $\mathcal{L}_1$ into $\mathcal{L}_2$ such that,

$$(\mathcal{M}, \mathsf{Def}^{kn}(\mathcal{M})) \models \forall X \tau^\sigma(X/R),$$

where $\mathrm{Def}^{kn}(\mathcal{M})$ denotes the set of parametrically $\mathcal{M}$-definable subsets of M^{kn}. In other words, for each $X \in \mathrm{Def}^{kn}(\mathcal{M})$, $(\mathcal{M}, X) \models \tau^\sigma(X/R)$. Note that the translation σ is allowed to use parameters from $\mathcal{M}$.

(**d**) We say that an $\mathcal{L}_2$-structure $\mathcal{M}$ *directly accommodates* a scheme template τ iff $\mathcal{M}$ accommodates τ via a direct translation σ. In such a case, $\mathcal{M}$ is also called a *strong model of* τ.

(**e**) We say that a theory T (*directly*) *accommodates* a scheme τ iff there is a translation σ such that for every model $\mathcal{M} \models T$, $\mathcal{M}$ (directly) accommodates τ via σ.

Example 62. Since in this paper we are primarily interested in sequential theories, we can always assume that for our purposes interpretations are one-dimensional. Assuming additionally that τ is equality-preserving, we can easily show that $\mathcal{M}$ accommodates an $\mathcal{L}$-scheme τ via a one-dimensional and equality-preserving translation just in case the following condition holds:

$$\text{There exists an } \mathcal{L}\text{-structure } \mathcal{N} \trianglelefteq_{\text{par}} \mathcal{M} \text{ such that for every parametrically } \mathcal{M}\text{-definable set } P \subseteq M, \\ (\mathcal{N}, P \cap N) \models \tau(P).$$

In particular, if $\mathcal{M}$ is a model of Z_2 (Second Order Arithmetic) or a model of Z (Zermelo set theory), then $\mathcal{M}$ accommodates the induction scheme τ_{Ind}. Similarly, every model of KM accommodates the replacement scheme τ_{Repl}.

Example 63. A structure $\mathcal{M}$ accommodates the induction scheme τ_{Ind}, just in case $\mathcal{M}$ parametrically interprets a model $\mathcal{N}$ of PA *with the additional property that $\mathcal{N}$ has no proper parametrically $\mathcal{M}$-definable proper cuts*. Note that under this scenario if the 'accommodating' structure $\mathcal{M} \models \mathsf{PA}$, then there is an $\mathcal{M}$-definable isomorphism between $\mathcal{M}$ and $\mathcal{N}$; this readily follows from the fact that PA is a minimalist theory (as indicated in part (a) of Remark 8).

Remark 64. We note that $\mathcal{M}$ is a strong model (in the sense of part (d) of Definition 61) if an $\mathcal{L}$-scheme τ just in case there is an extension (perhaps higher order) $\mathcal{L}^+ \supseteq \mathcal{L}$ and a definitional expansion $\mathcal{M}^+$ which is an $\mathcal{L}^+$ model, such that $\mathcal{M}^+ \models \tau[\mathcal{L}^+]$.

Definition 65. Suppose $\tau(P)$ is an $\mathcal{L}$-template.

(a) τ is *g-minimalist* (generalized minimalist) if the following holds for all strong models $\mathcal{M}$ and $\mathcal{N}$ of τ: If $\mathcal{M}$ parametrically accommodates τ via a translation σ_0, $\mathcal{N} \trianglelefteq_{\text{par}} \mathcal{M}$, and $\mathcal{N}$ accommodates τ via a translation σ_1, **then** there is a **unique** $\mathcal{M}$-definable embedding

$$F : \sigma_0(\mathcal{M}) \to \sigma_1(\mathcal{N}).$$

(b) τ is *g-solid*[28] (generalized solid) iff the following holds for all strong models $\mathcal{M}$, $\mathcal{N}$, and $\mathcal{M}^*$ of τ that respectively accommodate $\tau(P)$ via translations σ_l, σ_c and σ_r: If

$$\mathcal{M} \trianglerighteq_{\text{par}} \mathcal{N} \trianglerighteq_{\text{par}} \mathcal{M}^*,$$

and there is an $\mathcal{M}$-definable isomorphism $F : \sigma_l(\mathcal{M}) \to \sigma_r(\mathcal{M}^*)$, **then** there is an $\mathcal{N}$-definable isomorphism

$$G : \sigma_c(\mathcal{N}) \to \sigma_r(\mathcal{M}^*).$$

(c) τ is *e-solid* (expansion solid) iff τ satisfies the definition for g-solidity with the added requirement that σ_l, σ_c and σ_r are *direct* interpretations.

Remark 66. The e-solidity terminology is informed by the fact that e-solidity can be equivalently formulated in terms of the notion of *expansion* in model theory. Observe that an $\mathcal{L}$-template τ is e-solid just in case for all expansions $\mathcal{L}_1$, $\mathcal{L}_2$, $\mathcal{L}_3$ of $\mathcal{L}$, and all structures $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_3$ such that $\mathcal{M}_i \models \tau[\mathcal{L}_i]$ for $i \in \{1, 2, 3\}$, the following condition holds:

if $\mathcal{M}_1 \trianglerighteq_{\text{par}} \mathcal{M}_2 \trianglerighteq_{\text{par}} \mathcal{M}_3$ and there is an $\mathcal{M}_1$-definable isomorphism between $\mathcal{M}_1{\restriction_{\mathcal{L}}}$ and $\mathcal{M}_3{\restriction_{\mathcal{L}}}$, then there is an $\mathcal{M}_2$-definable isomorphism between $\mathcal{M}_2{\restriction_{\mathcal{L}}}$ and $\mathcal{M}_3{\restriction_{\mathcal{L}}}$.

In the above $\mathcal{M}{\restriction_{\mathcal{L}}}$ denotes the reduct of a model $\mathcal{M}$ to the language $\mathcal{L}$.

Remark 67. Note that the categoricity-like notions defined in Definition 65 are attributes of *scheme templates*, whereas those defined in Definition 7 are attributes of *first order theories*.

Definition 68. The scheme template $\tau_{\mathsf{Repl+Tarski}}(P)$ is the result of adjoining Tarski's undefinability of truth theorem to the usual axiomatization of ZF using τ_{Repl}, i.e.,

$$\tau_{\mathsf{Repl+Tarski}}(P) := \exists x \in \omega \neg \mathrm{Sat}(P, x) \wedge \tau_{\mathrm{Repl}}(P).$$

In the above, $\mathrm{Sat}(P, x)$ is the formula obtained by adding the unary predicate P to the language $\{\in\}$ of set theory that expresses that P satisfies Tarski conditions for a satisfaction predicate for all set-theoretical formulae of depth at most x.

Remark 69. $\tau_{\mathsf{Repl+Tarski}}$ ensures that replacement holds for all formulae, and no formula defines a full satisfaction predicate for $\mathcal{L} = \{\in\}$. Thanks to Tarski's undefinability of truth theorem $\tau_{\mathsf{Rep+Tarski}}$ axiomatizes ZF. However, note that if κ is a strongly inaccessible cardinal, then the natural Kelley-Morse model associated with V_κ whose classes are elements of $V_{\kappa+1}$ does not satisfy the Π^1_1-sentence $\forall X \tau_{\mathsf{Repl+Tarski}}(X)$.

Theorem 70. *The scheme-templates τ_{Ind} and $\tau_{\mathsf{Repl+Tarski}}$ are e-solid.*[29]

Proof. (Outline). A close examination of the proof of the solidity of PA presented in [12] shows that the same proof strategy succeeds in establishing the e-solidity of τ_{Ind}. The e-solidity of $\tau_{\mathsf{Repl+Tarski}}$ is established by using the proof strategy of the solidity proof of ZF presented in [12] (or the one presented in [21]) with the difference that the Tarski clause comes to the rescue at the point in the proof of solidity of ZF where Tarski's undefinability of truth is invoked. More specifically, the Tarski clause of $\tau_{\mathsf{Repl+Tarski}}$ makes sure that none of the extra predicates included in the expansion of the first of the three models of ZF (in the set-up of the e-solidity proof) is a satisfaction predicate for the first model of ZF. □

[28] Note that general solidity is a generalization of the notion of strong solidity that was shown in Remark 9 to be equivalent to solidity. The proof strategy of the equivalence does not seem to generalize to the generalized setting. We have opted for this stronger notion because (1) it is satisfied by the usual schematic representations of PA and ZF, and (2) it comes handy in the proof of solidity of KF_μ (see Theorem 88).

[29] Recall from Definition 42 that τ_{Ind} axiomatizes PA and τ_{Repl} axiomatizes ZF.

Example 71. For $T = \mathsf{PA}$, Theorem 70 can be rephrased as follows: Suppose $\mathcal{M}_1$, $\mathcal{M}_2$, and $\mathcal{M}_3 \models \mathsf{PA}$, and that $\mathcal{M}_i^+$ is an $\mathcal{L}_i$-structure that is an expansion of $\mathcal{M}_i$ and $\mathcal{M}_i^+ \models \mathsf{PA}(\mathcal{L}_i)$ for $i \in \{1, 2, 3\}$, where each $\mathcal{L}_i$ is an extension of the usual language of arithmetic. Suppose, furthermore, that $\mathcal{M}_1^+ \trianglerighteq \mathcal{M}_2^+ \trianglerighteq \mathcal{M}_3^+$ and there is an $\mathcal{M}_1^+$-definable isomorphism $i_0 : \mathcal{M}_1 \to \mathcal{M}_3$. **Then** there is an $\mathcal{M}_1^+$-definable isomorphism $i : \mathcal{M}_2 \to \mathcal{M}_3$.

Note that in the above, the existence of $i : \mathcal{M}_1 \to \mathcal{M}_2$ cannot in general be strengthened to the existence of $i : \mathcal{M}_1^+ \to \mathcal{M}_2^+$, e.g., let $\mathcal{M}_1^+ = (\mathcal{M}, D_1)$ and $\mathcal{M}_1^+ = (\mathcal{M}, D_2)$, where D_1 and D_2 are distinct $\mathcal{M}$-definable sets.

Example 72. τ_{Repl} is not e-solid. To see this, suppose κ and λ are strongly inaccessible cardinals with $\kappa < \lambda$. Let γ be an ordinal such that $\kappa \in \gamma \in \lambda$. By the Löwenheim-Skolem theorem, there is an elementary submodel $\mathcal{K} = (K, \in)$ of $(V_\lambda, \in)$ that $\gamma \in K$ and $V_\kappa \subseteq K$ and $|K| = |V_\kappa|$. Note that the order-type of the ordinals in K is higher than κ. Using a bijection g between K and V_κ there is a binary relation E on V_κ such that g is an isomorphism between (V_κ, E) and $(K, \in)$. Thus there is an embedding f of $(V_\kappa, \in)$ as a topped rank initial segment of (V_κ, E). By well-foundedness of (V_κ, E) there is some ordinal $\alpha \in V_\kappa$ such that $(V_\alpha, \in)$-as-calculated-in-(V_κ, E) is isomorphic to $(V_\kappa, \in)$; thus α is the first ordinal in (V_κ, E) that is not in the range of f (and the order-type of the E-predecessors of α in (V_κ, E) from an external point of view is precisely κ). Now let $\mathcal{M} = (V_\kappa, \in, E, f)$; this is obviously a model of $\mathsf{ZF}(E, f)$ since κ is strongly inaccessible. Also let $\mathcal{N} = (V_\kappa, E)$, and let $\mathcal{M}^* = (V_\alpha, \in)^{\mathcal{N}}$ (i.e., $(V_\alpha, \in)$-as-calculated-in-$\mathcal{N}$), where α is as define above. Note that:

1. $\mathcal{N}$ is a reduct of $\mathcal{M}$, so clearly $\mathcal{N}$ is parameter-free interpretable in $\mathcal{M}$.
2. $\mathcal{M}^*$ is interpretable in $\mathcal{N}$ with the use of the parameter α.[30]
3. There is an $\mathcal{N}$-definable isomorphism between $(V_\kappa, \in)$ and $\mathcal{M}^*$, thanks to having f available as one of the "oracles" in $\mathcal{M}$.[31]

This shows that τ_{Repl} is not e-solid since $(V_\kappa, \in)$ is not isomorphic to (V_κ, E) since the latter is isomorphic to $(K, \in)$ whose ordinal height exceeds the ordinal height of $(V_\kappa, \in)$, i.e., κ.

Theorem 73. *The template τ_{Ind} for the induction scheme is g-minimalist.*

Proof. The proof is analogous to the proof of Proposition 11. □

Theorem 74. *The following implications hold for an arbitrary $\mathcal{L}$-scheme template τ:*

$$\begin{tikzcd}
 & \tau \text{ is g-minimalist} \arrow[d] \arrow[dr] & \\
 & \tau \text{ is g-solid} \arrow[d] & \tau[\mathcal{L}] \text{ is minimalist} \\
 & \tau \text{ is e-solid} \arrow[dl] \arrow[dr] & \\
\tau \text{ is internally categorical} & & \tau[\mathcal{L}] \text{ is a solid theory}
\end{tikzcd}$$

Proof. Note that based on the relevant definitions, three of the implications are trivial, namely that g-solidity implies e-solidity, e-solidity implies solidity and g-minimalist implies minimalist.

Recall that by Remark 8 a minimalist theory is a solid theory (on the basis of the relevant definitions). A similar argument shows that if τ is g-minimalist, then τ is g-solid.

It remains to show that e-solidity entails internal categoricity. In the interest of notational clarity, we will present the proof for the template τ_{Ind} of the usual schematic presentation of PA; the same proof strategy works in general. We want to show that if $\mathcal{M}^{\text{duo}} := \left(M, +^{\mathsf{red}}, \cdot^{\mathsf{red}}, +^{\mathsf{blue}}, \cdot^{\mathsf{blue}}\right) \models \tau^{\mathsf{duo}}_{\mathsf{Ind}}[\mathcal{L}^{\mathsf{duo}}_{\mathsf{PA}}]$, then there is an $\mathcal{M}^{\text{duo}}$-definable isomorphism $i : \mathcal{M}^{\mathsf{red}} \to \mathcal{M}^{\mathsf{blue}}$. Define:

$$\mathcal{M}_1 = \mathcal{M}_2 = \mathcal{M}_3 := \mathcal{M}^{\text{duo}};$$
$$\sigma_1 = \sigma_3 := [+ \mapsto +^{\text{red}}, \cdot \mapsto \cdot^{\text{red}}], \sigma_2 := [+ \mapsto +^{\text{blue}}, \cdot \mapsto \cdot^{\text{blue}}].$$

Then clearly $\mathcal{M}_1 \trianglerighteq \mathcal{M}_2 \trianglerighteq \mathcal{M}_3$ and there is an $\mathcal{M}_1$-definable isomorphism $i_0 : \sigma_1(\mathcal{M}_1) \to \sigma_3(\mathcal{M}_3)$, so the desired $\mathcal{M}^{\mathsf{duo}}$ -definable isomorphism $i : \mathcal{M}^{\mathsf{red}} \to \mathcal{M}^{\mathsf{blue}}$ exists by e-solidity of τ_{Ind}. □

[30] If κ is chosen as the first strongly inaccessible cardinal, then α is definable in $\mathcal{N}$ as the first strongly inaccessible cardinal and therefore the interpretation of $\mathcal{M}^*$ in $\mathcal{N}$ becomes parameter-free.

[31] Indeed, with a little extra work one can show that f is definable in $(V_\kappa, \in, E)$, but this flourish is not needed for our purposes.

Remark 75. We suspect that the implication "g-solid implies e-solid" of Theorem 74 cannot be reversed; but we have not yet found a counterexample. The following show that none of the other implication arrows of Theorem 74 reverses.

1. Let $\tau_{\mathsf{CT[PA]}}$ be the template axiomatizing CT[PA] obtained conjuncting τ_{Ind} with finitely many axioms of compositional truth CT. Then by part (c) of Theorem 88 τ is g-solid, but not g-minimalist by Remark 89.
2. If τ is template for the Vaught schematic axiomatization of PA, and $\mathcal{L}$ is the language of arithmetic, then $\tau[\mathcal{L}]$ is solid by Theorem 10, but τ is not internally categorical by Theorem 52, and since e-solidity implies internal categoricity by Theorem 74, τ is not e-solid. A similar argument shows that the fact that $\tau[\mathcal{L}]$ is minimalist does not imply τ is g-minimalist.
3. As indicated in Example 72, τ_{Repl} is not e-solid, but by part (b) of Theorem 47 it is internally categorical.

Remark 76. In light of Remark 8 and Proposition 50 the bottom part of the diagram of implication arrows of Theorem 74 can be complemented with two more implications, one indicating that the solidity of $\tau[\mathcal{L}]$ implies the tightness of $\tau[\mathcal{L}]$; and the other indicating that the internal categoricity of $\tau[\mathcal{L}]$ implies the tightness of $\tau[\mathcal{L}]$ with the additional assumption that $\tau[\mathcal{L}]$ is a sequential theory that eliminates imaginaries.

Recall from part (b) of Theorem 47 that τ_{Repl} is not strongly internally categorical and not e-solid by Example 72. The result below shows that there is a schematic axiomatization of ZF that is both 'more internally categorical' and 'more solid' than the schematic axiomatization given by τ_{Repl}.

Theorem 77. *The scheme template $\tau_{\mathsf{Repl+Tarski}}$ (that axiomatizes* ZF*; see Definition 68 and the remark following it), is both strongly internally categorical and g-solid.*

Before presenting the proof of 77, we need some preliminary lemmas.

Lemma 78. *Suppose $\mathcal{K}$ and $\mathcal{M}$ are models of* ZF*, where $\mathcal{M}$ is a proper rank extension of $\mathcal{K}$ (i.e., the rank of every element in $M \setminus K$ is higher than the rank of every element of K). Assume furthermore that for every parametrically $\mathcal{M}$-definable subset D of M, $(\mathcal{K}, D \cap K) \models$ ZF(P), where P is a fresh unary predicate interpreted by $D \cap K$.* ***Then*** *there is some $\gamma \in \mathrm{Ord}^{\mathcal{M}} \setminus \mathrm{Ord}^{\mathcal{K}}$ such that:*

$$\mathcal{K} \preceq (V_\gamma, \in)^{\mathcal{M}}.$$

In particular, there is a parametrically $\mathcal{M}$-definable set that is a full satisfaction class for $\mathcal{K}$.

Proof. This follows immediately from Theorem 4.4 of [14] (whose proof is obtained by an adaptation of the proof of Theorem 3.3 of [9]). □

Lemma 79. *Suppose $\mathcal{M}$ and $\mathcal{K}$ are models of* ZF *such that the universe K of $\mathcal{K}$ is a subset of the universe M of $\mathcal{M}$ (and therefore the membership relation $\in^{\mathcal{K}}$ of $\mathcal{K}$ is a subset of M^2). Moreover, assume for every parametrically $\mathcal{M}$-definable subset D of M, $(\mathcal{K}, D \cap K) \models$ ZF(P), where P is a fresh unary predicate interpreted by $D \cap K$.* ***Then*** *precisely one of the following three conditions hold:*

(1) There is an $(\mathcal{M}, \mathcal{K})$-definable isomorphism between $\mathcal{K}$ and $(V_\kappa, \in)^{\mathcal{M}}$ for some strongly inaccessible cardinal κ of $\mathcal{M}$.

(2) There is an $(\mathcal{M}, \mathcal{K})$-definable isomorphism between $\mathcal{K}$ and $\mathcal{M}$.

(3) There is an $(\mathcal{M}, \mathcal{K})$-definable embedding of $\mathcal{M}$ as a proper rank initial segment of $\mathcal{K}$.

Proof. Using the assumptions on $\mathcal{K}$, it is easy to see that the membership relation $\in^{\mathcal{K}}$of $\mathcal{K}$ is well-founded from the point of view of $\mathcal{M}$. The proof strategy of solidity of ZF (as in [12]) can be used to show that (1) holds if K is a set in $\mathcal{M}$; (2) holds if K is a proper class in $\mathcal{M}$ such that $\in^{\mathcal{K}}$ is set-like[32] in the sense of $\mathcal{M}$; and (3) holds if K is a proper class in $\mathcal{M}$ such that $\in^{\mathcal{K}}$ is not set-like in the sense of $\mathcal{M}$. □

Proof of Theorem 77. We shall explain why $\tau_{\mathsf{Repl+Tarski}}$ is g-solid. A similar argument shows that it is also strongly internally categorical. Suppose $\mathcal{M}$, $\mathcal{M}_2$, and $\mathcal{M}_3$ are strong models of ZF, $\mathcal{M}_1 \trianglerighteq_{\mathsf{par}} \mathcal{M}_2 \trianglerighteq_{\mathsf{par}} \mathcal{M}_3$, and $\mathcal{M}_i$ accommodates $\tau_{\mathsf{Repl+Tarski}}$ for $i \in \{1, 2, 3\}$ via a model $\mathcal{K}_i$ such that for every parametrically $\mathcal{M}_i$-definable subset D of $\mathcal{M}_i$,

$$(\mathcal{K}_i, D \cap K_i) \models \tau_{\mathsf{Repl+Tarski}}(P).$$

Note that since ZF eliminates imaginaries, we can assume without loss of generality that the interpretation of each $\mathcal{K}_i$ in $\mathcal{M}_i$ is identity preserving. By Lemma 79, for each $i \in \{1, 2, 3\}$ either (1) there is an $\mathcal{M}_i$-definable isomorphism between $\mathcal{K}_i$ and $(V_\kappa, \in)^{\mathcal{M}_i}$ for some inaccessible cardinal κ of $\mathcal{M}_i$, or (2) there is an $\mathcal{M}_i$-definable

[32] In other words, for each $x \in K$, the collection $\{y \in K : y \in^{\mathcal{K}} x\}$ forms a set in $\mathcal{M}$ as opposed to a proper class.

isomorphism between $\mathcal{M}_i$ and $\mathcal{K}_i$, or (3) there is an $\mathcal{M}_i$-definable embedding of $\mathcal{M}_i$ into a proper rank initial segment of $\mathcal{K}_i$.

Thanks to $\tau_{\mathsf{Repl+Tarski}}$ and Lemma 78, possibility (1) is ruled out, and (3) is ruled out by Tarski's theorem on undefinability of truth and Lemma 78. So for each $i \in \{1,2,3\}$ there is a $\mathcal{M}_i$-definable isomorphism between $\mathcal{M}_i$ and $\mathcal{K}_i$. This reduces the initial problem to the assumptions of e-solidity of ZF. Hence the desired conclusion follows from Theorem 70. □

Below we consider two natural operations that preserve g-solidity and e-solidity of schemes. The first one (the comprehension scheme) is connected to the process of enriching a given language $\mathcal{L}$ with a new sort of sets (that are interpreted as subsets of the domain of discourse of $\mathcal{L}$); and the second (extremal schemes) enriches $\mathcal{L}$ with a new predicate.

Definition 80 (The comprehension scheme)**.** If $\mathcal{L}$ is any n-sorted language, then $\mathcal{L}^{+1}$ is a natural extension of $\mathcal{L}$ with the $n+1$-st sort.

Suppose that $\mathcal{L}$ is an n-sorted language and P a fresh predicate for the n-th sort. The comprehension scheme $\tau_{\mathsf{Comp}}(P)$ is the following statement of $\mathcal{L}^{+1}$:

$$\exists X^{n+1} \forall X^n \big(X^n \in_{n+1} X^{n+1} \equiv P(X^n)\big).$$

Example 81. If $\mathcal{L}$ is the 1-sorted language of PA, then $\tau_{\mathsf{Comp}}(P)$ is the template for the comprehension scheme of Z_2 (Second Order Arithmetic); and if $\mathcal{L}$ is the 2-sorted language of Z_2, then $\tau_{\mathsf{Comp}}(P)$ is the template for the comprehension scheme of Z_3 (Third Order Arithmetic).

The following result generalizes the fact that Z_2 and KM and their higher older analogues mentioned in Theorem 14 are solid. Recall that, as indicated in Theorem 70, the scheme templates for usual axiomatizations of PA and ZF are e-solid.

Theorem 82. *Let $*$ be either g or e. Suppose $\tau(P)$ is an $\mathcal{L}$-template that is *-solid. Then $\tau \wedge \tau_{\mathsf{Compr}}$ is *-solid (as an $\mathcal{L}^{+1}$ template).*

Proof. We show the proof for g-solidity. Assume that $\mathcal{L}$ is an n-order language. Assume that $\mathcal{M}$, $\mathcal{N}$ and $\mathcal{M}^*$ are strong models of $\tau \wedge \tau_{\mathsf{Compr}}$ which accommodate $\tau \wedge \tau_{\mathsf{Compr}}$ via the translations σ_l, σ_c and σ_r respectively and that I is an $\mathcal{M}$-definable isomorphism between $\sigma_l(\mathcal{M})$ and $\sigma_r(\mathcal{M}^*)$.

For the ease of exposition we assume that the translations are one -dimensional and equality-preserving, however the proof below can clearly be adapted to the more general setting. Let $\widehat{\sigma_l}, \widehat{\sigma_c}, \widehat{\sigma_r}, \widehat{I}$ be the natural restrictions of σ's and I to the first n-sorts. Hence $\mathcal{M}$, $\mathcal{N}$ and $\mathcal{M}^*$ are strong models of τ which accommodate τ through $\widehat{\sigma_l}, \widehat{\sigma_c}, \widehat{\sigma_r}$ and $\widehat{I}$ is an isomorphism between $\widehat{\sigma_l}(\mathcal{M})$ and $\widehat{\sigma_r}(\mathcal{M}^*)$. Hence, by g-solidity of τ we have an $\mathcal{N}$-definable isomorphism $\widehat{J}$ between $\widehat{\sigma_r}(\mathcal{M}^*)$ and $\widehat{\sigma_c}(\mathcal{N})$. $\widehat{J}$ can be canonically extended to an $\mathcal{N}$-definable embedding $J : \sigma_r(\mathcal{M}^*) \to \sigma_c(\mathcal{N})$ by putting, for an arbitrary $(n+1)$-st order object X

$$J(X) := \Big\{\widehat{J}(x) \in \widehat{\sigma_c}(\mathcal{N}) \ \mid \ x \in X\Big\} \, (= \widehat{J}[X]).$$

The fact that J is a well defined function between the $(n+1)$-st sorts of $\sigma_r(\mathcal{M}^*)$ and $\sigma_c(\mathcal{N})$ follows from the fact that $\mathcal{N}$ accommodates the full comprehension scheme through σ_c.

We claim that J is onto the $(n+1)$-st sort of $\sigma_c(\mathcal{N})$. Assume not and consider $K := J \circ I$. Then K is an $\mathcal{M}$-definable embedding of $\sigma_l(\mathcal{M})$ into $\sigma_c(\mathcal{N})$, which is not onto. Let X be outside the image of this embedding. Let $\widehat{K}$ be the restriction of K to the first n-sorts, hence $\widehat{K} = \widehat{J} \circ \widehat{I}$. Hence $\widehat{K}$ is an isomorphism between $\widehat{\sigma_l}(\mathcal{M})$ and $\widehat{\sigma_c}(\mathcal{N})$. Consider the $\mathcal{M}$-definable subset Y of $\widehat{\sigma_l}(\mathcal{M})$ given by $\widehat{K}^{-1}[X]$. We claim that $K(Y) = X$ contradicting the choice of X. Indeed, fix an arbitrary $x \in \widehat{\sigma_c}(\mathcal{N})$ and $y \in \widehat{\sigma_l}(\mathcal{M})$ such that $x = \widehat{K}(y)$. Clearly we have:

$$x \in X \Leftrightarrow \widehat{K}(y) \in X \Leftrightarrow y \in Y \Leftrightarrow K(y) \in K(Y) \Leftrightarrow x \in K(Y).$$

By Extensionality, this shows that $J(Y) = X$, as desired. □

Definition 83 (Extremal schemes)**.** Let $\mathcal{L}$ be any language containing a predicate P. Let α be any $\mathcal{L}$-sentence and β be any $\mathcal{L}$-formula. The (α, β)-*minimality* scheme, denoted $\mu_\beta(\alpha)$ is the following $\mathcal{L}$-sentence

$$\alpha \to \forall x \big(\beta(x) \to P(x)\big).$$

The (α, β)-*maximality* scheme is defined analogously as

$$\alpha \to \forall x \big(P(x) \to \beta(x)\big).$$

Example 84. Both the induction scheme and the Burgess minimality scheme [3] are instances of the minimality schemes for sufficiently chosen α and β. Indeed, to see the former, take $\alpha := P(0) \wedge \forall x\big(P(x) \to P(x+1)\big)$ and $\beta := x = x$. In this way, $\mu_\beta(\alpha)$ says that the universe is its least subset closed under successor and containing 0. To see the latter, take α to be the conjunction of the compositional axioms of KF (with the predicate P substituted for T) and $\beta := T(x)$. By considering the extension of KF with the (α, β)-maximality scheme (for α, β defined as above) one obtains the dual version of Burgess extension of KF.

Theorem 85. *Let $*$ be either g or e. Suppose that $\tau(P)$ is a $*$-solid scheme for a language $\mathcal{L}$ and let Q be a fresh predicate. Let α be any sentence of $\mathcal{L}_P$. Then the schemes $\tau \wedge \alpha[Q/P] \wedge \mu_Q(\alpha)$ and $\tau \wedge \alpha[Q/P] \wedge \nu_Q(\alpha)$ are both $*$-solid (as $\mathcal{L}_Q$ templates).*

Proof. We run the proofs for g and e in parallel. Pick $\mathcal{M}_l, \mathcal{M}_c, \mathcal{M}_r$ as in the definition of $*$-solidity for $\tau \wedge \alpha[Q/P] \wedge \mu_Q(\alpha)$ (i.e. $\mathcal{M}_l$, $\mathcal{M}_c$ and $\mathcal{M}_r$ rename $\mathcal{M}, \mathcal{N}$ and $\mathcal{M}^*$, respectively, for notational convenience).

Let σ_l, σ_c and σ_r witness the accommodations of $\tau \wedge \alpha[Q/P] \wedge \mu_Q(\alpha)$, respectively. Finally let F be the $\mathcal{M}_l$-definable isomorphism between $\sigma_l(\mathcal{M}_l)$ and $\sigma_r(\mathcal{M}_r)$. Let $\sigma'_l, \sigma'_c, \sigma'_r$ be the restrictions of these translations to the language $\mathcal{L}$. By definition, each σ'_i witnesses that $\mathcal{M}_i$ accommodates τ.

By the $*$-solidity of τ we obtain an $\mathcal{M}_c$-definable isomorphism $G_c : \sigma'_c(\mathcal{M}_c) \to \sigma'_r(\mathcal{M}_r)$ and consequently an $\mathcal{M}_l$-definable isomorphism $G_c^{-1} \circ F =: G_l : \sigma'_l(\mathcal{M}_l) \to \sigma'_c(\mathcal{M}_c)$. For $i \in \{l, r, c\}$, let Q_i be the interpretation of Q in $\sigma_i(\mathcal{M}_i)$.

We shall argue that G_c maps Q_c to Q_r and hence is also an isomorphism of $\sigma_c(\mathcal{M}_c)$ with $\sigma_r(\mathcal{M}_r)$. Since $G_l^{-1}[Q_c]$ is an $\mathcal{M}_l$-definable subset of $\sigma_l(\mathcal{M}_l)$ such that

$$(\sigma'_l(\mathcal{M}_l), G_l^{-1}[Q_c]) \models \alpha[Q/P],$$

by the minimality scheme in $\mathcal{M}_l$ it follows that $Q_l \subseteq G_l^{-1}[Q_c]$. Hence $F[Q_l] \subseteq G_c[Q_c]$. However, since F is an isomorphism of $\mathcal{L}_Q$ structures, then $F[Q_l] = Q_r$. By applying G_c^{-1} we get $G_c^{-1}[Q_r] \subseteq Q_c$. However, since

$$(\sigma'_c(\mathcal{M}_c), G_c^{-1}[Q_r]) \models \alpha[Q/P],$$

and $G_c^{-1}[Q_r]$ is $\mathcal{M}_c$-definable, by the minimality scheme we obtain $Q_c \subseteq G_c^{-1}[Q_r]$, which concludes the proof. □

For completeness we note the following solidity variant of Theorem 60, which is proved with a similar argument.

Theorem 86. *Let τ and σ be two scheme-templates and $\mathcal{L}$ be a language such that $\tau[\mathcal{L}] \equiv \sigma[\mathcal{L}]$. Denote this theory with T. Suppose further that $\mathsf{PC}(T) \vdash \forall X \tau[X] \to \forall X \sigma[X]$. Then if σ is $*$-solid (for $* \in \{e, g\}$), then so is τ.*

We close this section by applying our results to canonical axiomatic theories of truth.

Definition 87. (Truth theories)

(a) CT[PA] is the theory obtained by augmenting PA with CT (finitely many axioms of compositional truth over arithmetic) with all instances of induction in the extended language (i.e., the language of PA augmented with a truth predicate).

(b) KF_μ[PA] is Burgess' extension [4] of the Kripke-Feferman truth theory KF[PA]; the choice μ is informed by the fact that the truth theory corresponding to the *least* fixed point of the usual construction of a model of KF[PA] yields a model of KF_μ[PA].

(c) Let $\mathsf{CT}_\sigma[\tau]$ be the theory in the language $\mathcal{L}_S$ with a fresh binary predicate S extending the schematic theory generated by the template $\tau \wedge \tau_{\mathsf{Ind}}[\mathcal{L}_S]$ with axioms that are the natural counterparts of compositional axioms of CT in the language with the satisfaction predicate, in which:

(1) the syntactical operations are translated by σ;

(2) the quantifiers over syntactical objects are relativized to the domain of σ;

(3) the quantifiers over assignments are unrelativized.

(d) $\mathsf{KF}_{\mu,\sigma}[\tau]$ is analogously defined as $\mathsf{CT}_\sigma[\tau]$, as an extension of the schematic theory generated by the template $\tau \wedge \tau_{\mathsf{Ind}} \wedge \mu_T(\bigwedge \mathsf{KF})$, where $\bigwedge \mathsf{KF}$ denotes the conjunction of the compositional axioms of KF.

Theorem 88. *The following axiomatic truth theories are solid:*

(a) $\mathsf{CT}[\mathsf{PA}]$, *and each of its finite iterates* $\mathsf{CT}[\mathsf{CT}[\mathsf{PA}]]$, *etc.*

(b) $\mathsf{KF}_\mu[\mathsf{PA}]$.

(c) *More generally, suppose that* τ *is a* $*$*-solid (where* $*$ *is either* g *or* e*)* $\mathcal{L}$*-scheme such that* $\tau[\mathcal{L}]$ *accommdates the induction scheme* τ_{Ind} *via a translation* σ*.* ***Then*** $\mathsf{CT}_\sigma[\tau]$ *and* $\mathsf{KF}_{\mu,\sigma}[\tau]$ *are solid theories.*

Proof. The solidity of $\mathsf{CT}[\mathsf{PA}]$ (as well as its finite iterates) can be readily obtained through the g-solidity of the scheme $\bigwedge \mathsf{CT} \wedge \tau_{\mathsf{Ind}}$, , where $\bigwedge \mathsf{CT}$ is the conjunction of the finitely many compositional axioms of CT. Indeed, observe that over $\mathrm{PC}(\mathsf{CT}[\mathsf{PA}])$, the sentence $\forall X \tau_{\mathsf{Ind}}(X)$ implies the minimality of T, i.e. $\forall X \mu_T(\bigwedge \mathsf{CT})$. Since $\tau_{\mathsf{Ind}} \wedge \bigwedge \mathsf{CT} \wedge \mu_T(\bigwedge \mathsf{CT})$ is g-solid (Theorem 85), so is $\bigwedge \mathsf{CT} \wedge \tau_{\mathsf{Ind}}$. Hence CT is indeed solid.

For the proof of (b) it is enough to observe that Burgess' KF_μ is axiomatized by the scheme $\sigma := \tau_{\mathsf{Ind}} \wedge \bigwedge \mathsf{KF} \wedge \mu_T(\bigwedge \mathsf{KF})$, where $\bigwedge \mathsf{KF}$ is the conjunction of the finitely many compositional axioms of KF. τ_{IND} is g-minimalist, hence by Theorem 85 σ is g-solid. Thus $\sigma[\mathcal{L}_T]$ is solid (where $\mathcal{L}_T$ is the language extending the language of arithmetic with a fresh unary predicate T).

The proof of (c) follows from Theorem 85, which generalizes parts (a) and (b). □

Remark 89. The minimalist property of PA is not shared by $\mathsf{CT}[\mathsf{PA}]$. To see this, it is sufficient to observe that if $\mathcal{M}$ and $\mathcal{N}$ are models of $\mathsf{CT}[\mathsf{PA}]$ and $\mathcal{M}$ is a submodel of $\mathcal{N}$, then the arithmetical theories of the two models are the same. Hence, in order to demonstrate that $\mathsf{CT}[\mathsf{PA}]$ is not minimalist, it is enough to consider $\mathcal{M} \models \mathsf{CT}[\mathsf{PA}] + \mathsf{Con}_{\mathsf{CT}[\mathsf{PA}]}$. Then $\mathcal{M}$, by the formalized second Gödel's incompleteness theorem, $\mathcal{M} \models \mathsf{Con}_{\mathsf{CT}[\mathsf{PA}]+\neg\mathsf{Con}_{\mathsf{CT}[\mathsf{PA}]}}$, hence $\mathcal{M}$ interprets a model $\mathcal{N} \models \mathsf{CT}[\mathsf{PA}] + \neg\mathsf{Con}_{\mathsf{CT}[\mathsf{PA}]}$ via the Arithmetized Completeness Theorem. However there can be no embedding between $\mathcal{M}$ and $\mathcal{N}$ as the arithmetical theories of the two models are different.

5 Philosophical discussion

In this section we outline the philosophical implications of the formal results in the preceding sections that we view as relevant for contemporary debates and discussions in philosophy of science and the foundations of mathematics. We focus mostly on two topics: (1) The use of bi-interpretability and synonymy (definitional equivalence) as good explications of sameness of theories, a view present in discussions in the philosophy of science; and (2) The use of categoricity-style arguments in the debate over the determinacy of some core mathematical concepts, such as the notion of a number, or the notion of a set. These two conceptual areas naturally correspond to Sections 3 and 4 of our paper.

5.1 The meaning of bi-interpretability, solidity and tightness

Synonymy is typically treated as a good explication of the notion of sameness of theories, which works nicely for theories in different signatures.[33] Intuitively, a definitional extension of a theory T "says no more" than T itself, hence two theories that have a common definitional extension "say the same thing", and thus they should have the same content. However, despite the conceptual attractiveness of synonymy, this approach needs to contend with some notorious examples of theories that are synonymous, yet seem to be clearly different. One of the most famous examples is provided by two extensions of $\mathsf{KF}[\mathsf{PA}]$ with axioms expressing (Cons) "No sentence is both true and false" and (Comp) "Every sentence is either true or false". By trivially disjoining the signatures of $\mathsf{KF}[\mathsf{PA}] + \mathsf{Comp}$ and $\mathsf{KF}[\mathsf{PA}] + \mathsf{Cons}$, one easily finds a common definitional extension of the two theories.[34] However, $\mathsf{KF}[\mathsf{PA}] \vdash \mathsf{Cons} \rightarrow \neg\mathsf{Comp}$, so the two theories are inconsistent.

Section 3.4 provides new examples of such unexpected failures of sameness between synonymous theories: each restricted fragment of any of the canonical theories from among $\{\mathsf{PA}, \mathsf{ZF}, \mathsf{Z}_2, \mathsf{KM}\}$ has two different complete bi-interpretable extensions. In particular for each n, $\mathrm{I}\Sigma_n$ has a model in which $\mathsf{I}\Sigma_{n+1}$ induction fails and yet the model is bi-interpretable with the standard model of arithmetic. By examining the proof, one notices that in fact the interpretations used in the bi-interpretability do not involve parameters and are equality preserving, hence by the Visser-Friedman theorem [23] the theories of both models are synonymous (treated as theories in disjoint signatures). However, the theories disagree on whether the induction holds and it is hard to imagine a more striking example of a foundational disagreement.

One can counter the conclusions of the above paragraphs, by pointing out that we misinterpret the *content* of theories, as measured by synonymy. For if we agree that a definitional extension of a theory U says no more than

[33] It is worth adding that sometimes synonymy is taken to be too restrictive and a weaker notion, called Morita equivalence is preferred; as shown in [40] Morita equivalence implies bi-interpretability under very mild assumptions.

[34] One defines a complete/consistent truth predicate T_2 using a consistent/complete one, T_1 by putting $T_2(x) := \neg T_1(\dot{\neg} x)$.

U, then we implicitly assume that what we really care about is the structure of U-definable sets. We abstract from the complexity of the definitions and treat all the notions that can be expressed in U on a par. This explains the sameness of $\mathsf{KF}[\mathsf{PA}]+\mathsf{Cons}$ and $\mathsf{KF}[\mathsf{PA}]+\mathsf{Comp}$: using a consistent truth predicate we can indeed define a complete one and vice versa. Hence two theories are equally expressive, only they disagree on which truth predicate is "the basic one". We agree with this argument, however we think that for the philosophical use of axiomatic theories one would expect a more fine-grained notion. As we explain in the next paragraph, bi-interpretability over tight theories seems to preserve more information about what the theories are meant to axiomatize.

Positive results about solidity and tightness show that "counterexamples" of the above discussed type do not occur if we restrict attention to the extensions of appropriate theories: whenever U and V are theories in the same language which both extend any of the solid theories from the list including PA, ZF, Z_n, KM, $\mathsf{CT}[\mathsf{PA}]$, $\mathsf{KF}_\mu[\mathsf{PA}]\ldots$, then U is a retract of V implies $U \vdash V$. This leads to a conclusion that bi-interpretability and synonymy are more restrictive on "cones" originating from foundationally salient theories[35]

The above naturally leads to a question: What is so special about solid theories? Unlike internal categoricity, whose definition preserves many intuitions supporting the traditional notion of categoricity, solidity seems to lack such a 'conceptual charm'. We shall try to explain the intuitions behind this notion more carefully. First of all, let us observe that solidity speaks about the *local* behaviour of a first-order U across all models of U, i.e., it considers what happens in each of the models of the theory separately. Speaking very informally, each model of a solid theory thinks of itself as a special or distinguished model from among the ones that it can talk about. More formally: consider a solid theory U and a model $\mathcal{M} \models U$ and let $U(\mathcal{M})$ be the set of all models of U that are parametrically definable in $\mathcal{M}$ (considered up to the $\mathcal{M}$-definable isomorphism). Of course, $\mathcal{M}$ itself is an element of $U(\mathcal{M})$. Elements of $U(\mathcal{M})$ are naturally pre-ordered by the relation of parametric definability. Intuitively speaking $\mathcal{N}$ is (strictly) greater in this ordering than a model $\mathcal{K}$, if $\mathcal{N}$ can see $\mathcal{K}$, but not vice versa. Solidity simply says that $\mathcal{M}$ is the greatest element in this pre-order: whenever $\mathcal{N} \in U_{\mathcal{M}}$ (hence $\mathcal{M}$ sees $\mathcal{N}$) and $\mathcal{N}$ is also greater or equal to $\mathcal{M}$ ($\mathcal{N}$ sees an isomorphic copy of $\mathcal{M}$), then actually $\mathcal{N}$ is equal to $\mathcal{M}$ (there is an $\mathcal{M}$-definable isomorphism between $\mathcal{N}$ and $\mathcal{M}$). We think that this explanation should suffice to put solidity on the list of categoricity-like properties for first-order theories. Let us observe that, unlike internal categoricity, solidity is a categoricity-like notion which naturally applies to first-order theories as opposed to schemes.

Let us complete this subsection with a metaphor that guides our thinking about an important consequence of solidity, namely *neatness* (as in Definition 7), which links the notion of a retract with that of theory extension. One can "define" a natural or canonical theory as the one that "has a story to tell". If so, then one can see a neat theory to be akin to the first episode of a well-written series of stories, in which the characters, and the relationships and tensions between them, are well-developed. Such a successful first episode is self-standing in its own right, but also allows the next episodes to be developed on the basis of it, without being seen as its inevitable conclusion.

5.2 Internal categoricity: second order predicates, schemes vs. (first-order) theories

We shall start from a reconstruction of an extensively discussed philosophical puzzle: the structure of the natural numbers *seems to be* both (a) well determined (by the operation of the successor and basic recursive equations for addition and multiplication) and (b) first-order describable. Of course, a natural way of making the above claim mathematically precise demonstrates that in fact (a) contradicts (b): the full (complete) first-order theory of the natural numbers (the standard model) can be realized in various nonstandard models. One can react to this phenomenon by admitting that there is indeed an irreducible second-order factor in our understanding of the canonical structure of the natural numbers; or one might try to explicate the condition of "uniqueness" in some other way. The internal categoricity approach uncovers a path leading towards a philosophically intriguing explanation of the latter option. It starts with an observation that actually the categoricity of various prominent second-order systems is *essentially* a very concrete phenomenon (we shall refer to this loosely defined property as "concrete categoricity")[36]: the statement "Each two structures which satisfy the theory U are isomorphic" is provable in the canonical proof system for second-order logic (which is sound and complete w.r.t. Henkin semantics) for the cases when U is one of the canonical theories such as PA, $\mathsf{CT}[\mathsf{PA}]$, restricting to models of the same ordinal height in the case of ZF. Even though to obtain these results one makes use of the second-order logic with Henkin semantic (with full comprehension), the appeal to the full second-order semantics has been bypassed. Secondly, one can formulate a first-order version of this observation by further noticing that in each of the cases there is a procedure of finding the required isomorphism, which is uniform with respect to the definitions of the models for the given theory. In this way one arrives at the definitions of *internal categoricity* (for first-order theories) and what we here dubbed *strong internal categoricity* in Definition 44. In both the syntactic second-order and the first-order approaches one obtains *intolerance* results: If U is a *concretely* categorical theory, then for every sentence φ

[35] Recent unpublished work due to Piotr Gruza, Leszek Kołodziejczyk and the second-named-author show that in some of these cases, and most probably in all of them, the full strength of these theories is not necessary to eliminate pathologies in the behaviour of bi-interpretability.

[36] This is because we decided to fix a meaning of "internal categoricity" that seems to be used in this context.

separately one can prove that either φ holds in all U-structures or $\neg\varphi$ holds in all U-structures.[37] An interesting variation of the syntactic second-order approach and the first-order approach was put forward by Fischer and Zicchetti [20]: the authors use an axiomatic theory of truth to obtain a first-order intolerance in the form of a single universal sentence (as opposed to a scheme).

Our results from Section 4 bring more information about the peculiarities of each of the above introduced approaches to formalising the idea of concrete categoricity, i.e., the syntactic second-order approach, the first-order approach and the truth-theoretic approach.[38] First of all, we developed a simple hierarchy of categoricity-like notions (as in Theorem 74) in the first-order realm that can serve as a simple indicator of a 'concrete categoricity degree' of a given first-order portion of mathematics. The internal categoricity is at the bottom of this hierarchy. Secondly, we claim that the properties defining each level in this categoricity scale (including internal categoricity) are best seen as properties of *schemes*, as opposed to first-order *theories*. This is supported by the following observations:

(a) A first order theory may be axiomatized by schemes with very different "degrees of concrete categoricity", ranging from those that are not internally categorical (the Vaught schemes, see Theorem 52) up to those that are g-minimalist (in the case of the usual schematic axiomatization of PA).

(b) This perspective is fruitful: we can define a simple syntactic operations on schemes and show that some categoricity degrees are preserved by them thus explaining some regularities in the distribution of internally categorical theories.

What is more, there is a very clear philosophical idea explaining why schemes are better carriers of internal categoricity than first-order theories: as opposed to theories, schemes are potentially *open-ended*, which means that they are meaningfully applicable to any first-order language. Continuing along these lines, mathematical agents can accept a given scheme τ as an open-ended entity, which means that they commit to extending τ to any language they find comprehensible. Mathematically speaking, schemes are functions which, when applied to a first-order language, return a first-order theory. Seen from this perspective, Theorem 52 shows that two such functions may coincide on a given language, however behave very differently on another ones. Further degrees on our scale, e-solidity, g-solidity and g-minimalist, are underpinned by the same intuition: how competent a given scheme is when presented with sets definable in various other languages. However, in designing the formal definitions we were admittedly led by analogies with the categoricity notions of theories and interpretability theory (the distinction between g- and e-solidity reflects the difference between the direct and relative interpretation).

The above highlights the core difference between schemes and first-order theories. How about schemes and second order formulations of theories, present in the syntactic second-order approach? Here we point to one interesting phenomenon linked to the Tarski component added to various schemes. Consider the replacement scheme conjuncted with the scheme expressing "no formula defines the truth for the set-theoretic universe" (as in $\tau_{\mathsf{Repl+Tarski}}$ of Definition 68). As shown in Theorem 77 this scheme is, on our scale, at least one level 'more categorical' in relation to the usual schematic axiomatization of ZF.[39] Moreover, there is a clear philosophical idea why mathematical agents might like to accept such a scheme in an open-ended fashion. Let's think about a set-theorist claiming that there is one all-embracing universe of sets, V, and all of mathematics takes place within it. It would be natural to think that there is no external-to-V point of view. By the Tarski undefinability of truth theorem, the truth for V should be absolutely undefinable, a claim which is reflected in the acceptance of the Tarski component of our scheme.[40]

What can be said about the categoricity of $\tau_{\mathsf{Repl+Tarski}}$ in the context of second-order logic with Henkin semantics and full comprehension? Indeed, in this second order context $\tau_{\mathsf{Repl+Tarski}}$ is categorical (in the sense that any two models of the second order theory are isomorphic), provably in the second order logic. However, this is because of the fact that provably in second order logic there is no structure that satisfies the second order version of $\tau_{\mathsf{Repl+Tarski}}$ (for a scheme τ in the signature with one binary relational symbol and for two second order variables X, E, $\tau^{X,E}$ denotes the result of relativizing the quantifiers in τ to X and substituting E for the relational symbol):

Theorem 90. *The following sentence is provable in pure second order logic (with full comprehension):*

$$\forall X\ \forall E\ \neg\forall Y \subseteq E\ \ \tau^{X,E}_{\mathsf{Repl+Tarski}}(Y).$$

Proof. Suppose otherwise and consider a Henkin model $\mathcal{M}$ such that some pair of sets (X, E) satisfies $\tau_{\mathsf{Repl+Tarski}}$ with respect to all subsets of X which exists in $\mathcal{M}$. However, the Tarskian satisfaction class for (X, E) can be seen (via coding) as a subset of X, which is Σ^1_1-definable in (X, E), so by comprehension it exists as an element of $\mathcal{M}$. This contradicts the fact that (X, E) satisfies the Tarski component of $\tau_{\mathsf{Repl+Tarski}}$. □

[37] For the precise explanation of these results in each scenario, consult [36].

[38] Indeed, Fischer and Zicchetti consider also one more perspective based on the work of Feferman and Hellman.

[39] And two levels "more categorical" if g-solidity is strictly stronger than e-solidity, which we conjecture.

[40] The philosophical ramifications of the open-ended feature of schemes has been explored by Charles Parsons, see, e.g., [45].

We think that the above observation speaks in favour of the schematic approach to the phenomenon of concrete categoricity. Unlike in the syntactic second order approach, the schematic approach can do justice to the coherence of the standpoint motivating the acceptance of schemes with the Tarski component. The fact that this cannot be accomplished in the syntactic second-order framework is, in our opinion, due to the fact that the full second-order comprehension introduces a perspective that is (at least partially) external to second order theories under consideration. In our proof, the Tarskian satisfaction class, although undefinable in the structure (X, E) can be shown to exists by the use of Σ^1_1-comprehension, available in the full second-order logic. Let us observe that these arguments are in line with the known objections against making use of impredicative (Π^1_1) comprehension to explain Parson's problem on how can mathematicians know that they are talking about the same structure of natural numbers. We refer to [20] for a discussion.

Last but not least, we stress how our results can be used to extend the truth-theoretic approach, put forward in [20]. An important step in this approach was to show that the the usual scheme axiomatizing the theory $\mathsf{CT}[\mathsf{PA}]$ is strongly internally categorical. We show (Theorem 85) that also an untyped theory of truth, $\mathsf{KF}_\mu[\mathsf{PA}]$ is g-solid, and a similar argument can be applied to show that it is strongly internally categorical. Hence one can devise a truth-theoretic approach based on an untyped truth predicate.

6 Future work

There is still much to be learned about categoricity-like properties of first order theories. The following list of questions and conjectures provides a glimpse of the unexplored part of the territory.

1. **Question.** *Is there a consistent sequential theory that is maximalist?* The notion of a maximalist theory was introduced in Definition 7.
2. **Conjecture.** *There is no finitely axiomatizable sequential tight theory.* Note that by Theorem 38 the conjecture is confirmed for finitely axiomatized subtheories of the canonical theories in the list $\mathcal{S}$ of Theorem 14.
3. **Question.** *Is there a solid deductively closed proper subtheory of* ZF *that includes* KP (*Kripke-Platek set theory*) *and the axiom of infinity?* This question is informed by the solidity of ZF.
4. **Question.** *Is the* $\mathsf{PA}^- + \mathsf{Collection}$ *a tight theory?* Recall from Theorem 35 that $\mathsf{PA}^- + \mathsf{Collection}$ is not a solid theory.
5. **Conjecture.** *For every sequential r.e. theory T in a finite language $\mathcal{L}$ there is an $\mathcal{L}$-scheme template τ such that $\tau[\mathcal{L}] = T$ and τ is not internally categorical.* This conjecture is informed by Theorem 52.
6. **Question.** *What is the relationship between the notion of internal categoricity in the context of Henkin models of pure second order logic (as in [5]) and the categoricity-like concepts of scheme templates studied in this paper?* This question is motivated by juxtaposing the (proof of) Theorem 77 with Theorem 90.
7. **Question.** *What are the categoricity-like properties of the so-called restricted schemes (in the sense of Wilkie [61])?*
8. **Conjecture.** *There is an e-solid scheme template τ that is not g-solid and τ axiomatizes a sequential theory.*
9. **Question.** *Is there an arithmetical template that is e-solid, the theory T it generates includes* PA^-*, and the deductive closure of T is a proper subtheory of* PA*?*
10. **Question.** *Do the results in this paper shed any light on the Parsons' dilemma about the determinacy of the structure of the natural numbers?*

References

[1] Wilhelm Ackermann. Die Widerspruchsfreiheit der allgemeinen Mengenlehre. *Math. Ann.*, 114(1):305–315, 1937.

[2] Jon Barwise. *Admissible sets and structures*. Perspectives in Mathematical Logic. Springer-Verlag, Berlin-New York, 1975. An approach to definability theory.

[3] Tyler Burge. Our entitlement to self-knowledge. *Proceedings of the Aristotelian Society*, 96(1):91–116, 1996.

[4] John P. Burgess. Friedman and the axiomatization of Kripke's theory of truth. In *Foundational adventures*, volume 22 of *Tributes*, pages 125–148. Coll. Publ., London, 2014.

[5] Tim Button and Sean Walsh. *Philosophy and model theory*. Oxford University Press, Oxford, 2018. With a historical appendix by Wilfrid Hodges.

[6] Patrick Cégielski, Charalampos Cornaros, and Costas Dimitracopoulos, editors. *New studies in weak arithmetics*, volume 211 of *CSLI Lecture Notes*. CSLI Publications, Stanford, CA; Presses Universitaires du Pôle de Recherche et d'Enseignement Supérieur Paris-Est, Créteil, 2013.

[7] C. C. Chang and H. J. Keisler. *Model theory*, volume 73 of *Studies in Logic and the Foundations of Mathematics*. North-Holland Publishing Co., Amsterdam, third edition, 1990.

[8] Conden Chao and Payam Seraji. Gödel's second incompleteness theorem for Σ_n-definable theories. *Log. J. IGPL*, 26(2):255–257, 2018.

[9] Ali Enayat. Conservative extensions of models of set theory and generalizations. *J. Symbolic Logic*, 51(4):1005–1021, 1986.

[10] Ali Enayat. Leibnizian models of set theory. *J. Symbolic Logic*, 69(3):775–789, 2004.

[11] Ali Enayat. Models of set theory with definable ordinals. *Arch. Math. Logic*, 44(3):363–385, 2005.

[12] Ali Enayat. Variations on a Visserian theme. In *A tribute to Albert Visser*, volume 30 of *Tributes*, pages 99–110. Coll. Publ., London, 2016.

[13] Ali Enayat. Tight theories. *Online Workshop on Gödel Incompleteness Theorems*, 2021, http://yongcheng.whu.edu.cn/webPageContent/Goedel2021/Goedel-Enayat.pdf.

[14] Ali Enayat. Models of set theory: Extensions and dead-ends ,. *arXiv:1804.09526 [math.LO]*, 2024.

[15] Ali Enayat and Mateusz Łełyk. Axiomatizations of Peano Arithmetic: A truth-theoretic view. *J. Symb. Log.*, 88(4):1526–1555, 2023.

[16] Ali Enayat, Mateusz Łełyk, and Albert Visser. Completions of restricted complexity,. *to appear*, 2024.

[17] Ali Enayat and Zachiri McKenzie. Initial self-embeddings of models of set theory. *J. Symb. Log.*, 86(4):1584–1611, 2021.

[18] Ali Enayat, James H. Schmerl, and Albert Visser. ω-models of finite set theory. In *Set theory, arithmetic, and foundations of mathematics: theorems, philosophies*, volume 36 of *Lect. Notes Log.*, pages 43–65. Assoc. Symbol. Logic, La Jolla, CA, 2011.

[19] Solomon Feferman. Arithmetization of metamathematics in a general setting. *Fundamenta Mathematicae*, 1960.

[20] Martin Fischer and Matteo Zicchetti. Internal categoricity, truth and determinacy. *J. Philos. Logic*, 52(5):1295–1325, 2023.

[21] Alfredo Roque Freire and Joel David Hamkins. Bi-interpretation in weak set theories. *J. Symb. Log.*, 86(2):609–634, 2021.

[22] Alfredo Roque Freire and Kameryn J. Williams. Non-tightness in class theories and second order arithmetic. *The Journal of Symbolic Logic*, page 1–28, 2023.

[23] Harvey Friedman and Albert Visser. When bi-interpretation implies synonymy. *Logic Group Preprint Series*, 9(3):1–19, 2014.

[24] Victoria Gitman, Joel David Hamkins, and Thomas A. Johnstone. What is the theory ZFC without power set? *MLQ Math. Log. Q.*, 62(4-5):391–406, 2016.

[25] Petr Hájek and Pavel Pudlák. *Metamathematics of first-order arithmetic*. Springer, 1998.

[26] Wilfrid Hodges. *Model theory*, volume 42 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1993.

[27] Thomas Jech. *Set theory*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, millennium edition, 2003.

[28] Emil Jeřábek. Sequence encoding without induction. *MLQ Math. Log. Q.*, 58(3):244–248, 2012.

[29] Akihiro Kanamori. *The higher infinite*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, second edition, 2009. Large cardinals in set theory from their beginnings, Paperback reprint of the 2003 edition.

[30] Richard Kaye. *Models of Peano Arithmetic*. Oxford: Clarendon Press, 1991.

[31] Richard Kaye and Tin Lok Wong. On interpretations of arithmetic and set theory. *Notre Dame J. Formal Logic*, 48(4):497–510, 2007.

[32] Roman Kossak and James Schmerl. *The Structure of Models of Peano Arithmetic*. Oxford, England: Clarendon Press, 2006.

[33] G. Kreisel and A. Lévy. Reflection principles and their use for establishing the complexity of axiomatic systems. *Z. Math. Logik Grundlagen Math.*, 14:97–142, 1968.

[34] Kenneth Kunen. *Set theory*, volume 102 of *Studies in Logic and the Foundations of Mathematics*. North-Holland Publishing Co., Amsterdam-New York, 1980. An introduction to independence proofs.

[35] Koen Lefever and Gergely Székely. On generalization of definitional equivalence to non-disjoint languages. *J. Philos. Logic*, 48(4):709–729, 2019.

[36] Penelope Maddy and Jouko Väänänen. *Philosophical Uses of Categoricity Arguments*. Cambrversity University Press, 2023.

[37] W. Marek and P. Zbierski. On a class of models of the nth order arithmetic. In *Higher set theory* (*Proc. Conf., Math. Forschungsinst., Oberwolfach, 1977*), volume 669 of *Lecture Notes in Math.*, pages 361–374. Springer, Berlin, 1978.

[38] David Marker. End extensions of normal models of open induction. *Notre Dame J. Formal Logic*, 32(3):426–431, 1991.

[39] A. R. D. Mathias. The strength of Mac Lane set theory. *Ann. Pure Appl. Logic*, 110(1-3):107–234, 2001.

[40] Paul Anh McEldowney. On morita equivalence and interpretability. *The Review of Symbolic Logic*, 13:388 – 415, 2020.

[41] R. Montague. Semantical closure and non-finite axiomatizability. I. In *Infinitistic Methods* (*Proc. Sympos. Foundations of Math., Warsaw, 1959*), pages 45–69. Pergamon, Oxford-London-New York-Paris, 1961.

[42] Jan Mycielski. Definition der arithmetischen Operationen im Ackermannschen Modell. *Algebra Logika*, 3(5-6):64–65, 1964.

[43] Fedor Pakhomov and Albert Visser. On a question of Krajewski's. *The Journal of Symbolic Logic*, 84:343 – 358, 2019.

[44] J. B. Paris and L. A. S. Kirby. Σ_n-collection schemas in arithmetic. In *Logic Colloquium '77* (*Proc. Conf., Wrocław, 1977*), volume 96 of *Stud. Logic Found. Math.*, pages 199–209. North-Holland, Amsterdam-New York, 1978.

[45] Charles Parsons. *Mathematical thought and its objects*. Cambridge University Press, Cambridge, 2008.

[46] Julia Robinson. Definability and decision problems in arithmetic. *J. Symbolic Logic*, 14:98–114, 1949.

[47] Philipp Schlicht. Perfect subsets of generalized Baire spaces and long games. *J. Symb. Log.*, 82(4):1317–1355, 2017.

[48] Farmer Schultzenberg. Projective well-ordered sets, higher up. MathOverflow. URL:https://mathoverflow.net/q/460467 (version: 2023-12-20).

[49] Dana Scott. More on the axiom of extensionality. In *Essays on the foundations of mathematics*, pages 115–131. Magnes Press, The Hebrew University, Jerusalem, 1961.

[50] J. C. Shepherdson. A non-standard model for a free variable fragment of number theory. *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys.*, 12:79–86, 1964.

[51] Stephen G. Simpson. *Subsystems of second order arithmetic*. Perspectives in Logic. Cambridge University Press, Cambridge; Association for Symbolic Logic, Poughkeepsie, NY, second edition, 2009.

[52] Robert M. Solovay. A model of set-theory in which every set of reals is Lebesgue measurable. *Ann. of Math. (2)*, 92:1–56, 1970.

[53] Jouko Väänänen. Second order logic or set theory? *Bull. Symbolic Logic*, 18(1):91–121, 2012.

[54] Jouko Väänänen. An extension of a theorem of Zermelo. *Bull. Symb. Log.*, 25(2):208–212, 2019.

[55] Jouko Väänänen. Tracing internal categoricity. *Theoria*, 87(4):986–1000, 2021.

[56] Jouko Väänänen and Tong Wang. Internal categoricity in arithmetic and set theory. *Notre Dame J. Form. Log.*, 56(1):121–134, 2015.

[57] Robert L. Vaught. Axiomatizability by a schema. *J. Symbolic Logic*, 32:473–479, 1967.

[58] Albert Visser. Categories of theories and interpretations. In *Logic in Tehran*, volume 26 of *Lect. Notes Log.*, pages 284–341. Assoc. Symbol. Logic, La Jolla, CA, 2006.

[59] Albert Visser. What is sequentiality? In *New studies in weak arithmetics*, volume 211 of *CSLI Lecture Notes*, pages 229–269. CSLI Publ., Stanford, CA, 2013.

[60] Albert Visser. On Q. *Soft Computing*, 21(3):39–56, 2016.

[61] A. J. Wilkie. On schemes axiomatizing arithmetic. In *Proceedings of the International Congress of Mathematicians, Vol. 1, 2* (*Berkeley, Calif., 1986*), pages 331–337. Amer. Math. Soc., Providence, RI, 1987.

[62] Kameryn Williams. The structure of models of second order set theories. *arXiv:1804.09526* [*math.LO*], 2018.

THE FOURTH GRADE OF MODAL INVOLVEMENT[*]

Volker Halbach

7th July 2020

I present a logical framework for analyzing *de re* modalities in their full force. The sentential operators of modal logic and unary semantic predicates in Quine's sense are shown to fall short of capturing the full strength of modal discourse. Instead we employ binary modal predicates applying to formulæ (or relations) and variable assignments. Possible worlds semantics for such predicates is sketched. Finally, some applications to metaphysics, more specifically, actualism are explored.

FOUR GRADES OF MODAL INVOLVEMENT

Modalities are at the centre of many philosophical debates. They include apriority, analyticity, metaphysical, as well as logical and physical necessity, future and past truth, and being knowable or known. In philosophical logic the main tool for analyzing modal discourse is modal or intensional logic in the wide sense that includes temporal, epistemic, and deontic logic. The expressive weaknesses of modal logic have been known for a long time, and it is somewhat of an embarrassment of philosophical logic that the standard formalization of modality in its straightforward forms does not suffice as a framework for many well-known philosophical debates.

In this paper I give some hints for developing formal languages with possible worlds semantics that can analyze modal discourse in its full force. Much of what I am going to say is tentative and experimental. This is because the project is huge: There is at least as much scope for work on this approach than on quantified modal logic. Therefore, I can only sketch some basic considerations. There are many aspects that

[*] I am grateful to Beau Mount, Simon Nagler, Carlo Nicolai, Johannes Stern, Philip Welch, and Timothy Williamson for useful hints and suggestions.

will not be addressed at all. In particular, I will not say anything about extensions such as actuality operator, two-dimensional semantics, and so on.

I focus on metaphysical necessity as my main example of a modality; but the machinery sketched below is also adaptable to other modalities. Future and past truth should be straightforward, but epistemic modalities may be more difficult.

As a starting point, I revisit Quine's *Three grades of modal involvement* (1976, first published 1953). The three grades are three ways to view necessity, namely as a *semantical predicate*, a *statement operator*, or a *sentence operator*.

In modal logic modalities are conceived as *operators*: An operator $\Box$ is combined with a formula φ to obtain a new formula $\Box\varphi$. The difference between statement and sentence operators is that the latter allow quantification into the scope of the modal operator. Thus a sentence operator $\Box$ can be combined with a formula φ containing free variables into a formula $\Box\varphi$ that contains the same free variables as φ itself. Quine's terminology is somewhat misleading, because sentences are understood here as *open* sentences with free variables. A statement operator does not permit quantification into its scope. *De re* modalities can be analyzed with sentence, but not with statement operators.

According to Quine, the lowest grade of modal involvement permits only 'semantic predicates': A semantic predicate Nec for necessity is combined with a singular term into a formula.[1] Quine thinks of semantic predicates as applying to sentences, not propositions or other objects that Quine called 'creatures of darkness' elsewhere; but this is not the point here. It is rather that a semantic predicate can be applied to a variable: $\mathsf{Nec}\,x$ is a formula with x free.

Quine thought of semantic predicates as a light grade of modal involvement, but in another sense they are expressively richer than operators. Using modal predicates one can express quantifications. Quantified statements such as 'There are synthetic judgements a priori', 'All theorems of arithmetic are analytic', or 'There are necessary a posteriori truths' cannot be expressed using only an operator $\Box$, at least not in a straightforward way. However, if modalities are conceived as predicates, quantified statements can be easily expressed: Kant's rejection 'There are synthetic judgements a priori' of empiricism, becomes the sentence $\exists x\,(\mathsf{Syn}(x) \wedge \mathsf{Apriori}(x))$ (omitting the restriction to judgements).[2]

For these reasons we require modal predicates. They are required to analyze the full force of modal discourse, especially in philosophy. With a semantic predicate for a

[1] Quine used 'Nec' (upper case) for the predicate and 'nec' (lower case) for the operator. Here the latter is replaced with the more familiar $\Box$.

[2] For a recent detailed discussion of modal predicates see (Stern 2016). It has been argued that the full force of modal discourse can be restored by using propositional or 'substitutional' quantification or truth predicates (Halbach and Welch 2009). Some remarks on such extensions are given in the next section.

modality comes the expressive power of what Quine called a 'statement operator'. That is, under fairly general assumptions everything that can be expressed using a statement operator $\Box$ can be expressed also with a semantic predicate. As long as φ does not contain $\Box$, a sentence $\Box\varphi$ can be replaced with $\mathsf{Nec}\ulcorner\varphi\urcorner$ where $\ulcorner\varphi\urcorner$ is a quotation or structurally descriptive name for the *sentence* φ. This induces a translation from the entire language with a statement operator to a language with a modal predicate only. However, this reduction or translation does not work for what Quine called 'sentence operators'. When a formula $\Box\varphi$ with free variables is replaced with $\mathsf{Nec}\ulcorner\varphi\urcorner$, the free variables disappear; they are only mentioned, not used in $\mathsf{Nec}\ulcorner\varphi\urcorner$. In other words, Quine's semantic predicates do not permit *de re* modality.

Famously, Quine argued against *de re* modalities and sentence operators that apply to formulæ with free variables – and failed completely at convincing the philosophical community to reject them. *De re* modalities are nowadays used without qualms almost everywhere; and it is hard to imagine many discussions in contemporary philosophy without the additional expressive strength gained from using *de re* modalities. Among the *de re* modalities analyzed in modal or intensional logics are temporal modalities (future and past truth), different epistemic modalities such as knowledge, and several varieties of necessity, especially metaphysical necessity. However, semantic predicates in Quine's sense do not suffice for analyzing *de re* modality; in particular, sentence operators are not reducible to semantic predicates.

I would not like to miss the expressive power of quantifying over the bearers of necessity or of quantifying into modal contexts. The full strength of modal discourse requires both: a predicate conception of modality and *de re* modality. None of Quine's three grades of modal involvement offers the strength of modal predicates *and* that of sentence operators. There is a conspicuous omission from Quine's list of grades of modal involvement: There are no predicates expressing *de re* modalities.[3]

One could hope that nevertheless semantic predicates together with sentential operators in Quine's sense would suffice: For quantified statements of the kind mentioned above a predicate, and for *de re* modality an operator that allows quantifying-in could be used. This is in itself unsatisfactory, because we need to switch between quantified modal logic for dealing with *de re* modalities and modal predicates for dealing with general quantified claims. But there is a class of statements for which both, an operator or *de dicto* predicate are insufficient. For this kind of statement the fourth grade of modal involvement is required. An example is the following claim:

(E) There is an object *o* and there is a *P* such that *P* is necessary for *o*.

[3] Here I refer to *Three grades of modal involvement* only. Later, in *Word and Object* (1960, §41) he considers *de re* modalities as predicates; and earlier in §35 and in (1956) he discusses propositional attitudes *de re*.

This claim may be taken as a statement of some kind of essentialism. Of course metaphysicians will disagree what kind of object P can be: The traditional answer will be that P must be a universal or property, while Quine and others might view P as a formula.

The point is that neither a semantic predicate nor a sentential operator suffice to express (E) without further assumptions or additional devices. This can be seen as follows. If we use an operator as in modal logic, $\exists x \ \Box \ Px$ fails as a formalization of (E), because it only claims that the specific property P is necessary for some object. In order to quantify over P we need a modal predicate (or higher-order quantification or the like). A unary predicate Nec for *de dicto* necessity may also be insufficient, but for a different reason: With a semantic predicate we cannot quantify any longer easily 'into the modal context', that is, over the objects that may have a property necessarily, at least if the semantic predicates apply to sentences. We may be able to express that some sentence of the form Pt is necessary, where t is some closed term; but, of course, the claim (E) just says that P is necessary for some o, whether o has a name or not.[4] Quine's semantic predicates cannot capture this higher kind of modal involvement. Discussions on various versions of essentialism, *ante rem* and *in rebus* conceptions of universals, and so on require *de re* modalities as predicates. Nothing should keep us from quantifying at the same time into modal contexts and over properties or predicates. Only the fourth grade of modal involvement affords this in a straightforward way.

In the following diagram the four grades of modal involvement are ordered by their strength. An arrow from one grade to another means that the grade at the origin of the arrow is stronger than that at the target, that is, the grade at the end of an arrow is reducible to that at its origin.

$$\begin{tikzcd}
& \textit{de re}\text{ semantic predicate} \arrow[dl] \arrow[dr] & \\
\text{semantic predicate} \arrow[dr] & & \text{sentence operator} \arrow[dl] \\
& \text{statement operator} &
\end{tikzcd}$$

At the bottom, as the lowest form of modal involvement, is the conception of operator

[4] Instead of predicates of sentences we could use predicates of Russellian propositions. If we are able to express the operation of applying the property P to the object o, we can use a predicate to say that there is a property P and an object o such that the proposition resulting of applying P to o is necessary. This approach has its own problems, because now *de dicto* modalities are not directly expressible. I will not pursue this strategy any further without claiming that it is not viable. A more detailed discussion would require a formalized theory of Russellian propositions.

modal logic without any quantifying-in.[5] As explained above, treating a modality as a semantic predicate or permitting quantifying-in increases the expressive strength. Neither is a semantic predicate for a modality reducible to the corresponding sentence operator nor *vice versa*: Semantic predicates and sentence operators add different kinds of expressive strength. The latter add *de re* modality, the former the possibility of quantifying over objects to which the modality is ascribed. At the top sits the highest, the fourth grade of modal involvement under which all others can be subsumed, a device that gives both kinds of expressive strength.

A formal framework for this fourth grade of modal involvement will need to deal not only with the traditional problems of modal predicates and *de re* modality as in quantified modal logic; there are also a few additional puzzles. There is not much literature on modal *de re* predicates. George Bealer's contributions (1982, 1993, 1998) are probably the most sophisticated.

DE RE SEMANTIC PREDICATES

Different strategies have been tried to reach the fourth grade of modal involvement. One option is to use a sentence operator for necessity (permitting quantifying-in) and to expand the language with second-order order, substitutional, or propositional quantification. To emulate propositional quantification, quantifiers ranging over sets of possible worlds can be used, as suggested by Kripke (1959). Bull (1969), Fine (1970), Kaplan (1970) and subsequently many others elaborated on Kripke's suggestion. The simple approaches formulated over propositional logic have to be generalized to quantified modal logic to recover the strength of the fourth grade of modal involvement. This will yield some kind of second-order quantified modal logic as in (Williamson 2013). I cannot provide a discussion of these approaches here. I refer the reader to (Halbach and Leigh 2020) for a sketch of reasons why this approach is less promising than the strategy pursued here. One worry is that it will be difficult to express quantification over predicates of arbitrary arity: In higher-order logic one can quantify over propositions (0-place predicates), unary predicates, binary predicates, and so on; but simultaneously quantifying over all predicates of arbitrary arity is not easily possible without further resources.

Another option, with a similar problem, is the use of modal predicates of different arities: There would be a binary predicate that expresses that a single object has a necessary property or that a formula with one free variable is necessary for that object. A ternary predicate would be required for binary relations or formulæ with

[5] When Quine talks about a ordering of grades of modal involvement, he does not have an ordering according to reducibility in our sense in mind. As mention above, a semantic predicate is the lowest form of modal involvement for Quine.

two free variables, and so on. This becomes very clumsy, because there will be a necessity predicate for each arity (or one with variable arity). Moreover, we cannot easily express quantification over predicates with arbitrary arity. Examples of such quantifications will be given below.

The obvious solution consists in adapting Tarski's (1935) trick for truth to modalities. Tarski used a satisfaction predicate applying to formulæ and variable assignments to define the unary truth predicate. Corresponding to the truth predicate we have the unary *de dicto* predicate for necessity; and corresponding to the binary satisfaction predicate, applying to formulæ and variable assignments, we have the binary *de re* necessity predicate applying to formulæ (or universals) and variable assignments.

A language with such a binary modal predicate is expressively richer than one with modal predicates for each arity. By quantifying over formulæ (or corresponding universals) and variable assignments we are not restricted to a fixed arity. For instance, we can express essentialist claims in the style of (E) above in a more cautious way by saying that some objects necessarily stand in some relation. With this formulation we do not commit ourselves to a relation of a specific arity. By using a binary predicate applying to predicates of arbitrary arity and sequences of objects, this can be expressed with the formula $\exists x\, \exists y\, \mathsf{Nec}(x, y)$ or $\exists x\, \exists y\, (\mathsf{For}(x) \wedge \mathsf{As}(y) \wedge \mathsf{Nec}(x, y))$, where $\mathsf{For}(x)$ expresses that x is a formula (relation) and $\mathsf{As}(y)$ that y is a variable assignment.

Using a predicate like Nec we can also express *de re* factivity. This is the claim that if a formula (or relation) is necessary of some objects, then it is true of those objects. This can be parsed as follows: If a formula x is necessary of a string y of objects, then x is satisfied by y, or formally $\forall x\, \forall y\, (\mathsf{Nec}(x, y) \rightarrow \mathsf{Sat}(x, y))$. Again the quantifiers may be restricted to formulæ and assignments as above. The factivity of *de re* knowledge can be expressed by replacing the binary predicate for necessity with a corresponding binary predicate for *de re* knowledge. *De re* factivity implies *de dicto* factivity, that is, the claim that whatever is necessary is true, but the converse implication does not hold.

At this point it is worth emphasizing the following: I am far from claiming that binary modal predicates of the kind described above are close to the methods of expressing the fourth degree of modal involvement in natural languages. A linguistic analysis is not the target of the present paper, although there are clearly some very interesting differences between modalities in the natural languages. For instance, temporal notions are usually expressed in the languages with which I am familiar by modifying the verb, either directly or with an auxiliary verb. The tense of a verb is probably most faithfully represented in a formal language by a tense operator, that is, a sentence operator in Quine's sense, not by a primitive predicate for future or past truth. Moreover, in many cases modalities can be expressed in many different ways in natural languages. We have modal operators such as *necessarily* as well as predicate phrases such as *is necessary* in English. I aim to provide a formal framework that allows us to analyze philosophical modal discourse in its full strength; this requires quantification over universals or formulæ and variable assignments without fixed arity. If possible, I prefer to be parsimonious and not introduce devices that are reducible to binary predicate for *de re* modalities, even if there are analogues of these devices in natural languages.

All my claims so far about the strength of various devices for expressing modalities rely on appeals to some informal semantics. Of course, a plethora of various semantic systems exists for modal logics; but they are largely missing for corresponding binary predicates. Therefore, to substantiate my claims and to assess the prospects of analyzing *de re* modalities using binary predicates of the kind described, a formal semantics for languages with such predicates are needed.

POSSIBLE WORLDS SEMANTICS

In this section I extend and adapt the possible worlds semantics for a unary modal predicate developed by Halbach, Leitgeb and Welch (2003), Halbach and Welch (2009), and Halbach and Leigh (2020) to a semantics for the binary predicate $\mathsf{Nec}(x, y)$. I think of the semantics given below as a starting point. Several assumptions may be tweaked. Moreover, the account below can be generalized in many ways to different modalities and multimodal settings.

As pointed out at the end of the previous section, a formal semantics is needed in order to substantiate various claims about the strength of modal devices such as the binary modal predicate Nec, unary semantic predicates, the operator □ of modal logic as sentence and statement operators, and perhaps further expressions. This requires a certain degree of adequacy of the semantics for which I will not argue.

There are further reasons for developing possible worlds semantics for Nec. In particular, I would like to transfer insights that have been obtained using possible worlds semantics for the modal operator □. Giving up possible worlds semantics together with the operator conception of modality would be a huge loss. I do not intend to reject sentence and statement operators. My complaint is that they cannot capture the full force of modal talk; but as long as this force is not required, modal operators and their accompanying possible worlds semantics can shed light on many questions and puzzles.

The aim of this section is to define a possible worlds models such that $\mathsf{Nec}(\ulcorner\varphi(x_1,\dots,x_n)\urcorner, a)$ holds at a world w iff $\varphi(x_1,\dots,x_n)$ holds at all worlds v with wRv, if the free variables $x_1,\dots,x_n$ are assigned values $a(x_1),\dots,a(x_n)$. Here $\ulcorner\varphi(x_1,\dots,x_n)\urcorner$ is a name for the formula $\varphi(x_1,\dots,x_n)$, and $a(x_k)$ the value assigned to the variable x_k by a. I will be sloppy with notation and, for instance, use a in the object language, even though we might lack a name for it.

When setting up possible-worlds semantics for Nec, the same choices as for quantified operator modal logic have to be faced, and, as is well-known, there are many (see, for instance, Garson 2001). In particular, we have to decide whether to opt for a possibilist treatment of quantifiers with a fixed domain for all worlds or for actualism with domains that can vary between world.

There are also decisions that do not arise in quantified operator modal logic. On the predicate approach, we can quantify over the objects that can be necessary and they can inhabit possible worlds along with other objects (unlike propositions conceived as sets of possible worlds, which cannot themselves be elements of a world). Which of these objects exist would differ from world to world, if we decided to ascribe necessity to sentence tokens. I think of them as types and assume that they exist in all worlds. Metaphysical questions and decisions of this kind are suppressed in the usual variants of quantified modal logic; on the predicate approach they can be addressed in the object language.

Finite sequences of objects are needed as variable assignments for *de re* modalities. I assume that all worlds are closed under the formation of sequences. This is again a strong ontological assumption, which I endorse for metaphysical necessity, but not for other modalities. Only objects existing in the world can form part of a sequence in that world. I discuss potential problems arising from this assumption after having introduced formal possible worlds semantics for languages with modal predicates.

For my purposes here there is no need to be very specific about the language. The language should contain vocabulary that allows one to talk abouth either expressions in the language or corresponding 'universals'. I have a preference for expansions of languages described in (Halbach and Leigh 2020), but expansions of arithmetic or set theory are equally suitable. The language should provide vocabulary for talking about syntax or universals, finite sequences of objects, and possibly other 'ordinary' objects. I use the syntactic approach and terminology. The reader preferring universals will have to replace 'sentence' with 'proposition', 'formula with one free variable' with 'property', and so on. Given that we need arbitrarily long finite sequences of arbitrary objects as variable assignments anyway, we can understand expressions as a specific kind of such sequences, namely those with symbols as members of the sequence.

Each expression in the non-logical vocabulary falls into at least one of three groups:

1. *The syntactic vocabulary* contains at least a unary predicate that applies exactly to all expressions of the language. Quantifiers relativized to this predicate range exactly over all syntactic objects. The other syntactic vocabulary may express operations such as concatenation and substitution and allow one to define grammatical categories such as variables, predicate expressions, connectives etc.

2. *The sequence vocabulary* includes again a relativizing predicate that applies exactly to finite sequences. Further vocabulary may express concatenation of finite sequences, projections etc.

3. *The 'contingent' vocabulary* contains all remaining non-logical expressions. A relativizing predicate is not required, because 'contingent' objects are exactly those that are not syntactic or sequences.

As mentioned above, it is not assumed that finite sequences and expressions do not overlap. However, the contingent objects are exactly those that are neither expressions nor sequences. The use of the term 'contingent' may be somewhat misleading. It only means that no assumptions about the interpretation of the contingent vocabulary are made in possible worlds semantics. This does not rule out that the contingent vocabulary is interpreted at all worlds in the same way in a given model. For instance, some vocabulary of pure mathematics may form part of the contingent vocabulary, and we may confine our attention to interpretations that assign the same extensions to these expressions at all possible worlds. But this restriction is not imposed by possible worlds semantics, but rather by other considerations.

The following definition of a pre-model is close to the definition of a possible worlds model in modal logic, the main difference being that it is assumed that all expressions exist in every world and that worlds are closed under the formation of finite sequences.

DEFINITION 1. A triple $\langle W, R, I \rangle$ is a *pre-model* iff

1. W is a non-empty set.
2. R is a binary relation on W.
3. I is function that assigns each world $w \in W$ a domain $\mathcal{D}_w$. $\mathcal{D}_w$ contains (codes of) all strings of $\mathcal{L}_N$-symbols, possibly further 'contingent' objects and all finite sequences of all objects in $\mathcal{D}_w$. Moreover I provides interpretations of the contingent vocabulary over the domain $\mathcal{D}_w$.

Of course $\langle W, R \rangle$ is a frame in the usual sense in modal logic. Each world contains contains all $\mathcal{L}_N$-expressions and therefore infinitely many objects. Moreover, finite sequences can be formed from all objects in the world. Sequences themselves can be elements of sequences. Therefore each domain $\mathcal{D}_w$ is closed recursively under the formation of sequences. I do not make any assumption on whether expressions conceived as sequences of symbols are identical with these finite sequences.

Consequently, the domain $\mathcal{D}_w$ of each world $w \in W$ consists of three types of objects:

1. $\mathcal{D}_w$ contains all strings of symbols.
2. $\mathcal{D}_w$ contains arbitrarily many further 'contingent' objects.

3. $\mathcal{D}_w$ contains all finite sequences of objects from 1, 2, and 3. This includes mixed sequences that have syntactic, contingent, and finite sequences as entries. Thus the domain of each world is defined recursively.

A pre-model provides interpretations for the entire language at each world $w \in W$, except for the binary necessity predicate Nec. At any world the syntactic and sequence vocabulary is interpreted in the standard way. This means that the syntactic vocabulary is interpreted in the same way at all worlds. However, worlds differ with respect to the objects that exist in them and consequently also with respect to the sequences that exist in them. Thus the interpretation of the vocabulary about sequences at a world depends only on the domain $\mathcal{D}_w$ of that world, because only sequences whose ultimate components exist at that world also exist at that world.

If φ is a formula not containing the modal predicate Nec and a a variable assignment over $\mathcal{D}_w$, I write $\langle W, R, I \rangle \vDash_w \varphi[a]$ to indicate that the formula φ is satisfied by the variable assignment a at world w. As mentioned above, variable assignments are conceived as finite sequences and thus the question arises about situations where φ contains variables outside the domain of a. I could rule this out without real loss of generality, but prefer to stipulate that $a(x)$ is always a certain expression, say $\neg$ and thus effectively make every variable assignment a total function from the set of variables. By our stipulations, $\neg$ is in the domain of every world.

What is missing from this account is an interpretation of the *de re* modal predicate Nec. The interpretation B of the binary predicate symbol Nec is provided separately from I: B is a function that assigns to each world a binary relation (set of ordered pairs) on the set of objects that exist at w. I write $\langle W, R, I, B \rangle \vDash_w \varphi[a]$ iff φ holds at world w under the variable assignment a if Nec is interpreted according to $B(w)$. I now define what a possible-worlds model ('pw-model' for short) is.

DEFINITION 2. $\langle W, R, I, B \rangle$ is a pw-model iff $\langle W, R, I \rangle$ is a pre-model and B a function satisfying the following properties:

(i) B is a function assigning to each world $w \in W$ a set of pairs $\langle \varphi, a \rangle$ such that φ is a formula and a is a finite sequence in $\mathcal{D}_w$.

(ii) $\langle \varphi, a \rangle \in B(w)$ iff forall v (if wRv then $\langle W, R, I, B \rangle \vDash_v \varphi[a]$)

In (ii) the variable assignment a in $[a]$ belongs to the metatheory, while pair $\langle \varphi, a \rangle$ is a possible element of the extension $B(w)$ of Nec at w and thus a an element of $\mathcal{D}_w$. That is, variable assignments belong to both, the object and metatheory.

Condition (ii) of definition 2 forces the interpretation of Nec as truth in all accessible worlds. Suppressing the variable assignment, it means that a sentence φ is in the extension of Nec at a world w iff φ is true in all worlds accessible from w. By condition (i), if $\langle \varphi, a \rangle \in B(w)$, then $a \in \mathcal{D}_w$ and therefore $a(x) \in \mathcal{D}_w$ for all variables x. That is, a variable assignment in a world w assigns variables only values that exist in w. But of course $a(x)$ need not exist in another world, that is, we may have $a(x) \notin \mathcal{D}_v$ for some other world $v \neq w$. In particular, when we pass from a world w to all worlds v that can be seen by w, a variable assignment with $\langle \varphi, a \rangle \in B(w)$ might assign a value to the variable x that is not in $\mathcal{D}_v$. Thus we need to make a stipulation about how to understand the right-hand-side of (ii), that is, $\langle W, R, I, B \rangle \vDash_v \varphi[a]$ if φ contains a free variable x such that $a(x) \notin \mathcal{D}_v$. The situation is similar to free logic and individual constants that do not denote. There are various options. Here I adopt the policy of negative free logic. If φ is an atomic formula with a free variable x such that $a(x) \notin \mathcal{D}_v$, then $\langle W, R, I, B \rangle \nvDash_v \varphi[a]$. For instance we have $\langle W, R, I, B \rangle \nvDash_v x = x\ [a]$. Of course, we still have $\langle W, R, I, B \rangle \vDash_v \forall x\, x = x\ [a]$ as quantifiers range over objects in $\mathcal{D}_v$.

The situation is different from ordinary quantified modal logic, because we have variable assignments not only in the metatheory (as in quantified modal logic), but we are also talking about variable assignments in the object language. In $\langle W, R, I, B \rangle \vDash_v \varphi[a]$ the variable assignment a is in the metatheory of course, while in $\exists y\, \mathsf{Nec}(x, y)$ we are quantifying over variable assignments in the object language. The metatheory is extensional. Every variable assignment that exists at some world is also available in the metatheory. But of course not every variable assignment in our metatheory exists in every world: only if all values exist in w, the variable assignment exists in w.

In the usual possible-worlds semantics for modal operators, one defines recursively the semantics for $\Box$ from the interpretation of the other vocabulary at each world. Definition 2, in contrast, does *not* imply that for every pre-model $\langle W, R, I \rangle$ there is exactly one B such that $\langle W, R, I, B \rangle$ is a pw-model. In fact, both, existence uniqueness fail. Halbach et al. (2003) and Halbach and Leigh (2020) provide more information on existence and uniqueness conditions. I mention only some observations that apply to the present framework.

A relation R is converse wellfounded iff there is no infinitely ascending sequence $w_1 R w_2 R w_3 \ldots$ of objects.

LEMMA 3. If the accessibility relation R of a pre-model $\langle W, R, I \rangle$ is converse wellfounded on W, then there is a unique B such that $\langle W, R, I, B \rangle$ is a pw-model.

The lemma is proved by defining B on all $w \in W$ inductively on the converse of R, starting from the dead ends of R, that is, worlds $w \in W$ such that there is no $v \in W$ with wRv. At a dead end $w \in W$, we have $\langle \varphi, a \rangle \in B(w)$ for all formulae φ and variable assignments $a \in \mathcal{D}_w$, because trivially $\langle W, R, I, B \rangle \vDash_v \varphi[a]$ for all worlds v with wRv, as there are no such worlds. Once $B(v)$ has been defined for all worlds v with wRv, we can set

$$B(w) \coloneqq \bigcap_{wRv} \{ \langle \varphi, a \rangle : \langle W, R, I, B \rangle \vDash_v \varphi[a] \} \cap \mathcal{D}_w$$

The intersection with $\mathcal{D}_w$ ensures that we only include variable assignments that exist at world w. If R is converse wellfounded, this definition fixes $B(w)$ for all $w \in W$.

Lemma 3 gives us a sufficient condition for the existence of a pw-model $\langle W, R, I, B\rangle$ based on a given pre-model $\langle W, R, I\rangle$. I turn now to necessary conditions.

Under fairly general circumstances, there cannot be a pw-model $\langle W, R, I, B\rangle$ with a reflexive or symmetric accessibility relation R. This is a direct consequence of the paradoxes. Since φ is a sentence, the variable assignment a is irrelevant and $\forall a\, \mathsf{Nec}(\ulcorner\varphi\urcorner, a)$ expresses *de dicto* necessity of φ in the same way the unary truth predicate can be defined as the satisfaction of a sentence by all variable assignments. If R is reflexive, we have the following for all sentences φ and worlds $w \in W$, as in operator modal logic:

(1) $$\langle W, R, I, B\rangle \vDash_w \forall a\, \mathsf{Nec}(\ulcorner\varphi\urcorner, a) \rightarrow \varphi$$

Using the diagonal lemma, a sentence λ can be obtained such that $\lambda \leftrightarrow \neg\forall a\, \mathsf{Nec}(\ulcorner\lambda\urcorner, a)$ holds at all worlds. Of course, this is the liar sentence for the 'truth predicate' $\forall a\, \mathsf{Nec}(x, a)$. Combining this with (1) for λ yields the following for all $w \in W$:

(2) $$\langle W, R, I, B\rangle \vDash_w \lambda$$

Since λ holds at all worlds, $\forall a\, \mathsf{Nec}(\ulcorner\lambda\urcorner, a)$ holds at all worlds. This clashes with following consequence of the diagonal property of λ:

$$\langle W, R, I, B\rangle \vDash_w \neg\forall a\, \mathsf{Nec}(\ulcorner\lambda\urcorner, a)$$

This contradiction is only a variant of Montague's (1963) theorem, as explained in (Halbach and Leigh 2020). This can be generalized as follows:

THEOREM 4. If the relation R is converse illfounded on W, the syntactic vocabulary sufficiently expressive, and the contingent vocabulary contains at least one sentential parameter, then there is a I such that $\langle W, R, I\rangle$ is a pre-model, but there is no B such that $\langle W, R, I, B\rangle$ is a pw-model.

For a detailed proof the reader is referred to (Halbach and Leigh 2020). First it is shown that one can define the modal predicate Nec^* associated with the transitive closure of R. Using the diagonal lemma one proves that Löb's theorem holds for this predicate Nec^*. Löb's theorem is of course a principle of transfinite induction that may fail if R^* is not converse wellfounded. To show this, one can choose an interpretation I that makes the sentential parameter true at all converse wellfounded worlds, but false at the converse illfounded worlds. If there is no contingent vocabulary (the sentential parameter in the theorem), there can be a converse illfounded R such that $\langle W, R, I, B\rangle$ is a pw-model and the situation becomes complicated. See (Halbach et al. 2003).

Theorem 4 means that the accessibility relation cannot be total on all worlds; and we cannot expect modal predicates to satisfy the schemata of the modal systems T, B, S4, or S5 for all sentences (especially those sentences without equivalent in ordinary modal operator logic such as sentences obtained by Gödel's diagonal lemma). If we are confined to converse wellfounded frames for possible worlds semantics of predicates, then one might suspect that we cannot have a reasonable semantics for Nec as metaphysical necessity. Perhaps the accessibility relation need not be total and, perhaps, not even transitive for metaphysical necessity; but at last it should be reflexive. This is ruled out by Theorem 4.

The reader may wonder whether this shows that the predicate approach is to be rejected. In the light of the paradoxes Montague (1963) rejected 'syntactic' treatments of modality, that is, treatments of modality as predicates of sentences. The restrictions in theorem 4 only spell out the consequences of the paradoxes as restrictions on the accessibility relations and they thereby systematize them. Should one agree with Montague, abandon predicate approaches, and resort to operator modal logic? The price for this move will be a significant reduction in expressive power of our language. This matters most to philosophers, who are interested in the quantified claims mentioned in the first section.

The situation can be compared with that for truth: We can reject the predicate conception of truth and opt for an operator treatment. This will result the trivial modal logic with $\Box\varphi \leftrightarrow \varphi$ as characteristic axiom. Although this move immediately blocks the paradoxes, it has not been very popular. What saved truth from the fate of metaphysical necessity and other modalities is that the truth predicate becomes trivial and boring, once it is stripped of its ability to express generalizations. When metaphysical necessity is treated in the same way, there are still many interesting things to say. But that is not a good justification for accepting the weakening of expressive power caused by the transition from $\Box$ to Nec.

Rejecting modal predicates because of the paradoxes is as sensible as rejecting the theory of distances because of Zeno's paradoxes and rejecting set theory because of Russell's and Burali–Forti's paradox. We may have to revise or refine some of our naive expectation, but we will gain expressive strength in return. This strength is needed for general claims in philosophy. If we give it up in favour of modal operators, the way philosophy is done will have to be changed profoundly. Therefore, we should seek solutions to the paradoxes of modal predicates in the same spirit as in the case of truth.

Unless we are prepared to restrict ourselves to converse wellfounded accessibility relations, we need to adapt and tweak possible worlds semantics. Stern (2016) discussed various strategies. First, one can apply the usual techniques known from the semantic paradoxes. In particular, classical logic can be abandoned. Halbach and Welch (2009) employed Kripke's (1975) fixed-point semantics for their possible worlds semantics without really defending it as a real solution.

Alternatively, condition (ii) of Definition 2 can be weakened: We do no longer stipulate for *all* sentences that $\mathsf{Nec}\ulcorner\varphi\urcorner$ holds at w iff φ holds in all accessible worlds; we merely stipulate this for certain sentences. Condition (ii) should be satisfied at least for all those sentences φ that are expressible with an operator $\Box$. This guarantees that the possible worlds semantics for predicates is not 'worse' than possible worlds semantics for the operator $\Box$. The class of sentences expressible with an operator $\Box$ can be defined as follows using a recursively defined mapping I from the language with the operator $\Box$ to the language with the corresponding predicate Nec. Since we deal with *de re* modality, we need to take care of the free variables in the scope of $\Box$. Pick some fixed variable assignment a_0. The operation that changes the assignment of a given variable v_n, so that the object x is assigned to v_n can be expressed in the object language and I write $a_0(x/\ulcorner v_n\urcorner)$ for this operation.

(i) If φ does not contain $\Box$, then $I(\varphi) = \varphi$.

(ii) $I(\varphi \wedge \psi) = I(\varphi) \wedge I(\psi)$, $I(\neg\varphi) = \neg I(\varphi)$, $I(\forall x\, \varphi) = \forall x\, I(\varphi)$, and so on.

(iii) $I(\Box\varphi(x_1 \ldots x_n)) = \mathsf{Nec}(\ulcorner I(\varphi)\urcorner, a_0(x_1/\ulcorner x_1\urcorner, \ldots x_1/\ulcorner x_1\urcorner))$, where $a_0(x_1/\ulcorner x_1\urcorner, \ldots x_1/\ulcorner x_1\urcorner)$ is obtained by iterated application of the operation mentioned above. Only finitely many such applications are required, because φ contains only finitely many variables.

This embedding of the operator languages into the language with a predicate is well-known, and its origin can be traced back at least to (Carnap 1934, IV.B.e).[6] Here I have only added the free variables. The class of formulæ expressible with an operator $\Box$ is now simply the set of all $I(\varphi)$ such that φ is a formula of the language with the operator $\Box$ only. If condition (ii) of Definition 2 is restricted to such sentences, there are pw-models for arbitrary frames.

Sentences and formulæ generated with Gödel's diagonal construction cannot be obtained with I from sentences of operator modal logic, but also not general formulæ of the form $\forall x\,(\chi(x) \to \mathsf{Nec}\,x)$. Some but not all formulæ of the latter kind can be unproblematic and be included as permissible instances of condition (ii) of Definition 2.

A more comprehensive class of permissible instances of condition (ii) of Definition 2 can be obtained by singling out the 'grounded' formulæ and sentences, defined along the lines of Kripke's (1975). If we apply this account to formulæ in general, the notion

[6] Carnap's translations is different and there are several variations. What caused problems especially in early variants were problems with the iteration of Nec.

of groundedness relative to a variable assignment will have to be defined. In the brief following comments I restrict myself to sentences. The groundedness approach has the disadvantage that the set of such sentences is no longer definable in the object language (under fairly general assumptions). Moreover, whether a sentence is grounded may depend on contingent factors. If all pw-models with this restricted condition (ii) are considered, many general principles such as the distribution of necessity over → in pw-models without dead ends have to be abandoned; they will only hold for grounded sentences. This flies in the face of the predicate approach: After all, I adopted the predicate approach in order to express quantified principles. Not being able to get them as quantified generalizations seems odd. At any rate, the desirable models may be sought among those that are pw-models in the weaker sense of satisfying (ii) only for grounded sentences. A different option may be to introduce a new primitive predicate for determinacy or groundedness into the language as in (Fujimoto and Halbach 2019) for truth.

An elegant and philosophically useful way to obtain generalization without compromising too much on the expected properties of possible worlds semantics is given by Stern (2014*a*,*b*, 2016). The idea is to state at least quantified principles with a truth predicate, as we routinely do in informal philosophical discourse. For this strategy, however, a sophisticated theory of truth will be required that avoids the paradoxes.

The challenge of the paradoxes arises for *de dicto* and *de re* modalities, although some paradoxes require quantification. In particular, McGee's ω-inconsistency (1985) and the Yablo–Visser paradox are of this kind (Visser 1989, Yablo 1985, 1993). A binary predicate may lend itself to the study of these paradoxes. Finally, the move to a binary predicate can be used to recover diagonalization and other syntactic operations without any explicit syntax-theoretic axioms (Halbach and Zhang 2016, Halbach and Leigh 2020). I will not pursue this topic here, but rather turn to the application of the fourth grade of modal involvement to an issue in metaphysics.

A PROBLEM FOR ACTUALISM

In this section I illustrate my claim that using the fourth grade of modal involvement can shed new light on topics and discussions in metaphysics, using an argument by Bealer (1993). In many ways the predicate conception of modalities is a more versatile framework for modal metaphysics than first-order quantified modal logic. The ontology of the objects to which the modalities are ascribed is not hidden away in the metatheory. The expressive power generated by the predicate conception permits an analysis of many questions in the object theory instead of the metatheory, which is usually set-theoretic and extensional.

For instance, it has become common to assume that propositions, as the objects that can be necessary, do not exist in a possible world; as sets of possible worlds, they live in the 'modal æther' (Forster 2005) and are thereby incomparable to other normal objects. On the predicate conception, we talk about the objects that can be necessary in the object language, and we have to decide whether they exist in a world, as sets of worlds, or wherever. In the outline above they exist in all worlds. But this is only one option; moreover, they are assumed to have the structure of formulæ. A metaphysician more serious about proposition may structure them in a different way. Of course, the ontology of propositions has received much attention, and I will not pursue this topic here. Instead I focus on the second argument place of the binary modal predicate Nec, the variable assignments. When they are discussed in the usual set-theoretic metatheory, they receive little attention. If they can be talked about in the object language, they become interesting and puzzling. In particular, they can shed new light on the discussion between actualism and possibilism (or possibilism and necessitism in Williamson's 2013 terminology). The semantics outlined above is actualist in the sense that the domains $\mathcal{D}_w$ can vary between worlds w; at each w the quantifiers range only over $\mathcal{D}_w$. Possibilist semantics is a special case: Nothing rules out that $\mathcal{D}_v = \mathcal{D}_w$ for all $v, w \in W$, that is, constant domain semantics is a special case of the actualist semantics.

The argument in this section is not intended as a definitive argument in favour of possibilism or some other position; the section merely demonstrates the effects of injecting objects, in particular, variable assignments from the metalanguage into the object language and how this move can shed light on metaphysical questions.

Bealer (1982, 1993) and others have used arguments of the kind discussed in this section to argue for an *ante rem* conception of universals. I could follow Bealer here, but in the present setting – which is different from Bealer's – it would mean that variable assignments exist prior to their elements. There would be variable assignments that assign non-existing objects to some variables, which is hardly acceptable.

What is going to follow is a partial recantation of (Halbach and Sturm 2004). The discussions there and Bealer's (1993) paper suffer from the absence of formal semantics. With a framework such as the one outlined above this family of puzzles can now be discussed in a more rigorous way.

As I mentioned already, in quantified modal logic variable assignments are used only in the set-theoretic, extensional metatheory, where questions about possibilism and actualism cannot arise. In our semantics the variable assignments are pushed into the object language that contains modalities. In the presence of modality, the theory of sequences (and sets) becomes more complicated in actualist semantics: Sequences exist only at a world, if all its members exist at that world. This assumption is built into our semantics. In the following example our semantics may yield an unexpected result because of this assumption.

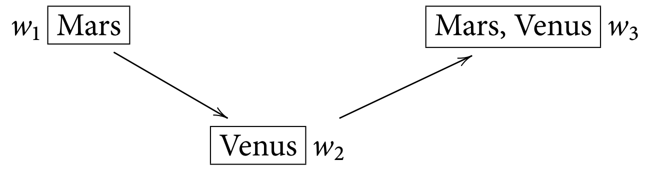
(PLA) There is a planet, and instead of this planet another planet could have existed which could have coexisted with the first.

There may be more than one reading of this sentence. The reading I have in mind, can be made explicit using possible worlds: In our world w_1 there is a planet A and there is a possible world w_2 accessible from our world where A does not exist, but planet B does; moreover, there is another world w_3 accessible from w_2 where both planets A and B exist. The reading is captured by the following formalization in quantified modal logic:

$$(\text{PLA}\Box) \qquad \exists x \Big(\mathrm{Pla}\, x \wedge \Diamond\big(\neg\exists z\, z = x \wedge \exists y\, (\mathrm{Pla}\, y \wedge \Diamond(\exists z\, z = x \wedge \exists z\, z = y))\big)\Big)$$

The formula $\exists z\, z = x$ expresses that x exists. That is, $\exists z\, z = x$ becomes false at a world w under a variable assignment a if $a(x) \notin \mathcal{D}_w$. This is the case for our semantics for Nec, and we assume that the same holds for the possible worlds semantics for the operator □.

A minimal possible worlds model $\mathcal{M}$ of (PLA□) looks as follows:



The three worlds of the model are related by the accessibility relation as in the diagram. The domain of w_1 is {Mars}, and so on for the other worlds. The unary predicate $\mathrm{Pla}\, x$ applies at a world to all objects that exist in that world. It is easily seen that then (PLA□) is true at w_1 in this model.

The crucial feature of the example is that the first quantifier $\exists x$ binds an occurrence of x in the scope of *two* possibility operators ◇. The witness of the existentially quantified sentence (PLA□) at w_1 is Mars, which exists at w_1, but not in the next world w_2; it exists again in the third world w_3. In operator modal logic this does not pose a problem; it is not required that x exists in the intermediate world w_2 for the existential quantifier $\exists x$ to bind an occurrence of x in the scope of two modal operators. As I will show, however, it does cause a problem if a modal predicate is used. The formalization of (PLA) with a binary predicate Nec instead of the modal operator □ is false at w_1 in the pw-model corresponding to the model $\mathcal{M}$ above, while (PLA□) is true at w_1, as pointed out above.

To substantiate my claim that the predicate formalization gives a different truth value, I specify the predicate formalization of PLA and the (predicate) pw-model corresponding to the operator model $\mathcal{M}$ above.

First, I sketch how to transform $\mathcal{M}$ into a pw-model $\langle W, R, I, B\rangle$ for the language with Nec instead of $\Box$. W and R stay the same. The domain $\mathcal{D}_w$ for one of the three worlds $w \in W$ is obtained by adding all syntactic objects to the domain of w in $\mathcal{M}$ and then closing under the formation of sequences. Thus $\mathcal{D}_{w_1}$ contains Mars, but not Venus, $\mathcal{D}_{w_2}$ only Venus, while $\mathcal{D}_{w_3}$ contains both. Consequently, $\mathcal{D}_{w_2}$ does not contain any sequences involving Mars. Pla applies at w exactly to all planets in w. The existence of such a model is guaranteed by lemma 3 above, because the accessibility relation is converse wellfounded.

The formalization of (PLA) with the model predicate Nec looks as follows, if $\mathsf{Pos}(x, y)$ is an abbreviation for $\neg\mathsf{Nec}(\underset{.}{\neg}x, y)$ where $\underset{.}{\neg}$ represents the function of negating a sentence. Thus, Pos stands for possibility.

(PLA N)
$$\exists x \left(\mathrm{Pla}\, x \wedge \mathsf{Pos}(\underbrace{\ulcorner\neg\exists z\, z=x \wedge \exists y\,(\mathrm{Pla} y \wedge \mathsf{Pos}(\ulcorner\exists z\, z=x \wedge \exists z\, z=y\urcorner, \langle \frac{x}{\ulcorner x\urcorner}, \frac{y}{\ulcorner y\urcorner}\rangle))\urcorner}, \langle \frac{x}{\ulcorner x\urcorner}\rangle)\right)$$

The expression $\langle \frac{x}{\ulcorner x\urcorner}, \frac{y}{\ulcorner y\urcorner}\rangle$ in the underbraced quotation name and the less complex $\langle \frac{x}{\ulcorner x\urcorner}\rangle$ require an explanation. I assume that the vocabulary permits to describe the variable assignment assigning an object to a specific variable. Here I have treated $\langle \frac{x}{x_1}, \frac{y}{y_1}\rangle$ like a function expression with four free variables x, x_1, y, and y_1 expressing the function that gives applied to objects x and y and variables x_1 and y_1 the relevant variable assignment; but this may have to circumscribed if no suitable function symbol is available. In addition, we need the quotation function sending an expression (here a variable) to a name for that expression. In arithmetical contexts the numeral function serves this purpose.

Since there are no further free variables, a variable assignment of length 2 in the underbraced name suffices. As mentioned above, $\neg$ is stipulated to be the value of all other variables. The 'outer' variable assignment $\langle \frac{x}{\ulcorner x\urcorner}\rangle$ has only length 1, because the preceding formula in corners has only x free. In $\frac{x}{\ulcorner x\urcorner}$ the upper occurrence of x is used, while $\ulcorner x\urcorner$ is only a mentioning of the same variable; of course, different variables could be used. If the bound variable x in the formula above is renamed as v and the relativization to planets is omitted for readability, the following formula is obtained:

$$\exists v\, \mathsf{Pos}(\ulcorner\neg\exists z\, z=x \wedge \exists y\, \mathsf{Pos}(\ulcorner\exists z\, z=x \wedge \exists z\, z=y\urcorner, \langle \frac{x}{\ulcorner x\urcorner}, \frac{y}{\ulcorner y\urcorner}\rangle)\urcorner, \langle \frac{v}{\ulcorner x\urcorner}\rangle)$$

Of course, any occurrences in the quotation name are not affected by the renaming and x is retained.

In a nutshell, the reason why (PLA N) fails at w_1 is the following: If (PLA N) were true at w_1, there would have to be a variable assignment in $\mathcal{D}_{w_2}$ that assigns Mars to x and Venus to y. But there is no such variable assignment in $\mathcal{D}_{w_2}$, because Mars does not exist in w_2, that is Mars is not an element of $\mathcal{D}_{w_2}$. In fact, such a variable assignment exists only at w_3 where both, Mars and Venus exist.

A more explicit version of this argument can be given as follows. Let $\langle W, R, I, B\rangle$ be the pw-model described above. To show that (PLA N) fails at w_1, assume to the contrary that it is true at w_1. Since Mars is the only potential witness of the existential quantifier, also the following must hold:

$$\langle W, R, I, B\rangle \vDash_{w_1} \mathsf{Pos}\big(\ulcorner \neg\exists z\, z{=}x \wedge \exists y\, (\mathrm{Pla}y \wedge \mathsf{Pos}(\ulcorner \exists z\, z{=}x \wedge \exists z\, z{=}y\urcorner, \langle \tfrac{x}{\ulcorner x\urcorner}, \tfrac{y}{\ulcorner y\urcorner}\rangle)))\urcorner, \langle \tfrac{x}{\ulcorner x\urcorner}\rangle\big) \quad [\langle \tfrac{\text{Mars}}{\ulcorner x\urcorner}\rangle]$$

That is, we assume that the formula $\mathsf{Pos}(\ldots)$ holds at world w_1 under the variable assignment that assigns Mars to x. The expression in square brackets is the variable assignment in the metalanguage, for which we conveniently use the same notation as in the object language. It follows from the assumption that the formula denoted by the quotation name must be satisfied by Mars for x at the world accessible from w_1:

$$\langle W, R, I, B\rangle \vDash_{w_2} \neg\exists z\, z{=}x \wedge \exists y\, \big(\mathrm{Pla}y \wedge \mathsf{Pos}(\ulcorner \exists z\, z{=}x \wedge \exists z\, z{=}y\urcorner, \langle \tfrac{x}{\ulcorner x\urcorner}, \tfrac{y}{\ulcorner y\urcorner}\rangle)\big) \quad [\langle \tfrac{\text{Mars}}{\ulcorner x\urcorner}\rangle]$$

Hence both conjuncts must hold at w_2. The first conjunct expresses that x does not exist. Because Mars fails to be in $\mathcal{D}_{w_2}$, the first conjunct does hold indeed.

The second conjunct is an existentially quantified sentence. The only witness that cold make the second conjunct true is Venus, because the witness must be a planet and Venus is the only planet that exists in $\mathcal{D}_{w_2}$. Therefore the following must obtain:

$$\langle W, R, I, B\rangle \vDash_{w_2} \mathrm{Pla}y \wedge \mathsf{Pos}(\ulcorner \exists z\, z{=}x \wedge \exists z\, z{=}y\urcorner, \langle \tfrac{x}{\ulcorner x\urcorner}, \tfrac{y}{\ulcorner y\urcorner}\rangle) \quad [\langle \tfrac{\text{Mars}}{\ulcorner x\urcorner}, \tfrac{\text{Venus}}{\ulcorner y\urcorner}\rangle]$$

This implies the following:

$$\langle W, R, I, B\rangle \vDash_{w_2} \exists z\, z = \langle \tfrac{x}{\ulcorner x\urcorner}, \tfrac{y}{\ulcorner y\urcorner}\rangle \quad [\langle \tfrac{\text{Mars}}{\ulcorner x\urcorner}, \tfrac{\text{Venus}}{\ulcorner y\urcorner}\rangle]$$

That is, at w_2 there is a finite sequence containing Mars and Venus, contrary to our assumption that finite sequences in a world can contain only objects existing in that world. Therefore, the initial assumption is refuted and, therefore, (PLA N) fails at w_1, while the operator version (PLA$\Box$) holds at w_1 in the corresponding operator model $\mathcal{M}$. Intuitively, the operator version yields the expected result, while the predicate version does not.

The problem for actualism is not caused by any restrictions on the accessibility relation. It cannot be avoided by replacing the accessibility relation R above with its transitive closure. We cannot have a total accessibility relation because of theorem 4; but even if diagonalization were banned and total accessibility relations permitted, the problem for actualism would remain.

Various strategies to address the problem were discussed by (Bealer 1993) and (Halbach and Sturm 2004). I do not provide a thorough discussion, but only sketch some of them.

First, an actuality operator or predicate cannot overcome the problem: The variable assignment assigning Mars to x and Venus to y could be claimed to exist in the actual world w_1 instead of the world w_2; but the variable assignment cannot exist at w_1 either, because Venus is not in w_1. Bealer (1993) made this observation already for universals rather than variable assignments.

The second strategy sounds desperate: Variable assignment assigning non-existing objects to variables could be permitted; variable assignments would exist *ante rem*. Such entities would be true creatures of darkness. I certainly could not think of them as functions from the set of variables into the domain of the world or as any mathematical entities. However, stranger entities have been dreamt in philosophy. If the point argument is made about universals instead of variable assignment, as Bealer (1993) does, an argument for *ante rem* universals is obtained.

Thirdly, actualism could be rejected in favour of possibilism, that is, constant domain semantics. The same objects exist in all worlds, but only some are 'instantiated'. Variable assignments could be formed using uninstantiated objects.

Finally, one could use a proxy b in w for an object a that does not exist at w and form variable assignments with proxies. All the worlds of a pw-model have an infinite domain, so this is not obviously ruled by cardinality constraints.[7]

I do not take a stance here. The problem just outlined is supposed to demonstrate how the additional expressive strength permits a discussion of issues of metaphysics in the object language. In first-order quantified modal logic the problem does not arise, because it is moved entirely into the metalanguage and there problems are solved by using a purely extensional, non-modal metatheory. One may even consider replacing modal talk with purely extensional discourse about possible worlds, as Lewis (1968, 1986) did. Of course, I cannot analyze such alternative approaches here, but one of the criteria would be whether they reach the expressive strength of the fourth grade of modal involvement.

THE ROAD AHEAD

I have outlined one way to capture the strength of the fourth grade of modal involvement in a formal language containing a binary predicate Nec and with a possible worlds semantics. But this is only one way to capture the full strength of the fourth grade of modal involvement. There are various ways to deviate from the approach in this paper, and I have only briefly sketched some reasons for my choices. However,

[7] I thank Beau Mount for bringing this point to my attention.

I hope I have succeeded in conveying a taste of how the fourth grade of modal involvement can help to shed light on issues such as the discussion between actualists and possibilists. The fourth grade of modal involvement may not open a Cantorian paradise, but at least it provides an expansive playground not only for metaphysics, but also for the analyses of various epistemic, alethic, logical, and further modalities.

BIBLIOGRAPHY

Bealer, George (1982), *Quality and Concept*, Clarendon Press, Oxford.

Bealer, George (1993), 'Universals', *Journal of Philosophy* 90, 5–32.

Bealer, George (1998), Universals and properties, *in* S.Laurence and C.Macdonald, eds, 'Contemporary Readings in the Foundations of Metaphysics', Blackwell Publishers, Oxford, pp. 131–47.

Bull, R.A. (1969), 'On modal logic with propositional quantifiers', *Journal of Symbolic Logic* 34, 257–263.

Carnap, Rudolf (1934), *Logische Syntax der Sprache*, Springer, Wien.

Fine, Kit (1970), 'Propositional quantifiers in modal logic', *Theoria* 36, 336–346.

Forster, Thomas (2005), The Modal Aether, *in* R.Kahle, ed., 'Intensionality', A K Peters, Wellesley, pp. 20–41.

Fujimoto, Kentaro and Volker Halbach (2019), 'Classical determinate truth'. draft.

Garson, James (2001), Quantification in modal logic, *in* 'Handbook of Philosophical Logic', Kluwer Academic Publishers, pp. 267–323.

Halbach, Volker and Graham Leigh (2020), *The Road to Paradox: A Guide to Syntax, Truth, and Modality*, Cambridge University Press. to appear.

Halbach, Volker, Hannes Leitgeb and Philip Welch (2003), 'Possible Worlds Semantics For Modal Notions Conceived As Predicates', *Journal of Philosophical Logic* 32, 179–223.

Halbach, Volker and Holger Sturm (2004), 'Bealers Masterargument: Ein Lehrstück zum Verhältnis von Metaphysik und Semantik', *Facta Philosophica* 6, 97–110.

Halbach, Volker and Philip Welch (2009), 'Necessities and Necessary Truths: A Prolegomenon to the Metaphysics of Modality', *Mind* 118, 71–100.

Halbach, Volker and Shuoying Zhang (2016), 'Yablo without Gödel', *Analysis* 76, 53–59.

Kaplan, David (1970), 'S5 with quantifiable propositional variables', *Journal of Symbolic Logic* 35, 355. abstract.

Kripke, Saul (1975), 'Outline of a Theory of Truth', *Journal of Philosophy* 72, 690–716. reprinted in **?**.

Kripke, Saul A. (1959), 'A completeness theorem in modal logic', *Journal of Symbolic Logic* 24, 1–14.

Lewis, David (1968), 'Counterpart Theory and Quantified Modal Logic', *Journal of Philosophy* 65, 113–126.

Lewis, David (1986), *On the Plurality of Worlds*, Blackwell Publishers, Malden MA.

McGee, Vann (1985), 'How Truthlike Can a Predicate Be? A Negative Result', *Journal of Philosophical Logic* 14, 399–410.

Montague, Richard (1963), 'Syntactical Treatments of Modality, with Corollaries on Reflexion Principles and Finite Axiomatizability', *Acta Philosophica Fennica* 16, 153–67. Reprinted in (**?**, 286–302).

Quine, Willard Van Orman (1956), 'Quantifiers and Propositional Attitudes', *Journal of Philosophy* 53, 177–187.

Quine, Willard Van Orman (1960), *Word and Object*, MIT Press, Cambridge (Mass.).

Quine, Willard Van Orman (1976), Three Grades of Modal Involvement, *in* 'The Ways of Paradox', revised and enlarged edn, Harvard University Press, Cambridge, Mass., pp. 158–176.

Stern, Johannes (2014*a*), 'Modality and Axiomatic Theories of Truth I: Friedman-Sheard', *Review of Symbolic Logic* 7, 273–298.

Stern, Johannes (2014*b*), 'Modality and Axiomatic Theories of Truth II: Kripke-Feferman', *Review of Symbolic Logic* 7, 299–318.

Stern, Johannes (2016), *Toward Predicate Approaches to Modality*, Vol. 44 of *Trend in Logic*, Springer, Cham.

Tarski, Alfred (1935), 'Der Wahrheitsbegriff in den formalisierten Sprachen', *Studia Philosophica Commentarii Societatis Philosophicae Polonorum* 1, 261–405. translated as 'The Concept of Truth in Formalized Languages' in (**?**, 152–278); page references are given for the translation.

Visser, Albert (1989), Semantics and the liar paradox, *in* D.Gabbay and F.Günthner, eds, 'Handbook of Philosophical Logic', Vol. 4, Reidel, Dordrecht, pp. 617–706.

Williamson, Timothy (2013), *Modal Logic as Metaphysics*, Oxford University Press.

Yablo, Stephen (1985), 'Truth and Reflection', *Journal of Philosophical Logic* 14, 297–349.

Yablo, Stephen (1993), 'Paradox Without Self-Reference', *Analysis* 53, 251–252.

Volker Halbach
New College
Oxford OX1 3BN
England
VOLKER.HALBACH@NEW.OX.AC.UK

HOW THE CONTINUUM HYPOTHESIS COULD HAVE BEEN A FUNDAMENTAL AXIOM

JOEL DAVID HAMKINS

ABSTRACT. I describe a simple historical thought experiment showing how we might have come to view the continuum hypothesis as a fundamental axiom, one necessary for mathematics, indispensable even for calculus.

1. INTRODUCTION

I should like to describe how our attitude toward the continuum hypothesis could easily have been very different than it is. If our mathematical history had been just a little different, I claim, if certain mathematical discoveries had been made in a slightly different order, then we would naturally view the continuum hypothesis as a fundamental axiom of set theory, one furthermore necessary for mathematics and indeed indispensable for making sense of the core ideas underlying calculus.

The continuum hypothesis (CH) is the assertion that the cardinality of the set of real numbers is the first uncountable infinity, or in other words, that $2^{\aleph_0} = \aleph_1$. This hypothesis is known to be independent of the Zermelo-Fraenkel ZFC axioms of set theory—it is neither provable nor refutable, if ZFC itself is consistent, and it remains independent even relative to any of the usual large cardinal axioms. Currently CH is not generally regarded as part of the standard axiomatization of set theory, but rather is taken as a separate supplemental hypothesis, one to be mentioned explicitly, assumed or denied, proved or refuted, in diverse circumstances for different purposes. The continuum hypothesis holds, for example, in the constructible universe L and in other canonical inner models, but we can make it fail (or hold) in forcing extensions; it is refuted by the forcing axioms PFA and MM, which settle the continuum as $\aleph_2$. In the study of the cardinal characteristics of the continuum, set theorists routinely work with $\neg$CH as the subject is trivialized in a sense under CH, since there would be no room for variation in the cardinal characteristics, although it is also trivialized in a different way under the forcing axioms, since these imply that they are all fully pushed up to value continuum.

Since the truth or falsity of CH cannot be settled on the basis of proof from the ZFC axioms, set theorists have offered various philosophical arguments aiming at a solution to the continuum problem, the problem of determining whether CH holds or its negation. (See my survey discussion in [Ham21, chapter 8].) For example, Chris Freiling [Fre86] advances an argument for $\neg$CH based on prereflective intuitions

Commentary on this article can be made on the author's blog at https://jdh.hamkins.org/how-ch-could-have-been-fundamental. This article follows up on a line of argument I had presented in [Ham21, p.78-79, Questions 2.18, 2.19, 2.20]. I am grateful to the anonymous referees for several helpful suggestions, as well as to Dan Isaacson, W. Hugh Woodin, Mel Fitting, Jason Chen, Tim Button, as well as to other participants at the talks I gave on this topic at the 2024 Eastern APA in New York, at UC Irvine, and at the University of Oslo.

about randomness as a primitive notion. W. Hugh Woodin made a case for $\neg$CH based on considerations of Ω-logic and forcing absoluteness (see the survey in [Koe23]). More recently, Woodin argues on the other side, making a case for CH based on features of his theory of Ultimate L, a canonical inner model accommodating even the largest large cardinals (see [Rit15] for an account of Woodin's change of heart). Defending set-theoretic pluralism, I argue in [Ham12] that it is incorrect to describe CH as an open question—the answer to CH, rather, is pluralist, consisting of the deep body of knowledge that we have concerning how it behaves in the set-theoretic multiverse, how we can force it or its negation while preserving diverse other set-theoretic features.

Many set theorists have yearned for what I call the *dream solution* to the continuum problem, by which we settle the CH once and for all by introducing a new set-theoretic principle, the "missing" axiom, which everyone agrees is fully consonant with the concept of set and which also provably settles CH. I argue in [Ham15], however, that this will never happen.

> Our situation with CH is not merely that CH is formally independent and we have no additional knowledge about whether it is true or not. Rather, we have an informed, deep understanding of how it could be that CH is true and how it could be that CH fails. We know how to build the CH and $\neg$CH worlds from one another. Set theorists today grew up in these worlds, comparing them and moving from one to another while controlling other subtle features about them. Consequently, if someone were to present a new set-theoretic principle Φ and prove that it implies $\neg$CH, say, then we could no longer look upon Φ as manifestly true for sets. To do so would negate our experience in the CH worlds, which we found to be perfectly set-theoretic. It would be like someone proposing a principle implying that only Brooklyn really exists, whereas we already know about Manhattan and the other boroughs. And similarly if Φ were to imply CH. We are simply too familiar with universes exhibiting both sides of CH for us ever to accept as a natural set-theoretic truth a principle that is false in some of them.

Nevertheless, I should like to explain in this article how it all might easily have been different. If our mathematical history had been slightly revised in a way I shall presently describe, if certain mathematical discoveries had been made in a slightly different order, then we might have come to look upon CH as fundamental principle for set theory, one necessary to make sense of mathematical ideas at the core of classical mathematics.

2. The thought experiment—two number realms

As a thought experiment, let us imagine that Newton and Leibniz in the early days of calculus provide somewhat fuller accounts of their ideas about infinitesimals. In the actual world, to be sure, a satisfactory account of the basic nature and features of infinitesimals was lacking—the foundations of calculus were famously mocked by Berkeley [Ber34] with withering criticism:

> And what are these same evanescent Increments? They are neither finite Quantities nor Quantities infinitely small, nor yet nothing. May we not call them the ghosts of departed quantities?

It was simply not clear enough in the early accounts of calculus what kind of thing the infinitesimals were and whether they were part of the ordinary number system or somehow transcending it, inhabiting a different larger realm of numbers.

According to Jesseph [Jes93, p.168], Berkeley argued that

> If infinitesimal magnitudes are introduced into analysis, the question arises whether they obey the ordinary laws of addition, subtraction, multiplication, and division.

And he found fault with both sides of the resulting dichotomy.

What I propose is that we imagine that Newton and Leibniz provide greater clarity concerning the conception of infinitesimals. Specifically, I would like to imagine that Newton and Leibniz conceive of the infinitesimals, as many do today, as living in a larger field of numbers, distinct from but extending the ordinary real numbers. Let us suppose that they posit two "realms" of numbers, the ordinary realm $\mathbb{R}$ of the real numbers and a further realm $\mathbb{R}^*$ consisting of what we might call the *hyperreal* numbers, to use the contemporary terminology, a transcendent number field accommodating the infinitesimals.

This idea alone, that infinitesimals inhabit another realm of numbers, immediately addresses the mocking Berkeley criticism, releasing the tension of the otherwise paradoxical claim that infinitesimals are positive yet also smaller than every positive number, for we need only claim that infinitesimals are smaller than every positive real number, of course, and not smaller than all the other infinitesimal numbers or themselves. The two-number-realms idea serves to clarify much of the early discussion surrounding infinitesimals, enabling a frank discussion of how the real numbers are related to the hyperreal numbers and what the hyperreal numbers are like.

3. Two specific clarifications of the nature of infinitesimals

Let us imagine that two further clarifying principles are introduced. First, in order to explain the nature and existence of infinitesimals, our imaginary Leibniz writes:

(1) *Every conceivable gap in the numbers is filled by infinitesimals.*

The gap between 0 and the positive real numbers, for example, is thus filled with the infinitesimal hyperreal numbers, and similarly there are hyperreal numbers at infinitesimal distance to $\sqrt{2}$ and to π. This idea, of course, amounts to an incipient form of saturation, expressing how the hyperreal number system transcends the real numbers. Namely, we know now the hyperreal numbers $\mathbb{R}^*$ of nonstandard analysis are countably saturated, which means that every countably specified gap

$$x_0 \leqslant x_1 \leqslant x_2 \leqslant \cdots \qquad\qquad \cdots \leqslant y_2 \leqslant y_1 \leqslant y_0$$

with $x_i < y_i$ is filled by some hyperreal number z strictly between

$$x_0 \leqslant x_1 \leqslant x_2 \leqslant \cdots \quad < z < \quad \cdots \leqslant y_2 \leqslant y_1 \leqslant y_0$$

Indeed, there will be many such z strictly in the gap, since the gap between the x_n and z itself will also get filled, as will the gap above z and below all the y_n.

In our actual history, the actual Leibniz was already inclined toward (1), considering higher orders of infinitesimality. According to [Jes93, p.173], Berkeley complains:

> Some mathematicians (notably Leibniz and L'Hopital) hold that there are infinitesimal quantities of all orders and "assert that there are infinitesimals of infinitesimals of infinitesimals, without ever coming to an end."

The historical Euler 1780 ([Eul80]; see also [BW18, p.87–88]) also was busy exploring the vast space of infinite orders of the infinitely large and the infinitely small, discovering the saturation-like manner in which gaps get filled. He observed that for any infinitely large quantity x the number x^2 will be infinitely larger still and x^3 infinitely larger than that. In contrast, $\sqrt{x}$ will be infinitely smaller than x, but still infinite, and the higher roots $\sqrt[3]{x}$, $\sqrt[4]{x}$, similarly get infinitely smaller with each step, while remaining infinite. Nevertheless, Euler demonstrates that we may find orders of infinity that are infinitely smaller than every $\sqrt[n]{x}$, but still infinite. One such number is $\ln x$, and then $(\ln x)^2$ will be infinitely larger than this, but still smaller than every $\sqrt[n]{x}$, and similarly $\sqrt{\ln x}$ smaller again. By taking reciprocals, he finds the same rich phenomenon amongst the orders of infinitesimality relevant for calculus. In this way, he makes a festive party out of filling gaps in the orders of infinity, realizing more and more instances of saturation and thereby supporting principle (1). And of course this kind of thinking fed into the much later work of Hardy [Har10] on the orders of infinity and Hausdorff's related work showing in his context that all countably specified gaps are filled.

Second, in order to justify his calculations with fluxions, the ultimate ratios, and evanescent increments, let us imagine that Newton writes:

> (2) *The two number realms fulfill all the same fundamental mathematical laws.*

According to the new dictum, the hyperreal numbers would thus fulfill the associativity and distributivity laws, or indeed any law that is true for the real numbers. From our perspective, of course, we can view this statement as an incipient form of the transfer principle, by which the hyperreal field is an elementary extension of the real field $\mathbb{R} \prec \mathbb{R}^*$, even in expansions of the field structure to include other functions and relations. In particular, $\mathbb{R}^*$ would be a real-closed field, an ordered field in which every positive number has a square root and every odd-degree polynomial has a root.

Jesseph [Jes93, p.135] writes that "Wallis [1685] took infinitesimal methods to be essentially the same as the method of exhaustion but shorter and more readily applied," which can be seen as a conservativity claim about the methods, and Button and Walsh [BW18, §4.7] describe Leibniz's philosophy of fictionalism about infinitesimals, expressed in his letter to Varignon, which can also be seen as a form of conservativity in that there is nothing really new in them. Newton himself in the *Principia* (1687) expresses conservativity for his methods, stating:

> These Lemmas are premised to avoid the tediousness of deducing perplexed demonstrations *ad absurdum*, according to the method of the ancient geometers. [New46, p.102]

Thus again, there is nothing mathematically new going on. This conservativity attitude fits with the proto-transfer-principle idea of statement (2), which also expresses a kind of conservativity, that there are no new mathematical rules arising in the new hyperreal number realm.

The proposal here with my thought experiment is definitely not that Newton and Leibniz have a full-blown well formulated theory of saturation and transfer, or even of the concept of a field (which only came much later, after Galois), but rather only that they have expressed the primitive idea of two distinct number realms, the real numbers and hyperreal numbers, with vaguely expressed ideas that appear from a contemporary perspective as incipient forms of saturation and the transfer principle. The thought experiment requires only very small initial steps towards the two-realms conception, since further development and rigor would naturally come in time, just as it did in our actual mathematical history.

Nevertheless, the proposal does call for us to imagine that mathematicians might have been a little more modern in their attitude toward number systems, moving beyond the historical understanding of numbers at that time as ratios of geometric magnitudes toward the idea of there being distinct number realms. To be sure, mathematicians have long distinguished at least between natural numbers and other kinds of magnitudes, such as when considering number-theoretic concepts such as even, odd, and prime, and the thought experiment calls for further analogous distinctions concerning the infinitesimals.

Meanwhile, the extent to which the historical Newton had used infinitesimals in the first place is a matter of discussion by historians of mathematics. He had renounced them explicitly, taking himself to mount instead his method of fluxions and the concept of ultimate ratios, which can be seen arguably either as a proto-limit concept or as disguised infinitesimals. In any case, for the purposes of my thought experiment we needn't be troubled by Newton's possible hesitancy towards infinitesimals, since the thought experiment is not about the historical Newton, but about the infinitesimal concept itself, which certainly did exist at the time. For the success of the thought experiment, therefore, if this is an issue we could simply imagine if necessary that it was mainly Leibniz or another Leibniz-like figure who had provided the somewhat fuller explication of the nature of infinitesimals that I am discussing.

4. The hyperreals are fundamentally coherent

The main initial point I should like to stress is that we know that these ideas about saturation and the transfer principle for the real and hyperreal numbers are fundamentally coherent and mathematically correct in light of the much later developments of nonstandard analysis, due to Abraham Robinson in the 1960s; they are definitely sufficient for a robust infinitesimal theory of calculus. A glance at Keisler's remarkable infinitesimals-based undergraduate calculus text [Kei00] shows what is possible when beginning even with only very elementary ideas—the entire classical theory can be developed on these notions. To be sure, in our actual mathematical history, the development of calculus proceeded to enormous success on the basis of very primitive infinitesimal ideas, without a fully rigorous foundation, and yet still achieved the key mathematical insights. The thought experiment I propose is that all those developments and insight would still occur, of course, and more, because even a slightly greater initial clarity in the infinitesimal concept would naturally lead to and support fruitful further analysis. In the imaginary history, the development of calculus would be something a little closer to the developments of what we now call nonstandard analysis, perhaps primitive at first, but with increasing sophistication and rigor. Precisely because we know that calculus can be successfully developed this way, the approach would not meet with any fundamental obstacle.

Kurt Gödel explains his views on nonstandard analysis after Robinson's talk on the subject at the IAS Princeton in 1973:

> There are good reasons to believe that non-standard analysis, in some version or other, will be the analysis of the future.
>
> One reason is the just mentioned simplification of proofs, since simplification facilitates discovery. Another, even more convincing reason, is the following: Arithmetic starts with the integers and proceeds by successively enlarging the number system by rational and negative numbers, irrational numbers, etc. But the next quite natural step after the reals, namely the introduction of infinitesimals, has simply been omitted. I think in coming centuries it will be considered a great oddity in the history of mathematics that the first exact theory of infinitesimals was developed 300 years after the invention of the differential calculus. [G9̈0, p. 311]

Gödel thus paints the picture that it is our own actual mathematical history that is odd and strange—the history of ideas in my imaginary thought experiment, in contrast, would be the more natural progression. I take this as truly very strong support for the fundamental reasonableness of my thought experiment.

5. The hyperreal numbers become a familiar mathematical structure

In this imaginary early history, therefore, the hyperreal numbers would be successful in the foundations of calculus, and as a result they would be taken seriously as a distinct realm of numbers, becoming a core part of the mathematical conceptions underlying the calculus. In the imaginary world, calculus would be founded fully on infinitesimals, without any need for $\forall\varepsilon\,\exists\delta$ limit concepts; the use of infinitesimals would become increasingly sophisticated and rigorous.

The hyperreal numbers would thus enter the Pantheon of number systems at the center of mathematics, the fundamental structures that mathematicians discovered and then used throughout their mathematical work.

$$\mathbb{N} \qquad \mathbb{Z} \qquad \mathbb{Q} \qquad \mathbb{R} \qquad \mathbb{C} \qquad \mathbb{R}^*$$

The hyperreal numbers would thereby find their place in mathematics alongside the other familiar standard mathematical number systems—the natural numbers, the integers, the rational numbers, the real numbers, the complex numbers—and in the world of my thought experiment, there would stand also the hyperreal numbers.

Reflecting on the move from the real numbers $\mathbb{R}$ to the hyperreal numbers $\mathbb{R}^*$, one might be encouraged to seek out saturated versions of all our favored mathematical structures. But actually, for the number systems I have mentioned in the Pantheon above, the hyperreals already provide this. By the transfer principle, we get the hypernatural numbers $\mathbb{N}^*$, the ring of hyperintegers $\mathbb{Z}^*$, the field of hyperrational numbers $\mathbb{Q}^*$, all sitting inside the hyperreal number field $\mathbb{R}^*$, as well as the hypercomplex numbers $\mathbb{C}^* = \mathbb{R}^*[i]$, consisting of numbers $a + bi$, where $a, b \in \mathbb{R}^*$ are hyperreal. In this way, we may view the move to the hyperreals simply as the move of saturating all our familiar structures.

6. On the necessity of categoricity for structuralism

Daniel Isaacson [Isa11], taking inspiration from Kriesel, describes the process by which mathematicians come to know their mathematical structures. Namely, as I describe it in [Ham21], by informal rigor "we become familiar with a structure; we find the essential features of that structure; and then we prove that those features axiomatically characterize the structure up to isomorphism. For Isaacson, this is what it means to identify a particular mathematical structure, such as the natural numbers, the integers, the real numbers, or indeed, even the set-theoretic universe." Isaacson says,

> ...the reality of mathematics turns ultimately on the reality of particular structures. The reality of a particular structure, constituting the subject matter of a branch of mathematics such as number theory or real analysis, is given by its categorical characterization, i.e. principles which determine this structure to within isomorphism. [Isa11, p. 2]

In this way, the categorical characterizations of our familiar particular structures become the framework of our mathematical reality.

This plays out in our actual mathematical history when Dedekind [Ded88] proves that the natural number structure $\mathbb{N}$ is uniquely specified up to isomorphism by his theory of the successor operation, leading directly to Peano's elegant development of elementary number theory in that framework. Building upon this, mathematicians provide categorical accounts of the integer ring $\mathbb{Z}$ and the rational field $\mathbb{Q}$. Cantor [Can95; Can97; Can52] proves that the rational order $\mathbb{Q}$ is characterized as the unique countable endless dense linear order. Huntington [Hun03] provides the categorical account of the real field $\mathbb{R}$ as the unique complete ordered field. The complex numbers are characterized as the algebraic closure of $\mathbb{R}$. The categorical characterizations of these core mathematical structures are the central results for the coherence of the mathematical enterprise, enabling us to refer to the various fundamental mathematical structures by their defining characteristics.

I have pointed to the categoricity results as the origin of the philosophy of structuralism in mathematics.

> Categoricity is central to structuralism because it shows that the essence of our familiar mathematical domains, including $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, and so on, are determined by structural features that we can identify and express. Indeed, how else could we ever pick out a definite mathematical structure, except by identifying a categorical theory that is true in it? Because of categoricity, we need not set up a standard canonical copy of the natural numbers, like the iron rod kept in Paris that defined the standard meter; rather, we can investigate independently whether any given structure exhibits the right structural features by investigating whether it fulfills the categorical characterization. [Ham21, p. 31]

In short, having categorical accounts of all our core mathematical structures is necessary for mathematical reference and supports a structuralist mathematical practice.

In set theory, this phenomenon arises after the improvement of Zermelo's flawed initial set theory to Zermelo-Fraenkel set theory, with the addition of the replacement and foundation axioms, an improvement that makes possible Zermelo's famous quasi-categoricity results of [Zer30], showing that the models of second-order set theory ZFC_2 agree with one-another on initial segments. The new ZFC_2 theory thus enjoys a measure of categoricity for the intended set-theoretic universe, leaving open only how high the ordinals will grow. To my way of thinking, categoricity should be a bigger part of the conversation concerning Zermelo's original set theory versus Zermelo-Fraenkel set theory. The models of this theory ZFC_2 are exactly the uncountable Grothendieck-Zermelo universes, now used pervasively in the foundations of category theory, and of course set theorists study them in connection with the inaccessible cardinals. Many of these models admit fully categorical characterizations, and Robin Solberg and I [HS20] explore the spectrum of fully categorical extensions of ZFC_2, while considering the curious tension between categoricity and reflection principles in the foundations of set theory.

7. The key imaginary event

In the world of my historical thought experiment, we shall similarly have categorical characterizations of all the various principal mathematical structures at the end of the 19th century and early 20th century, including the natural numbers, the integers, the real numbers, and the complex numbers.

But what about the hyperreal numbers? In the imaginary history, after all, the hyperreal numbers $\mathbb{R}^*$ have become a core mathematical structure alongside all the others, situated at the very foundations of calculus, present from the start of that subject. Mathematicians would demand an account of the definitive underlying theory of the hyperreal numbers, which would require a categorical characterization like all the others. Naturally, one would expect this characterization to involve the key features already recognized as characteristic of the hyperreal numbers, the saturation ideas and the transfer principle, just as the characterization of the natural numbers involves induction and that of the real numbers involves the least-upper-bound completeness principle.

Thus we come to the key imaginary event of the thought experiment. Namely, let us imagine that in the early 20th century, a Zermelo-like figure formulates a sufficient theory—the theory ZFC + CH suffices—able to prove a categorical characterization of the hyperreal number field $\mathbb{R}^*$, similar to how the actual Zermelo introduced his set theory as an explanation of his proof of the well-order theorem.

The theory ZFC + CH is indeed able to provide the desired characterization of the hyperreal field $\mathbb{R}^*$ as follows, which shows under CH how $\mathbb{R}^*$ is characterized by refined versions of the two ideas we had attributed to Newton and Leibniz in the thought experiment.

Hyperreal categoricity theorem. *Assume* ZFC + CH. *Then there is up to isomorphism a unique smallest countably saturated real-closed field.*

This theorem is by now a standard result in elementary model theory, proved using a back-and-forth argument in the style of Cantor's famous argument about the rational order, except that here the back-and-forth construction proceeds transfinitely through ω_1 many steps rather than just countably many (see Erdős, Gillman, and Henriksen [EGH55]). A model is *countably saturated*, that is, $\aleph_1$-saturated, if it realizes every finitely satisfiable type with countably many parameters, but in the theorem we actually need to require only that the order is saturated—all countably described gaps should be filled. The general fact in play here is that any two saturated models of the same complete theory and size are isomorphic, and that would be the situation for our hyperreal fields under the hypotheses of the categoricity theorem. The smallest possible size in question would be the continuum, since countably saturated real-closed fields must have size at least continuum, and there are such fields of size continuum.

Indeed, we have several various constructions of a countably saturated real-closed field of size continuum. Namely, (1) we can construct the ultrapowers $\mathbb{R}^{\mathbb{N}}/\mu$ of the real field by any nonprincipal ultrafilter μ on $\mathbb{N}$, and this is always a countably saturated real-closed field of size continuum; (2) we can proceed via Hahn series, a generalized kind of power series, to construct a countably saturated real-closed field of size continuum; (3) we can undertake a general model-theoretic construction, successively realizing types in a transfinite elementary chain, to produce a countably saturated model of any given consistent theory, including the theory of real-closed fields; (4) perhaps exemplifying this construction in an attractive, concrete general manner, we can undertake the Conway construction of the surreal field through all countable ordinal birthdays, filling all possible gaps that arise—the result is $\mathbb{No}(\omega_1)$, a countably saturated real-closed field of size continuum.

The hyperreal categoricity theorem shows under ZFC+CH that all these various constructions give rise to exactly the same hyperreal field, which we may thereby regard as the canonical structure of the hyperreal field $\mathbb{R}^*$. The situation for the hyperreals under CH is thus rather like that of the real field in ZFC, for which we also have a variety of constructions, proceeding with Dedekind-cuts in the rationals or equivalence classes of Cauchy sequences and so forth. All the various presentations of complete ordered fields are provably isomorphic in ZFC, and this categoricity enables a structuralist treatment of the real field as the unique complete ordered field. In a sense ZFC is aimed at providing the satisfactory theory of the real numbers, for the real categoricity theorem is not provable in weaker systems, such as constructive mathematics, where mathematicians must treat the Dedekind reals as a distinct conception from the Cauchy reals.

Similarly, in ZFC+CH, all the various constructions of the hyperreal field give rise to the same underlying canonical structure, thereby enabling a structuralist account of infinitesimals—the hyperreals are the unique smallest countably saturated real-closed field.

8. CH is required

I should like to call attention to a key feature of the thought experiment, namely, the mathematical fact that the CH is required. We can provide the categorical characterization of the hyperreal numbers in ZFC+CH as stated in the hyperreal categoricity theorem, but this is not possible in ZFC alone. Judith Roitman [Roi82] showed that it is relatively consistent with ZFC + $\neg$CH that there are multiple non-isomorphic hyperreal fields arising as ultrapowers $\mathbb{R}^\omega/\mu$, and these are always countably saturated real-closed fields of size continuum. Alan Dow [Dow84] showed that whenever CH fails, then indeed there are multiple non-isomorphic ultrapowers $\mathbb{R}^\omega/\mu$, non-isomorphic even merely in their order structure. Thus, CH is outright equivalent to the hyperreal categoricity assertion that there is a unique smallest countably saturated real-closed field (see also [Est77]).

These results show that the phrase “the hyperreal numbers” is not generally meaningful in ZFC, because in ZFC there is not necessarily just one mathematical structure fitting the description. But in ZFC+CH, there is. With the continuum hypothesis, we can specify the hyperreal numbers up to isomorphism as a canonical structure, the unique smallest countably saturated real-closed field.

9. How CH gets on the list

So this is how CH gets on the list of fundamental axioms. The thought experiment, at bottom, is that the hyperreal field $\mathbb{R}^*$ is long a core mathematical idea, pre-rigorous at first, but then with increasing rigor and sophistication. To give a foundational account of the hyperreal number system and thus of the theory of infinitesimals, the Zermelo-like figure provides an existence proof and categorical characterization, introducing the fundamental axioms of ZFC + CH in order to do so. We know that this is possible and, furthermore, that CH cannot be omitted. So CH gets onto the list of fundamental axioms, being necessary to establish the basic coherence or even (in the Isaacson sense) the reality of the hyperreal numbers and thus indispensable for the foundations of calculus.

10. Extrinsic and intrinsic support for CH

The developments would provide enormous extrinsic support for CH, similar to the extrinsic justification ZFC currently enjoys in light of its robust foundational account of the real numbers $\mathbb{R}$. The theory ZFC + CH would be seen as similarly successful regarding the theory of the hyperreal numbers.

In the imaginary history, I would find it quite likely that after the CH had found its extrinsic justification in this way as a mathematical necessity, then intrinsic justifications would also begin to find their appeal, similar to how the axiom of choice is often viewed as extrinsically justified by its widespread use and important central consequences, but set theorists also point to its intrinsic justification under the concept of arbitrary set existence. In the case of the continuum hypothesis, the intrinsic justification I imagine is that CH asserts that the two methods of achieving uncountability agree, that is, the process of going to the next higher cardinal gives the same result as taking the power set; in short, $\aleph_1 = \beth_1$. This can be seen as a unifying, explanatory principle of the uncountable, and therefore an intrinsic justification for CH.

We might also reflect on the fact that by basic human rationalizing nature, one naturally finds it easier to be convinced by arguments for the intrinsic truth of an axiom, once one has already been convinced of the axiom's extrinsic necessity.

11. A generalized thought experiment and the generalized continuum hypothesis

The hyperreal categoricity theorem is that under CH, the hyperreal field is the unique smallest countably saturated real-closed field. Perhaps a critic objects that the smallest-size requirement is ad hoc—wouldn't it be more natural to relax this and consider hyperreal fields of other sizes? And isn't it unnatural to require only countable saturation, rather than full saturation?

Let me address this initially by defending countable saturation as a rich, natural notion. Countable saturation is both easy to express and understand, and suffices for the robust existence of infinitesimals in a vast hierarchy of orders. It leads to a rich, successful theory, without the need for higher levels of saturation. Furthermore, the construction of a countably saturated real-closed field, as with the surreal construction through the countable birthdays, is both clear and natural. So there is very little lacking in our conception of the hyperreals as a countably saturated real-closed field. And Hausdorff's 1909 proof that $\mathbb{N}^{\mathbb{N}}/\mathrm{Fin}$ is countably saturated but has an unfilled (ω_1, ω_1) gap would tend to encourage a greater focus specifically on countable saturation.

Meanwhile, to be sure, full saturation does imply categoricity in any given cardinality in which it occurs by the back-and-forth construction. And furthermore, the existence of fully saturated models does not require CH. The existence of saturated real-closed field of size continuum is equivalent merely to $\mathfrak{c}^{<\mathfrak{c}} = \mathfrak{c}$, which can occur say, even if $\mathfrak{c} = 2^{\aleph_0} = 2^{\aleph_1} = \aleph_2$, and in many other kinds of cases. There will be a fully saturated real-closed field of uncountable size κ if and only if $\kappa^{<\kappa} = \kappa$, which is independent of the CH and GCH, and it can hold consistently by forcing with any particular regular uncountable cardinal, although it holds necessarily of any inaccessible cardinal. And yet, even in these cases, where CH fails and there is a saturated field of size continuum or more, there will also be numerous non-isomorphic countably saturated such fields, which may detract from the sense of uniqueness for the hyperreal conception.

Nevertheless, let me take on board both of the critical objections at once with a generalized thought experiment. Let me imagine that the Zermelo-like figure adopts an expansive attitude toward the hyperreals, proving instead the following generalized hyperreal categoricity theorem, on the basis of the *generalized continuum hypothesis* (GCH), which asserts that $2^\kappa = \kappa^+$ for every infinite cardinal κ.

Generalized hyperreal categoricity theorem. *Assume* ZFC + GCH. *Then there are up to isomorphism unique saturated real-closed fields in every uncountable regular cardinality.*

The theorem can be proved by observing that saturated models of the same size and theory are unique up to isomorphism by the back-and-forth construction, as we have mentioned, and you get existence of saturated models from the GCH by realizing types in a transfinite elementary tower. In fact the converse of this theorem also is true—that is, the GCH is required for the generalized categoricity here—since $\kappa^{<\kappa} = \kappa$ is necessary for there to be a κ-saturated model of size κ, and this implies the GCH if it holds for every uncountable regular cardinal. Thus, the generalized hyperreal categoricity principle is itself equivalent over ZFC to the GCH.

On the basis of the GCH we thereby find ourselves blessed with canonical hyperreal fields in every desired size, a transfinite tower of orders of infinitesimality continuing to all higher cardinals. This would be a natural continuation of the saturation ideas originating with Leibniz and continuing with Euler's exploration of the diverse orders of infinity, and then with Hardy and into contemporary times. And since the GCH is required, this is how the GCH could also get on the list of fundamental axioms. These higher uncountable hyperreal fields would be seen as converging in a vast elementary chain ultimately to the surreal numbers, another core number concept with a proper-class categoricity characterization.

12. Foundationalism and a priori knowledge

Williamson [Wil16, §III] speculates on how it would be that other beings, with physical and mathematical powers similar to humans, might settle the continuum hypothesis.

> One normal mathematical process, even if a comparatively uncommon one, is adopting a new axiom. If set theorists finally resolve CH, that is how they will do it. Of course, just arbitrarily assigning some formula the status of an axiom does not count as a normal mathematical process, because doing so fails to make the formula part of mathematical knowledge. In particular, we cannot resolve CH simply by tossing a coin and adding CH as an axiom to ZFC if it comes up heads, $\sim$CH if it comes up tails. We want to know whether CH holds, not merely to have a true or false belief one way or the other (even if we could get ourselves to believe the new axiom). Thus the question arises: when does acceptance of an axiom constitute mathematical knowledge?

My thought experiment engages with Williamson's challenge by describing the richer context and process that would lead mathematicians to the CH. In the alternative world I describe, mathematicians have gained increasing familiarity over the centuries with the hyperreal number system, embedding it at the center of classical mathematical developments, especially in the calculus, and thus they have gained increasing confidence in their use of the hyperreals as a particular, familiar mathematical structure. Seeking a more thorough underlying mathematical explanation of it, including a categorical account, as we do with all our particular mathematical structures, they would find it in a theory including CH, which can prove hyperreal categoricity, and we know that CH is necessary for this. In this way, my thought-experiment mathematicians are led to the CH by undertaking what Williamson calls the "normal mathematical process." The axiom they accept in effect is hyperreal categoricity, since it is necessary for the coherence and reality of their mathematical practice, and this principle is equivalent to the CH.

In another thought experiment, Berry [Ber13] introduces the mathoids, creatures who look upon Fermat's last theorem as intuitively and immediately obvious, a foundational belief, without feeling any need to justify this stance with an argument that we would find convincing; similar imaginary creatures are discussed in [Wil16, §III]. This is part of her investigation into foundationalism about a priori mathematical knowledge, in which she highlights a problematic conclusion, namely, "that nothing in the current literature lets us draw a principled distinction between what these creatures are doing and paradigmatic cases of good a priori reasoning."

I take myself to sidestep that particular debate in my thought experiment, precisely because I took pains to describe mathematical characters that we *would* find convincing. My imaginary Newton and Leibniz and the imaginary Zermelo-like figure seem perfectly reasonable given what we know about the underlying mathematics—far more reasonable than the mathoids, who strike us (and this is important for Berry's point) as unreasonable in their mathematical beliefs. My thought experiment, after all, is simply to reorder certain mathematical discoveries that we know are correct, in such a way that the hyperreals would naturally become far more central in classical mathematics than they are in our actual history, with the consequence that the need for a categorical account of them would become more urgent. This would lead us, I have argued, to a very different attitude towards the foundational theories able to provide such an account, and these theories would have to include CH.

13. Imaginary later history

Let us continue the thought experiment by considering how later mathematical developments would be received in the world of the imaginary history I have proposed. Gödel proves that ZFC+CH is true in the constructible universe L, which would of course be very welcome confirmation of the main theory. This could even lend some extrinsic support for $V = L$, more so than it enjoys currently. Indeed, the fact that GCH holds in L would provide further support for $V = L$ in light of the generalized hyperreal categoricity result in section 11. Solovay's theorem [Sol06], showing that under $V = L$ every finitely axiomatizable complete second-order theory is categorical, might be welcomed as a corresponding fundamental principle in the same light, rather than a curiosity about L as currently. (Solovay's theorem has been generalized to $L[\mu]$ and the large cardinal context by [SVW24], who also show under the axiom of projective determinacy, a unifying consequence of large cardinals, that every finitely axiomatizable complete second-order theory with a countable model is categorical.)

A commitment to a robust theory of the hyperreals presumes in certain ways a commitment to the axiom of choice or at least fragments of it. One needs the prime ideal theorem to have suitable ultrafilters μ with which to form the ultrapower $\mathbb{R}^{\mathbb{N}}/\mu$, and one needs countable choice in order to know that indeed these satisfy the transfer principle and are countably saturated. Conversely, the existence of a hyperreal field with infinitesimals and the transfer principle outright implies the existence of nonprincipal ultrafilters on the natural numbers, since for any fixed infinite number $N \in \mathbb{N}^*$ we can define $X \in \mu \leftrightarrow X \subseteq \mathbb{N}$ and $N \in X^*$, in effect defining that X is μ-large when it expresses a property that N exhibits. Meanwhile, the existence of ultrafilters implies the existence of non-Lebesgue measurable sets in the real numbers, an often-mentioned consequence of the axiom of choice. In this way, my thought-experiment inhabitants take the *modus tollens* to Alan Connes's criticism of nonstandard analysis (see [GS07] and my related discussion [Ham21, p. 81]). In this way, a commitment to the hyperreal numbers carries a small accompanying commitment to the axiom of choice.

Meanwhile, the discovery via forcing that without CH there can be multiple non-isomorphic hyperreal fields, as mentioned in section 8, would be seen as chaotic and bizarre, perhaps a little like current attitudes about models of ZF with strange failures of the axiom of choice. For example, it is known to be relatively consistent with ZF without the axiom of choice that the rational field $\mathbb{Q}$ can have multiple distinct non-isomorphic algebraic closures, a countable one as well as an uncountable one [Lä62; Hod76]. Mathematicians often find this situation very strange, and many regard this model as getting something fundamentally wrong about the algebraic numbers. I believe similarly that the mathematicians in the imaginary world would find it very odd to have multiple non-isomorphic hyperreal fields, and this would reinforce the view that $\mathrm{ZFC} + \mathrm{CH}$ is the right theory.

The method of forcing would be received differently in the imaginary world than in our own world—it would be perceived as a little less successful. For us, one of the attractive central features of forcing is that it necessarily preserves ZFC. Every forcing extension of a model of ZFC is another model of ZFC. But in the imaginary world, the main standard theory is ZFC+CH, and the corresponding feature is not true of this theory, since forcing can destroy CH. Indeed, this was Cohen's main initial application, to produce a model of $\mathrm{ZFC} + \neg\mathrm{CH}$. Perhaps the forcing method would be viewed in a similar way to how some people currently view the symmetric model constructions, which preserve ZF, but not necessarily the axiom of choice. We often use the symmetric model construction to build strange badly behaved models of ZF, enabling us to see how things can go awry when one doesn't have the axiom of choice. Similarly, in the imaginary world I propose, forcing could be used to build "strange" models of ZFC with $\neg$CH, showing how things can go awry when one doesn't have the continuum hypothesis.

In particular, my imaginary mathematicians would have a very different attitude concerning the dream-solution situation I described in the introduction, since in the thought experiment, they have adopted CH as fundamental before having had the experience via forcing of the $\neg$CH worlds, which they would view as strange. My thought experiment shows in effect how it could have been that the dream solution was achieved, even though it is no longer possible for us to achieve it. For my imaginary mathematicians, CH is a consequence of categoricity for the hyperreal numbers, and that is for them an instance of the dream solution.

In the imaginary world, of course eventually the $\forall\varepsilon\,\exists\delta$ accounts of limits and continuity would be discovered, but these would be seen as complicated and unnecessary abstractions, in light of the straightforward use of infinitesimals. The situation would be an inversion of current attitudes towards nonstandard analysis and ultrapowers.

A challenging counterpoint would eventually flow from the discovery by Esterle [Est77] (see also [Ehr12, theorem 17]) that in ZFC there is an initial countably saturated real-closed field, that is, one that embeds isomorphically into all other such fields, without requiring CH. By stratifying the field as a union of fields with all cuts having countable cofinality, one can mount a subtle back-and-forth argument along that hierarchy to show that this field is unique, again without CH. Indeed, this initial field is simply the field $\mathbb{N}\mathrm{o}(\omega_1)$ of the surreal numbers born at a countable ordinal birthday. This is a highly canonical mathematical structure, a countably saturated real-closed field that admits a very natural construction and furthermore embeds isomorphically into all such fields—it is therefore "smallest" in a stronger way than mere cardinality as in the hyperreal categoricity theorem, and furthermore its categorical characterization does not require CH. Would this undermine the status of the hyperreal categoricity theorem in my imaginary world? Perhaps people would want to replace the previous understanding of what the hyperreal numbers are with a new categorical account—they are the unique initial countably saturated real-closed field.

I would have several responses to this. First, the situation when CH fails, as I have mentioned, would remain chaotic, with numerous non-isomorphic minimal-size countably saturated real-closed fields, arising as ultrapowers $\mathbb{R}^{\mathbb{N}}/\mu$ and by other constructions, including $\mathbb{N}\mathrm{o}(\omega_1)$. Second, while the initiality of $\mathbb{N}\mathrm{o}(\omega_1)$ is a very natural property, I totally agree, meanwhile there would be other hyperreal field candidates with other very natural properties. For example, if $\mathfrak{c}^{<\mathfrak{c}} = \mathfrak{c}$ there would also be a fully saturated real-closed field of size continuum, which would definitely be different than $\mathbb{N}\mathrm{o}(\omega_1)$ when CH fails, and yet this alternative would also be a natural, canonical structure. Which would be the real hyperreals? It recalls the situation of the real numbers in constructive mathematics, where one must distinguish between the Dedekind reals and the Cauchy reals and so forth, and these are not constructively isomorphic. Which are the real reals in constructive mathematics? Similarly, which are the real hyperreals in $\mathrm{ZFC} + \neg\mathrm{CH}$? Under CH, all the various candidates are isomorphic, which largely dissolves the issue, and in this sense, the categoricity situation is simply much better with CH. But meanwhile, perhaps in my imaginary world there would be a community of mathematicians proving theorems about the various kinds of hyperreal fields in mere ZFC, in the same way that we currently have a community of mathematicians proving theorems about the various real fields in constructive mathematics.

14. Hyperreal hesitancy in the actual world

To my way of thinking (see [Ham21, p. 78-79, Questions 2.18, 2.19, 2.20]), the lack of a categoricity result for the hyperreal numbers in ZFC is a principal part of the explanation for the hesitancy amongst many mathematicians to take up nonstandard analysis. Mathematicians are naturally loathe to mount a fundamental mathematical theory like calculus with an underspecified mathematical structure at its core. We don't want to found calculus on an unknown hyperreal structure, if there are multiple non-isomorphic hyperreal structures to choose from. Which one do we choose? What if the resulting theory and features would depend on the particular choice? And how are we to refer anyway to a particular one of the hyperreal fields in the first place, if we lack a categorical description that picks it out from the alternatives? The need for categorical characterizations of all our fundamental mathematical structures seems necessary for a coherent structuralist mathematical practice of referring to them.

This explanation of the observed hyperreal hesitancy amounts to the contrapositive of the claim that we require categorical accounts for all our core mathematical structures. In our actual mathematical history, ZFC is already established as the core foundational theory before hyperreals are discovered, and so the lack of categoricity for the hyperreals means that we cannot accept them as a core structure.

One might inquire whether the proposed argument for CH could be convincing in our actual world, now that we know about the hyperreal number systems. One problem is that in our actual history, the hyperreal numbers never became a core mathematical structure in the first place, and so there is perhaps no pressing need to provide a categorical account of them. In actual history, we made calculus rigorous with the $\forall\varepsilon\,\exists\delta$ limit formalism of Bolzano and Weierstrass, and the hyperreal number systems were a later discovery, a logical curiosity, superfluous for calculus. This is why the hyperreal-categoricity argument for CH is more compelling in the imaginary thought-experiment world, where hyperreals were a core structure from the start, and less so in the actual world, where they were not.

15. Historical contingency

Penelope Maddy [Mad88] argued that there is a certain historical contingency to the ZFC axioms of set theory.

> The fact that these few [ZFC] axioms are commonly enshrined in the opening pages of mathematics texts should be viewed as an historical accident, not a sign of their privileged epistemological or metaphysical status. [Mad88]

She was concerned mainly with the large cardinal extensions of ZFC and the accompanying determinacy principles, seeking reasons to justify their incorporation into our basic conception of set theory.

I have argued in this paper similarly for the historical contingency of ZFC, describing how it could have been that we view the continuum hypothesis itself as a basic principle, necessary for the success of mathematics. While this kind of contingency for ZFC is a consequence of my argument, my main conclusion is not about contingency as such, but rather specifically that we might easily have had very different views about the continuum hypothesis.

Nevertheless, perhaps there would be other historical thought experiments by which we might have come to view $\neg$CH as fundamental, although the current examples I know of strike me as less compelling than the example I have described in this paper. Solovay (in personal discussions) has defended a vision of the real continuum that it should be a real-valued measurable cardinal, and this would necessarily involve an outsized failure of the CH, by which the continuum is extremely large, but also there would be a fully saturated hyperreal field of size continuum. Moore [Moo10] has defended the vision of set theory under the forcing axiom PFA, and others with MM and MM^+, all of which imply the continuum is $\aleph_2$ and that there is a fully saturated hyperreal field of size continuum. Moore says, "Forcing axioms have proved very effective in classifying and developing the theory of objects of an uncountable or non separable nature," and it would seem possible to expand this with a similar kind of thought experiment by which PFA was essential for the resolution of some fundamental mathematical commitment. Similarly, I have described in [Ham03; HW05] forcing axioms such as the c.c.c. maximality principle MP_{ccc} and a generalization to the necessary c.c.c. maximality principle $\Box\mathrm{MP}_{ccc}$, which imply that the continuum is larger than any cardinal that we can describe in any way that would be absolute to c.c.c. forcing extensions. I can imagine similar thought experiments by which these principles could be taken as fundamental.

If indeed such thought experiments are possible, or indeed only on the basis of the thought experiment of this paper in comparison with the standard ZFC-only foundations, I am inclined to take them all as an argument against the view that mathematical foundations have a necessary nature. Rather, for various sound reasons, mathematicians might have come to various different and perhaps incompatible conclusions about what they will take to be the central mathematical principles around which they intend to organize their mathematical investigations. For this reason, I regard my thought experiment here as supporting pluralism in the foundations of mathematics, by showing how it could naturally have been that we take a different theory as fundamental than we do currently.

Nevertheless, I recognize that for set theorists taking the universe view, by which there is a unique determinate set-theoretic reality, to have an argument that CH could have been a fundamental truth is in effect to have an argument that it is in fact a fundamental truth. And for this, I shall simply place my thought-experiment argument alongside the other CH arguments on offer. Namely, taken this way, as a proposal to those with the universe view, my argument is that CH is true because it provides the categorical theory of the infinitesimal numbers and indeed it is necessary for this.

16. Conclusion

I have described how we could have come to have a very different perspective on the continuum hypothesis. It could easily have been that the early theory of calculus had been a little more clear about infinitesimals, positing that they inhabit a distinct further realm, a system of numbers we might call the hyperreal numbers. This hyperreal number system would thereby have come to be embedded as a necessary component at the core of calculus. Clarifying the relation between the real and hyperreal numbers, early incipient forms of saturation and the transfer principle would have been put forth, vaguely at first, but then with increasing sophistication. We know from nonstandard analysis that these ideas are capable of serving as foundational in the development of infinitesimals-based calculus, and so the resulting theory would have been robust and successful. With the rise of rigor at the end of the 19th and early 20th centuries, when mathematicians were providing categorical accounts of all the familiar mathematical structures, a Zermelo-like figure would have provided such a characterization of the hyperreal numbers. The theory would be something like ZFC+CH, which we know suffices, and also we know that CH cannot be omitted for this. This would have provided enormous extrinsic justification for the continuum hypothesis, for this axiom would be seen as necessary for making sense of one of the core number systems underlying calculus. Thus, we would view CH as a fundamental principle necessary for mathematics and indispensable in the foundations of calculus.

To be sure, I am not arguing that CH already is or should be considered this way as fundamental, but rather only that it could have been. Thus, I claim, we must face a certain degree of contingency in our fundamental theories. What we consider to be bedrock foundational principles could have been different—we could have seen the continuum hypothesis as fundamental.

References

[Ber13] Sharon Berry. "Default reasonableness and the mathoids". *Synthese* 190.17 (2013), pp. 3695–3713. ISSN: 00397857, 15730964. http://www.jstor.org/stable/24019906.

[Ber34] George Berkeley. *A Discourse Addressed to an Infidel Mathematician.* The Strand, 1734. https://en.wikisource.org/wiki/The_Analyst:_a_Discourse_addressed_to_an_Infidel_Mathematician.

[BW18] Tim Button and Sean Walsh. *Philosophy and Model Theory.* Oxford University Press, 2018, pp. xvi+517. ISBN: 978-0-19-879040-2; 978-0-19-879039-6. DOI: 10.1093/oso/9780198790396.001.0001.

[Can52] Georg Cantor. *Contributions to the founding of the theory of transfinite numbers.* Translated, and provided with an introduction and notes, by Philip E. B. Jourdain. Dover Publications, Inc., New York, 1952, pp. ix+211.

[Can95] Georg Cantor. "Beiträge zur Begründung der transfiniten Mengenlehre". *Mathematische Annalen* 46 (1895). (German), pp. 481–512. DOI: 10.1007/BF02124929.

[Can97] Georg Cantor. "Beiträge zur Begründung der transfiniten Mengenlehre". *Mathematische Annalen* 49 (1897). (German), pp. 207–246. DOI: 10.1007/BF01444205.

[Ded88] Richard Dedekind. "Was sind und was sollen die Zahlen? (What are numbers and what should they be?)" (1888). Available in Ewald, William B. 1996. *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, Vol. 2, 787–832. Oxford University Press, pp. 787–832.

[Dow84] Alan Dow. "On ultra powers of Boolean algebras". *Topology Proceedings* 9.2 (1984), pp. 269–291.

[EGH55] P. Erdős, L. Gillman, and M. Henriksen. "An isomorphism theorem for real-closed fields". *Ann. of Math. (2)* 61 (1955), pp. 542–554. ISSN: 0003-486X. DOI: 10.2307/1969812. https://doi.org/10.2307/1969812.

[Ehr12] Philip Ehrlich. "The absolute arithmetic continuum and the unification of all numbers great and small". *Bulletin of Symbolic Logic* 18.1 (2012), pp. 1–45. ISSN: 1079-8986. DOI: 10.2178/bsl/1327328438.

[Est77] J. Esterle. "Solution d'un problème d'Erdős, Gillman et Henriksen et application à l'étude des homomorphismes de $\mathcal{C}(K)$". *Acta Math. Acad. Sci. Hungar.* 30.1-2 (1977), pp. 113–127. ISSN: 0001-5954,1588-2632. DOI: 10.1007/BF01895655. https://doi.org/10.1007/BF01895655.

[Eul80] Leonhard Euler. "On the infinity of infinities of orders of the infinitely large and infinitely small" (1780). Translation by Jordan Bell, 2009. arXiv:0905.2254[math.HO].

[Ewa96] William Bragg Ewald. *From Kant to Hilbert.* Vol. 2. A Source Book in the Foundations of Mathematics. Oxford University Press, 1996.

[Fre86] Chris Freiling. “Axioms of symmetry: throwing darts at the real number line”. *J. Symbolic Logic* 51.1 (1986), pp. 190–200. ISSN: 0022-4812. DOI: 10.2307/2273955.

[G9̈0] Kurt Gödel. *Collected works. Vol. II.* Publications 1938–1974, Edited and with a preface by Solomon Feferman. The Clarendon Press, Oxford University Press, New York, 1990, pp. xviii+407. ISBN: 0-19-503972-6.

[GS07] Catherine Goldstein and Georges Skandalis. “Interview with A. Connes”. *European Mathematical Society Newsletter* 63 (2007), pp. 25–31.

[Ham03] Joel David Hamkins. “A simple maximality principle”. *Journal of Symbolic Logic* 68.2 (2003), pp. 527–550. ISSN: 0022-4812. DOI: 10.2178/jsl/1052669062. arXiv:math/0009240[math.LO]. http://wp.me/p5M0LV-2v.

[Ham12] Joel David Hamkins. “The set-theoretic multiverse”. *Review of Symbolic Logic* 5 (2012), pp. 416–449. DOI: 10.1017/S1755020311000359. arXiv:1108.4223[math.LO]. http://jdh.hamkins.org/themultiverse.

[Ham15] Joel David Hamkins. “Is the dream solution of the continuum hypothesis attainable?” *Notre Dame Journal of Formal Logic* 56.1 (2015), pp. 135–145. ISSN: 0029-4527. DOI: 10.1215/00294527-2835047. arXiv:1203.4026[math.LO]. http://jdh.hamkins.org/dream-solution-of-ch.

[Ham21] Joel David Hamkins. *Lectures on the Philosophy of Mathematics.* MIT Press, 2021. ISBN: 9780262542234. https://mitpress.mit.edu/books/lectures-philosophy-mathematics.

[Har10] G. H. Hardy. *Orders of infinity, the 'Infinitärcalcül' of Paul Du Bois-Reymond.* available on the Internet Archive at https://archive.org/embed/ordersofinfinity00harduoft. Cambridge University Press, 1910.

[Hod76] Wilfrid Hodges. “Läuchli's algebraic closure of Q”. *Mathematical Proceedings of the Cambridge Philosophical Society* 79.2 (1976), 289–297. DOI: 10.1017/S0305004100052282.

[HS20] Joel David Hamkins and Robin Solberg. “Categorical large cardinals and the tension between categoricity and set-theoretic reflection”. *Mathematics arXiv* (2020). Under review. arXiv:2009.07164[math.LO]. http://jdh.hamkins.org/categorical-large-cardinals/.

[Hun03] Edward V. Huntington. “Complete Sets of Postulates for the Theory of Real Quantities”. *Transactions of the American Mathematical Society* 4.3 (1903), pp. 358–370. ISSN: 00029947. http://www.jstor.org/stable/1986269.

[HW05] Joel David Hamkins and W. Hugh Woodin. “The necessary maximality principle for c.c.c. forcing is equiconsistent with a weakly compact cardinal”. *Math. Logic Q.* 51.5 (2005), pp. 493–498. ISSN: 0942-5616. DOI: 10.1002/malq.200410045. arXiv:math/0403165[math.LO]. http://wp.me/s5M0LV-nmpccc.

[Isa11] Daniel Isaacson. “The reality of mathematics and the case of set theory”. In: *Truth, Reference, and Realism.* Ed. by Zsolt Novak and Andras Simonyi. Central European University Press, 2011, pp. 1–76.

[Jes93] Douglas M. Jesseph. *Berkeley's Philosophy of Mathematics.* 1st. Science and Its Conceptual Foundations series. University of Chicago Press, 1993. ISBN: 0226398978; 9780226398976.

[Kei00] H. Jerome Keisler. *Elementary Calculus: An Infinitesimal Approach.* Earlier editions 1976, 1986 by Prindle, Weber, and Schmidt; free electronic edition available. 2000. https://www.math.wisc.edu/~keisler/calc.html.

[Koe23] Peter Koellner. “The Continuum Hypothesis”. In: *The Stanford Encyclopedia of Philosophy.* Ed. by Edward N. Zalta and Uri Nodelman. Winter 2023. Metaphysics Research Lab, Stanford University, 2023.

[Lä62] H. Läuchli. “Auswahlaxiom in der Algebra”. *Commentarii mathematici Helvetici* 37 (1962/63). (German), pp. 1–18. http://eudml.org/doc/139241.

[Mad88] Penelope Maddy. “Believing the Axioms, I”. *The Journal of Symbolic Logic* 53.2 (1988), pp. 481–511.

[Moo10] Justin Tatch Moore. “The proper forcing axiom”. In: *Proceedings of the International Congress of Mathematicians. Volume II.* Hindustan Book Agency, New Delhi, 2010, pp. 3–29. ISBN: 978-81-85931-08-3; 978-981-4324-32-8; 981-4324-32-9.

[New46] Sir Isaac Newton. *Newton’s Principia : the mathematical principles of natural philosophy.* English translation by Andrew Motte. Published by Daniel Adee, 1846. https://archive.org/details/newtonspmathema00newtrich.

[Rit15] Colin J. Rittberg. “How Woodin changed his mind: new thoughts on the Continuum Hypothesis”. *Archive for History of Exact Sciences* 69.2 (2015), pp. 125–151. ISSN: 00039519, 14320657. http://www.jstor.org/stable/24569622 (version 20 January 2024).

[Roi82] J. Roitman. “Non-isomorphic hyper-real fields from non-isomorphic ultrapowers”. *Mathematische Zeitschrift* 181 (1982), pp. 93–96. DOI: 10.1007/BF01214984.

[Sol06] Robert Solovay. *An example of an axiomatizable second order theory that is complete but non-categorical?* FOM post. 2006. http://cs.nyu.edu/pipermail/fom/2006-May/010561.html.

[SVW24] Tapio Saarinen, Jouko Väänänen, and William Hugh Woodin. *On the categoricity of complete second order theories.* 2024. arXiv:2405.03428[math.LO]. https://arxiv.org/abs/2405.03428.

[Wil16] Timothy Williamson. “Absolute provability and safe knowledge of axioms”. In: *Gödel’s disjunction.* Oxford Univ. Press, Oxford, 2016, pp. 243–253. ISBN: 978-0-19-875959-1.

[Zer30] Ernst Zermelo. “Über Grenzzahlen und Mengenbereiche”. *Fundamenta Mathematicae* 16 (1930). Translated in [Ewa96], pp. 29–47.

(Joel David Hamkins) O’Hara Professor of Logic, University of Notre Dame, 100 Malloy Hall, Notre Dame, IN 46556 USA, & V. Research Fellow, Mathematical Institute, University of Oxford, UK.

Email address: jdhamkins@nd.edu

URL: http://jdh.hamkins.org

Aristotle meets Frege: from potentialism to Frege arithmetic

Stewart Shapiro

1 Introduction

The abstractionist, neo-logicist program in the philosophy of mathematics began with Crispin Wright's seminal [26]. Bob Hale [9] joined the cause, and it continues through many extensions, objections, and replies to objections (see Hale and Wright [10]). The program's overall plan is to develop branches of established mathematics using abstraction principles in the form:

$$\forall a \forall b(\Sigma(a) = \Sigma(b) \leftrightarrow E(a, b)), \tag{ABS}$$

where a and b are variables of a given type (typically first-order or monadic second-order), Σ is an operator, denoting a function from items of the given type to objects in the range of the first-order variables, and E is an equivalence relation over items of the given type.

Gottlob Frege [6], [7] employed at least three equations in the form (ABS). One of them, used for illustration, comes from geometry:

> The direction of l_1 is identical to the direction of l_2 if and only if l_1 is parallel to l_2.

The second was dubbed $N^=$ in [26] and is now called *Hume's Principle*:

$$\forall F \forall G(\#F = \#G \leftrightarrow F \approx G), \tag{HP}$$

where $F \approx G$ is an abbreviation of the second-order statement that there is a one-to-one relation mapping the F's onto the G's. In words, (HP) states that the number of F is identical to the number of G if and only if F is equinumerous with G. Georg Cantor deployed this principle to obtain extensive and profound results, especially concerning the transfinite.[1]

Unlike the direction-principle, the relevant variables, F, G here are second-order. We will follow the literature and refer to items in the range of these variables as "Fregean concepts", or sometimes just "concepts", essentially properties construed extensionally (and not necessarily identified with mental phenomena).

Frege's [7] third exemplar of an abstraction principle is the infamous Basic Law V:

$$\forall F \forall G(\epsilon F = \epsilon G \leftrightarrow \forall x(Fx \leftrightarrow Gx)). \tag{BLV}$$

Like Hume's Principle, Basic Law V is second-order, but unlike Hume's Principle, it is inconsistent (at least with classical or intuitionistic logic).

As is now well-known, Frege's *Grundlagen* [6] and *Grundgesetze* [7] contain the essentials of a derivation of the Dedekind-Peano postulates from Hume's Principle, plus some more or less straightforward definitions.[2] This result, now called *Frege's Theorem*, reveals that Hume's Principle, together with suitable definitions, entails that there are infinitely many natural numbers. The development of arithmetic from (HP) is sometimes called *Frege arithmetic*. This theory is taken to be the first success story of abstractionist neo-logicism. The underlying theme is that one can introduce (HP) as a sort of stipulative, implicit definition of the "number-of" operator, and develop arithmetic from that. There is an ongoing program of

[1] More details to follow. By rights, this abstraction principle should be called "Cantor's Principle", but the name "Hume's Principle" has caught on.

[2] Frege [7] used Basic Law V to derive the two conditionals in (HP). The rest of the Dedekind-Peano postulates follow from those.

attempting to found other, richer mathematical theories on abstraction principles. Here, we will only be concerned with arithmetic and (HP).

In their informal discussion of the abstractionist program, Wright and Hale occasionally speak of mathematical objects as "generated" by the abstraction principle (e.g., [10], pp. 19, 224, 237n, 278, 289, 412, 414), but for them, this term is only a metaphor. Their version of abstractionism is not a potentialist enterprise, as the quantifiers in the description of the equivalence relation on the right hand side are explicitly intended to include the "generated" abstracts, cardinal numbers in this case. Frege's Theorem depends on this. Sometimes, the word "generated" appears in scare-quotes, as in:

> One obvious danger here arises from the fact an equivalence relation defined on the concepts on a specified underlying domain of objects may partition those concepts into more equivalence classes than there are objects in the underlying domain, so that a second-order abstraction may 'generate' a domain of abstracts strictly larger than the initial domain of objects. This, in itself, need be no bad thing—indeed, it is essential, if there is to be a neo-Fregean abstractionist route to (classical) analysis. ([10], p. 19])

The plan here is to present a genuinely potentialist account of Frege arithmetic. The (cardinal) numbers are not generated from (HP), but rather from more or less standard principles of potentialism. The relevant version of (HP) is a principle stating a condition for numbers to be identical with each other. So the account is not an abstractionist one. Essentially, (HP) tells us what we are generating—cardinal numbers—but the generation does not go through (HP) itself.

The perspective here is that of an Aristotelian potentialist who rejects even the possibility of an actual infinity (as presented in Linnebo and Shapiro [17]). We also develop an Aristotelian, potentialist set theory—in effect, a theory of hereditarily finite sets—a theory that is definitionally equivalent to Dedekind-Peano arithmetic. The orientation is deductive: we articulate an axiomatic (higher-order and plural) language in which to express the main principles, and explore what can be deduced in order to express and sustain the potentialist insights (if that is what they are).

This is in contrast with the lovely study Stafford [23], which draws on Hodes [11]. Like the present project, Stafford treats potentialist arithmetic, drawing on (HP), but its orientation is "semantic", i.e., model-theoretic. Using a background set theory, presumably full ZFC (or perhaps Zermelo set theory), he develops Kripke structures for a modal language, structures in which each world is finite. He shows how to interpret a standard first-order (but not second-order or plural) version of Dedekind-Peano arithmetic in the model-theoretic structures. This particular model-theoretic perspective is, of course, not available to an Aristotelian, since the meta-theory makes heavy use of actual infinity. In the developed framework, each world is finite, but the Kripke structure itself has infinitely many worlds.

Stafford notes that he leaves the "development of a deductive theory for future work" (p. 557). Although the present project is deductive, it does not recapitulate Stafford's theory. The target here is full second-order Dedekind-Peano arithmetic, but as noted, Stafford's theory does not satisfy that.[3]

The present paper is self-contained. The next Section 2 provides a brief overview of the potentialist perspective—for more details, see [17]. We present an axiom, called "(Aristotle)", that entails that infinite concepts and pluralities are impossible—all worlds are (Dedekind) finite. Section 3 provides a sketch of Frege's own development of arithmetic (based on Hume's Principle (HP)), giving the usual definitions. Section 4 shows how to formulate (HP) in the modal, potentialist setting, along with axioms entailing the possible existence of various numbers. Then, in Section 5, we formulate definitions and show how to derive the relevant analogues of the Dedekind-Peano axioms. Section 6 drops the (Aristotle) axiom. At that point, we do not assert the possibility of an actual infinity; rather, the theory is officially neutral on whether there is or there could be an actual infinity. The development there is a bit closer to Frege's own treatment, along with that of the abstractionist neo-logicists. The final Section 7 presents an Aristotelian set theory (of hereditarily finite sets), and shows that arithmetic can be interpreted in that theory in the usual way.

[3] Thanks to a referee for pressing this comparison, and to Tim Button for an insightful exchange on the issues.

2 Potential infinity — a crash course

Aristotle famously rejected the actual infinite—the existence of any complete collection with infinitely many members. He argued that the only sensible notion is that of potential infinity. In *Physics* 3.6 (206a27-29), he wrote:

> For generally the infinite is as follows: there is always another and another to be taken. And the thing taken will always be finite, but always different (2o6a27-29).

As Richard Sorabji [22] (pp. 322-3) once put it, for Aristotle, "infinity is an extended finitude"(see also [14]). This attitude toward the infinite was expressed by the vast majority of mathematicians and philosophers at least until late in the nineteenth century. In 1831, for example, Gauss [8] wrote:

> I protest against the use of infinite magnitude as something completed, which is never permissible in mathematics. Infinity is merely a way of speaking.

In line with the mathematicians of his day, Aristotle did accept what is sometimes called *potential infinity*, against the ancient atomists (see [19]). Mathematicians in antiquity followed this, and, indeed, made brilliant use of potential infinity. But what is potential infinity?

Either directly or indirectly, the idea seems to be that potential infinity is tied to certain *procedures* that can be repeated indefinitely. A nice example is provided by Aristotle's claim, against the atomists, that matter is infinitely divisible. Consider a body of mud. However many times one has divided the mud, it is always possible to divide it again—or so it is assumed.

As indicated by the term "it is possible", the thesis here can be explicated in a modal way.[4] This yields the following analysis of the infinite divisibility of the body of mud s:

$$\Box\forall x(Pxs \rightarrow \Diamond\exists y Pyx), \tag{1}$$

where the variables range over parts of the given body of mud, and Pxy means that x is a *proper* part[5] of y. If the parts of the mud formed an actual infinity, the following would hold:

$$\forall x(Pxs \rightarrow \exists y Pyx). \tag{2}$$

According to Aristotle, it is impossible for there to be infinitely many divisions of the mud, existing all at once:

$$\neg\Diamond\forall x(Pxs \rightarrow \exists y Pyx). \tag{3}$$

By endorsing both (1) and (3), one is is asserting that the divisions of the mud are *merely potentially infinite*.

As noted, present concern is with mathematics, and the natural numbers in particular. According an Aristotelian, the sequence of natural numbers is merely potentially infinite. This can be represented as the conjunction of the following theses:

$$\Box\forall m\Diamond\exists n\, \textsc{Succ}(n,m) \tag{4}$$

$$\neg\Diamond\forall m\exists n\, \textsc{Succ}(n,m), \tag{5}$$

where $\textsc{Succ}(n,m)$ states that n comes right after m. The modal language thus provides a nice way to distinguish the merely potential infinite from the actual infinite.

Linnebo and Shapiro [17] develop an account that can accept some actually infinite collections, and still leaves room to insist that some other "totalities" are merely potentially infinite. The analysis provides a framework in which actual and potential infinity can live side by side, sometimes in the very same system. Here the focus is on a more Aristotelian perspective that allows no actually infinite collections. Our goal is to vindicate, for arithmetic,[6] Aristotle's claim about geometry:

[4] I make use of contemporary modal notions here. There is no attempt to recapitulate what Aristotle himself says about modality.

[5] A referee notes that this presupposes that the "parts" of the mud all have non-zero measure. In particular, the mud that occupies an extensionless point would have no proper parts, thus violating (1). However, Aristotle explicitly insisted that points are not parts of anything—see, for example, [12], Chapter 1.

[6] There is some anachronism here. As far as we know, there was nothing like full Dedekind-Peano arithmetic in Aristotle's day.

> Our account does not rob the mathematicians of their study, by disproving the actual existence of the infinite in the direction of increase, in the sense of the untraversable. In point of fact they do not need the <actually> infinite and do not use it. They postulate only that the finite straight line may be produced as far as they wish. (207b27-30)

2.1 Three orientations towards the infinite

It is useful to distinguish different orientations towards a given infinite totality, such as the natural numbers or the parts of a given body of mud (according to Aristotle). *Actualism* accepts actual infinities, of the given kind, and thus finds no use for modal notions—or at least no use that is specific to the analysis of the infinity in question. Actualists maintain that the non-modal language of ordinary mathematics is already fully explicit and thus deny that we need a translation into some modal language. Furthermore, actualists accept classical logic when reasoning about the infinite (or the infinite in question).

Potentialism is the orientation that stands opposed to actualism. Accordingly, the objects with which mathematics is concerned—or some of the objects with which mathematics is concerned—are generated successively, and at least some of these generative processes cannot be completed. Present concern is with the natural numbers. Our potentialist thinks of numbers as generated, presumably one at a time.

There are (at least) two different forms of potentialism. As characterized above, potentialism is the view that some or, in the present case, all of the objects with which mathematics is concerned are successively generated. What about *the truths* of mathematics? Of course, on any form of potentialism, these are modal truths concerned with certain generative processes. But how should these modal truths be understood?

Liberal potentialists regard the modal truths as unproblematic, adopting bivalence for the modal language, and excluded middle for the underlying modal logic. Consider Goldbach's conjecture. As potentialists interpret it, the conjecture says that necessarily any even natural number, greater than two, that is ever generated can be written as a sum of two primes. Liberal potentialists maintain that this modal statement has a truth-value—it is either true or false. Their approach to modal theorizing in mathematics is thus much like a realist approach to modal theorizing in general: there are objective truths about the relevant modal aspects of reality, and this objectivity warrants the use of some classical form of modal logic.

Strict potentialists differ from liberal potentialists by requiring, not only that every object generated at some stage of a process, but also that every truth be "made true" at some stage. Consider, again, the Goldbach conjecture. If there are counterexamples to the conjecture, then its negation will presumably be "made true" at the stage where the first counterexample is generated. But suppose there are no counterexamples. Since the conjecture is concerned with *all* the natural numbers, it is hard to see how it could be "made true" without completing the generation of natural numbers. This completion would, however, violate the strict potentialists' requirement that any truth be made true at some stage of the process.

Linnebo and Shapiro [17] argue that strict potentialists should adopt a modal logic whose underlying logic is intuitionistic (or intermediate between classical and intuitionistic logic). This allows them to adopt a conception of universal generality which does not presuppose that all the instances are ever "available" at a stage. In particular, strict potentialists should not accept every instance of the law of excluded middle in the background modal language. For the most part, we adopt the liberal perspective here, at least partly because we wish to recapitulate a version of classical arithmetic.

2.2 The modality and the modal logic

It is useful here to invoke the contemporary heuristic of possible worlds when discussing the modality in question. But it is insisted that this is *only* heuristic, as a manner-of-speaking (unlike the treatment in [23], see note 1). The theory is formulated in the modal language, with (one or both of) the modal operators as primitive. The modal language is not explained or defined in terms of anything else.

The potentialist does, of course, reject the now common thesis that mathematical objects exist of necessity—if they exist at all. To invoke the heuristic, the now common thesis is that all mathematical objects exist in all worlds. The potentialist gives that up. There is no world with all of the objects in question—all natural numbers in the present case. Nor, of course, is there an actual infinity of possible worlds (again, unlike [23]).

What about the philosophical nature of the modality invoked in the analysis of potentiality? For the Aristotelian, it can perhaps be an ordinary metaphysical modality invoked in contemporary philosophy (or perhaps defined from that notion)—waiving the now widely held thesis that mathematical objects exist necessarily if they exist at all. For that matter, the modality can also be a very ordinary circumstantial modality, as studied in linguistic semantics (suitably idealized, of course).

The idea is that natural numbers are generated in time. At any stage—in any world—there are finitely many natural numbers, but each such world has access to another where some more numbers have been generated. Given enough time, any natural number can be generated, even though there is no time when they are all generated.

Following the heuristic, we assume that every possible world is finite, in the sense that it contains only finitely many objects. For convenience, we wish to avoid invoking a free logic, at least here. So we do not wish to countenance objects—mathematical or otherwise—going out of existence. To paraphrase Aristotle, we study generation, but not corruption. This entails that the domains of the possible worlds grow (or, better, are non-decreasing) along the accessibility relation. So we assume:

$$w_1 \leq w_2 \rightarrow D(w_1) \subseteq D(w_2) \tag{6}$$

where '$w_1 \leq w_2$' says that w_2 is accessible from w_1, and for each world w, $D(w)$ is the domain of w.

Again, for convenience, in this initial foray, one can think of a possible world as determined completely by the objects it contains. So we assume the converse of (6). This motivates the following principle:

> **Partial ordering**: The accessibility relation $\leq$ is a partial order. That is, it is reflexive, transitive, and anti-symmetric.

So the underlying modal logic is at least S4. As is well-known, the conditional (6) entails that the converse Barcan formula is valid. That is,

$$\exists x \Diamond \phi(x) \rightarrow \Diamond \exists x \varphi(x). \tag{CBF}$$

So far, then, we have S4 plus (CBF).[7]

We also assume, for simplicity (and convenience) that the only things generated are mathematical. If we think of the natural numbers as generated one at a time, in order, it is perhaps natural to assume that the possible worlds have a linear ordering. But it is useful to tie the present framework in with other, more general ones, in order to invoke, or eat least discuss, some existing results. At any stage in a process of construction, we generally have a choice of which objects to generate. For some types of construction, but not all, it makes sense to require that a license to generate objects is not revoked at accessible worlds. Intuitively, geometric construction is like this. For example, we might have, at some stage, two intervals that don't yet have bisections. We can choose to bisect one or the other of them, or perhaps to bisect both simultaneously. Assume we are at a world w_0 where we can choose to generate objects, in different ways, so as to arrive at either w_1 or w_2. Say at w_1 we bisect an interval i and at w_2 we bisect another interval j. It seems plausible to require that the license to bisect i can be executed at w_2 or any later world. In other words, nothing we do can prevent us from being able to bisect the other interval.

This corresponds to a requirement that any two worlds w_1 and w_2 accessible from a common world have a common extension w_3. This is a directedness property known as *convergence* and formalized as follows:

$$\forall w_0 \forall w_1 \forall w_2 (w_0 \leq w_1 \wedge w_0 \leq w_2 \rightarrow \exists w_3 (w_1 \leq w_3 \ \wedge \ w_2 \leq w_3))$$

For constructions that have this property, then, we adopt the following principle:

> **Convergence**: The accessibility relation $\leq$ is convergent.

This principle ensures that, whenever we have a choice of mathematical objects to generate, the order in which we choose to proceed is irrelevant. Whichever object(s) we choose to generate first, the other(s)

[7] Recall that S4 and (non-free) first-order logic entails (CBF). We can also require the accessibility relation to be well-founded, on the grounds that all mathematical construction has to start somewhere. However, nothing of substance turns on this here.

can always be generated later. Unless $\leq$ is convergent, our choice whether to extend the ontology of w_0 to that of w_1 or that of w_2 might have an enduring effect.[8]

It is well known that the convergence of $\leq$ ensures the soundness of the following principle:

$$\Diamond\Box p \to \Box\Diamond p. \tag{G}$$

The modal propositional logic that results from adding this principle to a complete axiomatization of S4 is known as S4.2.

2.3 The logic of potential infinity

What is the correct logic when reasoning about potentially infinite collections? Informal glosses aside, the language of mathematics is usually non-modal. We thus need a translation to serve as a bridge connecting the non-modal language in which mathematics is ordinarily formulated with the modal language in which our analysis of potentiality is developed. Suppose we adopt a translation $*$ from a non-modal language $\mathcal{L}$ to a corresponding modal language $\mathcal{L}^{\Diamond}$. The question of the right logic of potential infinity is the question of which entailment relations obtain in $\mathcal{L}$.

To determine whether $\varphi_1, \ldots, \varphi_n$ entail ψ, we need to (i) apply the translation and (ii) ask whether $\varphi_1^*, \ldots, \varphi_n^*$ entail ψ^* in the modal system. This means that the right logic of potential infinity depends on several factors. First, the logic depends on the bridge that we choose to connect the non-modal language of ordinary mathematics with the modal language in which our analysis of potential infinity is given. Second, the logic obviously depends on our modal analysis of potential infinity; in particular, on the modal logic that is used in this analysis—in particular, on whether the underlying logic of the modal language is classical or intuitionistic. Let us now turn to the first factor.

The heart of potentialism, or at least the present explication of potentialism, is the idea that the existential quantifier of ordinary non-modal mathematics has an implicit modal aspect. In the developed interpretive program, a statement that a given number has a successor is interpreted as a statement that this number *potentially* has a successor—that it is *possible* to generate a successor. This suggests that the right translation of $\exists$ is $\Diamond\exists$.

Similarly, a potentialist statement that a given property holds of all objects (of a certain sort) is interpreted as a statement that the property holds of all objects (of that sort) *whenever they are generated.* This suggests that $\forall$ be translated as $\Box\forall$.

Thus understood, the quantifiers of ordinary non-modal mathematics are devices for generalizing over absolutely all objects, not only the ones available at some stage, but also any that we may go on to generate. In our modal language, these generalizations are effected by the strings $\Box\forall$ and $\Diamond\exists$. Although these strings are strictly speaking composites of a modal operator and a quantifier proper, they behave logically just like quantifiers ranging over all entities at all (future) worlds.

The proposal is thus that each quantifier of the non-modal language is translated as the corresponding modalized quantifier. Each connective is translated as itself. Let us call this the *potentialist translation*, and let $\varphi^{\Diamond}$ represent the translation of φ. We say that a formula is *fully modalized* just in case all of its quantifiers are modalized. Clearly, the potentialist translation of any non-modal formula is fully modalized.

Say that a formula φ is *stable* if the necessitations of the universal closures of the following two conditionals hold:

$$\varphi \to \Box\varphi \qquad\qquad \neg\varphi \to \Box\neg\varphi$$

This sets the stage for two key results, which answer the question about the correct logic for those kinds of potentiality that enjoy the above convergence property.

For the first, let $\vdash$ be the relation of classical deducibility in a non-modal first-order language $\mathcal{L}$. Let $\mathcal{L}^{\Diamond}$ be the corresponding modal language, and let $\vdash^{\Diamond}$ be deducibility in the modal system consisting of classical $\vdash$, S4.2, and axioms asserting the stability of all atomic predicates of $\mathcal{L}$.

[8] Other types of "generation" are not like this. Suppose for example, that I can bake bread or I can bake a cake, since I have enough time and ingredients to do either. But it may be that if I bake bread, then I can no longer bake a cake, since I may have used up the needed ingredients or I won't have enough time. Or suppose that a country has a law that a couple can have only two children. If a given couple already has one, then it is possible for them to have a boy and it is possible for them to have a girl. But if they do one, they will not be allowed to go on and do the other.

Classical potentialist mirroring. For any formulas $\varphi_1, \ldots, \varphi_n, \psi$ of $\mathcal{L}$, we have:

$$\varphi_1, \ldots, \varphi_n \vdash \psi \quad \text{if and only if} \quad \varphi_1^\Diamond, \ldots, \varphi_n^\Diamond \vdash^\Diamond \psi^\Diamond.$$

(See [15] for a proof.)

The theorem has a simple moral. Suppose we are interested in logical relations between formulas in the range of the potentialist translation, in a classical, first-order modal theory that includes S4.2 and the stability axioms. Then we may delete all the modal operators and proceed by the ordinary non-modal logic underlying $\vdash$. In particular, under the stated assumptions, the modalized quantifiers $\Box\forall$ and $\Diamond\exists$ behave logically just as ordinary quantifiers, except that they generalize across all (accessible) possible worlds rather than a single world.[9]

As noted above, however, Linnebo and Shapiro [17] argue that a stricter form of potentialism pushes in the direction of intuitionistic logic. There is a second mirroring theorem. In a system governed by intuitionistic logic, a formula φ is said to be *decidable* if the universal closure of $\varphi \vee \neg\varphi$ is deducible in that theory. Let $\vdash_{\text{int}}$ be the relation of intuitionistic deducibility in a first-order language $\mathcal{L}$, together with axioms stating the decidability of all atomic formulas, and let $\vdash_{\text{int}}^\Diamond$ be deducibility in the modal language corresponding to $\mathcal{L}$, by $\vdash_{\text{int}}$, S4.2 and the stability axioms for all atomic predicates of $\mathcal{L}$.[10] The conclusion is then the same as above:

Intuitionistic potentialist mirroring For any formulas $\varphi_1, \ldots, \varphi_n, \psi$ of $\mathcal{L}$, we have:

$$\varphi_1, \ldots, \varphi_n \vdash_{\text{int}} \psi \quad \text{if and only if} \quad \varphi_1^\Diamond, \ldots, \varphi_n^\Diamond \vdash_{\text{int}}^\Diamond \psi^\Diamond.$$

(See [17] for a proof.)

Our interest won't always be limited to formulas in the range of the potentialist translation. One can often use the extra expressive resources afforded by the modal language to engage in reasoning that takes us outside of this range. The modal language allows us to look at the subject matter under a finer resolution, which can be turned on and off, according to our needs.

The final order of business for this section is a case in point. We state an axiom that enforces the Aristotelian thesis that—to invoke the heuristic—all worlds are finite. Our choice is that it is necessary that for any Fregean concept X and for any relation R on X that is a one-to-one function on X, R is a surjection on X (i.e., if Xx then x has an R predecessor):

$$\begin{aligned}\Box\forall R\forall X[\forall x(((Xx \to \exists y\forall z(Xz \wedge Rxz) \leftrightarrow y = z)) \wedge \forall x_1\forall x_2\forall y((Rx_1y \wedge Rx_2y) \to x_1 = x_2))\\ \to \forall y(Xy \to \exists x(Xx \wedge Rxy)))] \quad \text{(Aristotle)}\end{aligned}$$

This entails that all worlds are at least Dedekind finite.

Intuitively, a Fregean concept, or a possible world, is finite if its cardinality is a natural number. To be sure, it would be premature to invoke that here, since we have not yet defined the natural numbers. But we can express a statement that a concept or world is actually finite (and not just Dedekind finite). Let R be any binary relation, Frege [5] defined the (weak) *ancestral* R^* of R.[11] Using a more contemporary framework the definition is as follows:

$$R^*xy \leftrightarrow_{\text{def}} [\forall X(Xx \wedge (\forall z\forall w((Xz \wedge Rzw) \to Xw))) \to Xy] \qquad \text{(Ancestral)}$$

In words, R^*xy holds if y has every Fregean concept that holds of x and is closed under R. In effect, R^*xy holds just in case either $x = y$ or there is a finite sequence $a_1 \ldots a_{n+1}$ such that $a_1 = x, a_{n+1} = y$ and for each m such that $1 \leq m < n$, $R_{a_m a_{m+1}}$.

[9] Note the restriction to first-order languages. The result will apply to higher-order languages if the modal translations of the instances of the comprehension scheme are deducible in the modal language.

[10] A referee points out that identity is not decidable in intuitionistic analysis (or in smooth infinitesimal analysis) and membership is not decidable in some intuitionistic set theories. The relevant moral is that this mirroring theorem does not hold for those theories.

[11] From here on, when we speak of the ancestral or an ancestor, we mean the weak ancestral and a weak ancestor.

We can state a principle that asserts, in effect, that all worlds are finite as follows:

$$\Box\exists R(\forall x\forall y_1\forall y_2((Rxy_1 \wedge Rxy_2) \to y_1 = y_2) \wedge \exists x\forall z R^*(x,z) \wedge \exists y\forall z(\neg R(y,z))) \qquad \text{(finite)}$$

In words, (finite) says that necessarily (i.e., in every world) there is one-to-one relation R such that (i) there something x such that everything (in that world) is an R-ancestor of x, and (ii) there is something z that does not bear R to anything (in that world). Intuitively, it follows that in every world, every concept is actually finite in that world. This principle is probably closer to Fregean concerns, and it allows many of the proofs to be constructive.

Nevertheless, we shall stick with the weaker (Aristotle) here. After some other axioms are added, we can derive (finite), showing that all worlds and thus all Fregean concepts are actually finite. It follows that, necessarily (i.e., in every world w), there could be a natural number (possibly in an accessible world w') that is the number of objects (in w).

3 A sketch of Frege's Theorem

Recall the ill-named Hume's Principle:

$$\forall F\forall G(\#F = \#G \leftrightarrow F \approx G), \qquad \text{(HP)}$$

where $F \approx G$ is an abbreviation of the second-order statement that there is a one-to-one relation mapping the F's onto the G's. That is:

$$F \approx G \leftrightarrow_{\text{def}} \exists R[\forall x(Fx \to \exists! y(Gy \wedge Rxy)) \wedge \forall y(Gy \to \exists! x(Xx \wedge Rxy))]$$

A central component of (this phase of) the abstractionist program is to establish Frege's Theorem: a derivation of the Dedekind-Peano axioms from (HP) and reasonable definitions of the primitive arithmetical vocabulary. This, it is argued, provides a logical and epistemological foundation for arithmetic. The purpose of this section is to provide a sketch of Frege's Theorem, or at least of some key steps along the way. Then we turn to a potentialist version of the result.

The abstractionist does not presuppose, at the outset, just as a matter of syntax and semantics, that every Fregean concept F has a (unique) number $\#F$. So the proper background would be a free logic. Their practice is to start with a concept F, note the trivial consequence that $F \approx F$, and then *conclude* that F has a number. So perhaps the proper background is a negative free logic in which identity statements, or perhaps atomic statements, entail existence, so that $a = b$ entails that a and b denote something.

Frege defines *zero* to be the number of the Fregean concept of being not self-identical, the concept characterized by the open formula $x \neq x$. The next item is the successor relation. Frege [6], §76 writes:[12]

> I now propose to define the relation in which every two adjacent members of the series of [cardinal] numbers stand to each other. The proposition:
>
> > "there exists a concept G, and an object falling under it x, such that the Number which belongs to the concept G is n and the Number which belongs to the concept 'falling under G but not identical with x' is m"
>
> is to mean the same as
>
> > "n follows in the series of ... numbers directly after m".

Frege's definition, then, is that n follows directly after m just in case

$$\exists F\exists G((m = \#F \wedge n = \#G) \wedge \exists x(Gx \wedge \forall y(Fy \leftrightarrow (y \neq x \wedge Gy)))) \qquad \text{(Frege-successor)}$$

The same definition is employed in Frege [7] as well as in Wright [26], p. 37.

[12] The notation is tweaked slightly.

The following is equivalent (via classical logic): the number n follows directly after the number m just in case:

$$\exists F \exists G\big(m = \#F \wedge n = \#G \wedge \exists x(\neg Fx \wedge \forall y(Gy \leftrightarrow (y = x \vee Fy)))\big) \qquad \text{(Successor)}$$

To put it in Fregean words, n follows directly after m if and only if:

> There exists a concept F, and an object x not falling under F, such that the Number which belongs to the concept F is m and the Number which belongs to the concept 'either falling under F or identical with x' is n.

Let us write $\mathsf{S}(x, y)$ for "y bears (Successor) to x", or, in Fregean terms, "y follows directly after x".

In [6], §76, Frege immediately notes that he does not use the expression "*the* Number following next after m", since he had not yet shown that each number has exactly one successor. Of course, one can, and Frege does, prove uniqueness—using classical logic. Finally, Frege defines a *natural number* to be an ancestor of zero under (Frege-successor):

$$\mathbb{N}x \leftrightarrow_{\text{def}} \mathsf{S}^*(0, x)$$

With these definitions, the usual Dedekind-Peano axioms follow. Every natural number has a unique successor, and so the successor relation is a function; this function is one-to-one; zero has no successor, and full, second-order induction holds:

$$\forall X[(X0 \wedge \forall x \forall y((Xx \wedge \mathsf{S}(x, y)) \to Xy)) \to \forall x(\mathbb{N}x \to Xx)]$$

Induction is a straightforward consequence of the use of the ancestral in the definition of $\mathbb{N}$.

Where in the above, more or less standard development of this instance of abstractionism, do we run afoul of the Aristotelian dictum to reject any and all actual infinities, any completed infinite totalities? From the background actualism, we have that there are infinitely many natural numbers. That follows from the Dedekind-Peano principles established by Frege's Theorem(s). So the Fregean concept expressed by the formula $\mathsf{S}^*(0, x)$ holds of infinitely many things, all of the natural numbers. Moreover, an application of the '#' operator to $\mathbb{N}$ would give us the existence of an infinite number, one that Frege called "*endlos*". In modern terms, that is the cardinal number $\aleph_0$.

The ancestral relation is an essential part of the development and, in particular, the proof that the natural numbers satisfy the induction principle. Recall that for any binary relation R, R^*xy holds if *every* concept that holds of x and is closed under R also holds of y. But every concept that holds of 0 and is closed under S holds of infinitely many things. So, for our Aristotelian, there are no such concepts and so $\mathbb{N}$ vacuously holds of everything! This motivates the development of a potentialist, account of arithmetic, along Aristotelian and Fregean lines.[13]

4 Potentialist arithmetic

Linnebo [15] and Chapter 3 of [16] presents a (consistent) account of set theory based on a potentialist version of Frege's Basic Law V. Instead of saying that every Fregean concept has an extension (which is inconsistent, given the usual comprehension principles of higher-order logic), we say that at least some Fregean concepts X *could* have an extension.

$$\Diamond \exists x(x = \epsilon(X))$$

[13] As is well-known, there are issues concerning how the underlying higher-order logic is to be developed in the framework. One issue is the analogue of impredicative comprehension. Sean Walsh [25] (Corollary 92) shows that full first-order Dedekind-Peano arithmetic PA cannot be interpreted in predicative Frege arithmetic—regardless of our definitions of the arithmetical primitives. It follows from [25], Corollary 92 and Proposition 6, that the interpretability strength of the system of second-order arithmetic known as ACA_0 is strictly above that of (HP) with Δ^1_1-comprehension, and hence also strictly above that of (HP) with predicative comprehension. These issues are not broached here.

To invoke the heuristic of possible worlds, if all of the objects that X holds of are in a single world w, then there is an accessible world that contains an extension whose members are all and only the objects that X holds of in w.

To formulate this, Linnebo uses the language of plural logic. It is stipulated that, like sets, pluralities are *rigid*: the same objects are among a given plurality xx in all worlds in which xx exists.[14] As Linnebo ([15] [16]) shows, this rigidity can be expressed in the underlying modal and plural language.

The set-existence principle is

$$\Box\forall xx\Diamond\exists y\forall z(z \in y \leftrightarrow z \prec xx). \qquad \text{(set exists)}$$

In words, for any objects, there could be a set whose members are just those objects. On pain of Russell's paradox, there is no plurality of all possible sets, no plurality of all possible von Neumann ordinals, and the like.

We do not adopt (set exists) here since we are not (yet) interested in a potentialist set theory (but see §7 below). The present framework has both higher-order variables (ranging over Fregean concepts) and plural variables (ranging over rigid pluralities). Concerning Fregean concepts, the framework has full, unrestricted comprehension: each instance of the universal closure of the following holds:

$$\Box\exists R\Box\forall x_1 \dots \forall x_n(Rx_1 \dots x_n \leftrightarrow \Psi),$$

where Ψ is any formula that does not have R free. That is, every formula in the language defines a Fregean concept.[15]

Note that a Fregean concept can have different extensions in different worlds. For example, the concept of being self-identical (defined by "$x = x$") holds of different things in different worlds. The concept of being the largest number also holds of different numbers in different worlds. The same goes for a concept like being a living dog—dogs come and, alas, go. In short, unlike pluralities, concepts are not rigid.

There also is a comprehension principle for pluralities. Each instance of the universal closure of the following holds:

$$\Box(\exists x\Psi \rightarrow \exists xx\forall x(x \prec xx \leftrightarrow \Psi)),$$

where Ψ is a formula that does not contain xx free. That is, every formula in the language that holds of something defines a (rigid) plurality in each world. So if $\exists x\Psi$ holds in a world w, then the instance of plural comprehension defines a plurality of all objects that satisfy Ψ in w. Of course, those objects need not satisfy Ψ in a different world.

To avoid a free logic, we replace the abstractionist number-of operator with a relation. So for a Fregean concept F, '$\#(F, n)$' can be read as "n is a number of F". And similarly '$\#(mm, n)$' can be read as "n is a number of mm".[16]

As noted, for simplicity (and to keep the logic unfree), we also assume that objects are never destroyed—if an object exists in a given world, it exists in all accessible worlds. So the accessibility relation will be at least that of S4, and we adopt the above policy and take the accessibility relation to be convergent (if not linear).

4.1 Numbers of pluralities

In §2 we formulated and adopted a principle (Aristotle) that entails that, in effect, all worlds are Dedekind finite. It follows that all pluralities are Dedekind finite. We thus adapt (set exists) to numbers and adopt:

$$\Box\forall xx\Diamond\exists y\#(xx, y). \qquad \text{(Num Exists)}$$

In words, every plurality could have a number.

[14] Clearly, if something is rigid then it is stable, but the converse typically fails.

[15] Since there are no restrictions on the formula Ψ, the system allows impredicative definitions. The same goes for plural comprehension. This is taken to be of-a-piece with a liberal view of potentialism. In any case the goal here is to sanction full induction on the natural numbers, and thus an analogue of full second-order arithmetic. Full impredicative comprehension facilitates that.

[16] Some readers (and one referee) find it awkward or annoying or infelicitous to use the same symbol ('#') for both numbers of pluralities and numbers of Fregean concepts, even though context always settles which is meant. It is, of course, simple and straightforward to use two different symbols instead.

Define equinumerosity on pluralities as follows:

$$xx \approx yy \leftrightarrow_{\text{def}} \exists R(\forall x(x \prec xx \to \exists! y(Rxy \wedge y \prec yy)) \wedge \forall y(y \prec yy \to \exists! x(Rxy \wedge x \prec xx))).$$

We do not have that every plurality has a number (see below for more details on this). The relevant version of Hume's Principle is that if two pluralities each have a number, then these numbers are identical just in case the pluralities are equinumerous:

$$\forall xx \forall yy \forall x \forall y((\#(xx, x) \wedge \#(yy, y)) \to (x = y \leftrightarrow xx \approx yy)). \qquad (\mathsf{HP})$$

It follows that each plurality has at most one number:

$$\Box \forall xx \forall y \forall z((\#(xx, y) \wedge \#(xx, z)) \to y = z).$$

We add an axiom that if two pluralities are equinumerous and one of them has a number, then so does the other:

$$\forall xx \forall yy \forall x((xx \approx yy \wedge \#(xx, x)) \to \exists y \#(yy, y)). \qquad (\mathsf{HP}')$$

Of course, it follows from this and (HP) that these two numbers are identical: if two pluralities are equinumerous and one has a number, then the other has the same number.[17]

It seems reasonable to add a principle that if a certain number exists (in a given world) then so do all smaller numbers:

$$\Box \forall xx \forall yy \forall x((\#(xx, x) \wedge \forall z(z \prec yy \to z \prec xx)) \to \exists w \#(yy, w)). \qquad \text{(Closure Down)}$$

The idea is that the numbers are generated one at a time, in their natural order.

One does not have to think of the numbers as generated this way. One might instead allow a given plurality to get a number before some of the its sub-pluralities have numbers. In that case, the relevant principle would be that if a plurality has a number, then it *could be* that all of its sub-pluralities have numbers:

$$\Box \forall xx \forall x(\#(xx, x) \to \Diamond \forall yy(\forall z(z \prec yy \to z \prec xx) \to \exists w \#(yy, w))). \qquad \text{(potential closure down)}$$

The principle (potential closure down) is a kind of generalization on the (G) axiom. From (G) and (Num Exists) we have that for any given finite list of pluralities in a given world, there is an accessible world that contains numbers for all of them. But, it seems, one cannot prove this generalization, at least as it is stated. Moreover, our (Aristotle) principle only entails that all worlds are Dedekind finite. And we certainly have no guarantee (yet) that the reasoning extends to Dedekind finite collections of pluralities.[18]

So far, we have the possibility of a world that contains only numbers, and which has a number for every plurality in that world. Consider, for example, a world whose domain is the first five numbers, starting with one. But we have such a world (or such a possibility) only because we have not yet introduced zero as a number. Following standard practice, there is no "empty" plurality, and so there is no plurality whose number is zero.

4.2 Numbers of Fregean concepts

One reason that we treat numbers of Fregean concepts here is to explicate Frege's account of how arithmetic is applied. Frege [6], §§45-46 asks what is it that (cardinal) numbers are numbers of? What is it that we count? What do we apply number words to? His answer is that we apply numbers to concepts:

> While looking at one and the same external phenomenon, I can say with equal truth both "It is a copse" and "It is five trees", or both "Here are four companies" and "Here are 500 men". Now what changes here from one judgment to the other is neither any individual object, nor the whole, the agglomeration of them, but rather my terminology. But that is itself only a sign that one concept has been substituted for another. This suggests as the answer [to the question is] that the content of a statement of number is an assertion about a concept. This is perhaps clearest with the number 0. (§46)

[17] There is no need for an analogous axiom Linnebo's [15] [16] treatment of set theory. If two pluralities are coextensive, then they are the same plurality.

[18] Thanks to Øystein Linnebo for these observations.

This account is also adopted by the abstractionists.

We introduce zero as the number of a Fregean concept. One option is to add a constant "0" to the language, along with the axiom:

$$\Box\forall F(\#(F, 0) \leftrightarrow \neg\exists xFx).$$

This entails that zero exists in every world or, in other words, zero necessarily exists. Since we are not presupposing a free logic, we have that every constant denotes something—the same thing—in every world.

To be sure, having numbers—any numbers—that exist of necessity is contrary to the spirit of potentialism. But perhaps the necessary existence of zero is at least relatively harmless: zero is the *only* number that exists necessarily.

We opt to avoid even this. Instead, we define a concept Z that, necessarily, holds of nothing. Following Frege, say that Z is defined by the formula $x \neq x$:

$$\Box(\forall x(\mathrm{Z}x \leftrightarrow x \neq x)).$$

Since we have comprehension, this concept constant is eliminable, but it is convenient to have it in the formal language. We add an axiom that it is necessary that zero could exist:

$$\Box\Diamond\exists x\#(\mathrm{Z}, x). \tag{Zero}$$

Following an informal Hume's Principle, we add an axiom stating that, necessarily, any number of Z is not the number of a plurality:

$$\Box\forall x\forall xx\neg(\#(Z, x) \land \#(xx, x). \tag{Zero Not Plural}$$

And we extend (Closure Down) to a statement that if any number exists, then zero does:[19]

$$\Box(\exists xx\exists y\#(xx, y) \rightarrow \exists x\#(\mathrm{Z}, x)). \tag{Closure Down+}$$

Let E be a concept constant defined by the formula "$x = x$". It is the concept of being self-identical. Like Z, this concept constant is eliminable. It is straightforward to see that E cannot have a number in any world that is actually finite. Suppose that a given world w has exactly n elements in its domain. If n exists in w then by (Closure Down) and (Closure Down+), every number smaller than n is also in w. There are thus $n + 1$ numbers in W, which is a contradiction. Moreover, the plurality of all objects in this world w also has no number.

Consider an interpretation in which every world is actually finite. Then E cannot have a number in any world in that interpretation. So this interpretation satisfies:

$$\Box\neg\exists x\#(\mathrm{E}, x),$$

or, in other words:

$$\Box\forall xx(\forall x(x \prec xx) \rightarrow \neg\exists x\#(xx, x)).$$

Recall our principle (Num Exists) stating that every plurality could have a number:

$$\Box\forall xx\Diamond\exists y\#(xx, y).$$

The concept-analogue of this would be:

$$\Box\forall X\Diamond\exists y\#(X, y).$$

As we have just seen, this is false at least in interpretations in which every world is actually finite. Our identity concept E is a counterexample to this principle in any such interpretation.

It might prove instructive to dwell on this. The problem that, unlike pluralities (as conceived here), at least some Fregean concepts are not rigid—and some are not stable. Let X be a concept, and focus on a particular world w. The above formula says that there is a world w', accessible from w that contains a number a number of X. What we would get, if the principle were correct, is the existence of a number of X *in* w', and *that* number may be different from the number of X in w. Let's apply this to our concept E.

[19] It is straightforward to modify (potential closure down) in an analogous way.

Suppose that a given world w has n members. The above principle would tell us that there is a world w' in which E has a number. But this would be the number of elements in w', not the number of elements of w. Again, there is no such number in w'.

We add the following, connecting the numbers of non-empty Fregean concepts (in a given world) to pluralities in that world, the number of the objects that the concept holds of in that world:

$$\Box\forall X\forall xx\forall y(\#(xx,y)\wedge(\forall z(Xz\leftrightarrow z\prec xx))\rightarrow\#(X,y))$$

In words, and invoking the heuristic, if xx consists of the objects that X holds of in a given world, and if the number of xx is y, then the number of X is y in that world. Of course, the concept X may not have that same number (or any number) in a different world.

We add a converse:

$$\forall X\forall y(\#(X,y)\rightarrow\forall xx(\forall z(Xz\leftrightarrow z\prec xx)\rightarrow\#(xx,y)))$$

This says that if a given Fregean concept has a number (in a world) and a given plurality is co-extensive with the concept (in that world) then the plurality has the same number as the concept.

Define equinumerosity on Fregean concepts as follows:

$$F\approx G\leftrightarrow_{\text{def}}\exists R(\forall x(Fx\rightarrow\exists!y(Rxy\wedge Gy))\wedge\forall y(Gy\rightarrow\exists!x(Rxy\wedge Fx))).$$

then (HP), plus the usual properties of identity, entails:

$$\forall X\forall Y\forall x\forall y((\#(X,x)\wedge\#(Y,y))\rightarrow(x=y\leftrightarrow X\approx Y)).$$

In words, if two Fregean concepts have the same number (in a given world) then they are equinumerous (in that world). This, of course, is a concept version of (HP). This and (HP') entails that if two concepts are equinumerous and one of them has a number, then the other one has the same number:

$$\forall X\forall Y\forall z(((X\approx Y\wedge\#(X,z))\rightarrow\forall w((\#(Y,w)))\leftrightarrow z=w)).$$

5 Frege's Theorem potentialized

5.1 The Dedekind-Peano axioms

A typical articulation of the (second-order) Dedekind-Peano axioms uses a language with a predicate N for being a natural number, a constant 0 for zero, and a successor function s. The axioms are that zero is a number, that the successor of a number is a number, that there is no number whose successor is zero, that the successor function is one to one on the numbers, and an axiom of induction—for any Fregean concept X, if X holds of zero and if X is closed under the successor function, then X holds of every number:

1. $\mathsf{N}(0)$
2. $\forall x(\mathsf{N}(x)\rightarrow\mathsf{N}(sx))$
3. $\neg\exists x(\mathsf{N}(x)\wedge sx=0)$
4. $\forall x\forall y((\mathsf{N}(x)\wedge\mathsf{N}(y)\wedge sx=sy)\rightarrow x=y)$
5. $\forall X((X0\wedge(\forall x((\mathsf{N}(x)\wedge Xx)\rightarrow Xsx)))\rightarrow(\forall x(\mathsf{N}(x)\rightarrow Xx)))$

Recall that, in the present modal potentialist setting, we avoid a free logic. Since we do not want any individual constants or any closed terms in the formal language, we follow Frege and employ a successor *relation*: $S(x,y)$ says that y is a successor of x. We use the following as our basic (non-modal) Dedekind-Peano axioms:

1. $\exists y(\mathsf{N}(y)\wedge\forall x\neg(\mathsf{N}(x)\wedge S(x,y)))$

2. $\forall x(\mathsf{N}(x) \rightarrow \exists y \forall z((\mathsf{N}(z) \wedge S(x,z)) \leftrightarrow y = z))$
3. $\forall x \forall y \forall z((\mathsf{N}(x) \wedge \mathsf{N}(y) \wedge \mathsf{N}(z) \wedge S(x,z) \wedge S(y,z)) \rightarrow x = y)$
4. $\forall X(\forall x((\mathsf{N}(x) \wedge \forall y \neg(\mathsf{N}(y) \wedge S(y,x)) \rightarrow Xx) \wedge (\forall x((\mathsf{N}(x) \wedge Xx \wedge S(x,y)) \rightarrow Xy))) \rightarrow (\forall x(\mathsf{N}(x) \rightarrow Xx)))$

The first axiom says that there is a number that is not a successor of anything; the second says that the successor relation is a function on numbers; the third says that this function is one to one; and the fourth is induction.

5.2 Definitions

Our next chore is to say what it is to be a natural number, in the potentialist setting, and to define the successor relation and some related items. Then we give potentialist versions of the Dedekind-Peano axioms, and establish those, along the lines of Frege's Theorem.

As noted, Frege defines a natural number to be an ancestor of zero under the successor relation. So, for Frege, a natural number is a finite cardinal number. Recall our axiom (Aristotle) designed to entail that, to invoke the heuristic, all worlds are Dedekind finite. There are no Dedekind infinite pluralities nor are there any Fregean concepts that apply to Dedekind infinitely many things (in any world). So here we take *all* (cardinal) numbers to be natural numbers. So define an object to be a *natural number* if it is the number of a Fregean concept:

$$\mathsf{NN}(x) \leftrightarrow_{\text{def}} \exists F \#(F, x)$$

It is straightforward to show that x is a natural number just in case either it is the number of the concept Z of being not self-identical or it is the number of a plurality:

$$\mathsf{NN}(x) \leftrightarrow (\#(\mathrm{Z}, x) \vee \exists xx \#(xx, x)).$$

Let O be the concept of being the number of a plurality of just one thing:

$$\mathrm{O}(x) \leftrightarrow_{\text{def}} \exists xx((\#(xx, x) \wedge \forall y \forall z((y \prec xx \wedge z \prec xx) \rightarrow y = z)))$$

By (HP), if $\mathrm{O}(x)$ and $\mathrm{O}(y)$, then $x = y$. In effect, O is the concept of being the number one.

Frege and the abstractionist neo-logicists define the number one to be the number of the concept of being identical to zero. This can be captured here. Recall our concept Z of being not self identical. The following follows:

$$\forall x \forall y \forall xx((\#(\mathrm{Z}, x) \wedge \forall z(z \prec xx \leftrightarrow z = x) \wedge \#(xx, y)) \rightarrow \mathrm{O}(y)).$$

In words, if y is the number of a plurality of only a number of Z, then $\mathrm{O}(y)$.

The next thing to be defined is the successor relation on numbers. Above, we gave a Fregean version of a successor relation, plus another one that is classically equivalent but perhaps more natural. Number n follows directly after number m just in case:

$$\exists F \exists G((m = \#F \wedge n = \#G) \wedge \exists x(Gx \wedge \forall y(Fy \leftrightarrow (y \neq x \wedge Gy)))) \qquad \text{(Frege-successor)}$$

$$\exists F \exists G\big(m = \#F \wedge n = \#G \wedge \exists x(\neg Fx \wedge \forall y(Gy \leftrightarrow (y = x \vee Fy))\big)) \qquad \text{(Successor)}$$

Those can be captured here. We use a version of the latter:

$$\mathsf{S}(m, n) \leftrightarrow_{\text{def}} \exists F \exists G(\#(F, m) \wedge \#(G, n) \wedge \exists x(\neg Fx \wedge \forall y(Gy \leftrightarrow (y = x \vee Fy))))$$

It can be shown that S is stable, since the concepts F and G can themselves be chosen to be stable, say as either the concept Z of being not self identical or the concept of being one of a given plurality. The following can be shown:

$$\mathsf{S}(m, n) \leftrightarrow (\#(\mathrm{Z}, m) \wedge \mathrm{O}(n)) \vee \exists xx \exists yy(\#(xx, m) \wedge \#(yy, n) \wedge \exists x(x \not\prec xx \wedge \forall y(y \prec yy \leftrightarrow (y = x \vee y \prec xx)))).$$

In words, n is a successor of m just in case either m is a number of the concept of being not self identical and n is a number of a plurality of just one thing, or m is a number of a plurality xx and n is a number of a plurality of the xx and one more thing.

It will prove convenient to define the inequality relations on natural numbers:

$$x \leq y \leftrightarrow_{\text{def}} \exists F \exists G(\#(F,x) \land \#(G,x) \land \forall z(Fx \to Gx))$$

$$x < y \leftrightarrow_{\text{def}} (x \leq y \land x \neq y)$$

Again, we can see that these relations are stable since one can choose stable Fregean concepts F and G, either our empty concept Z or the concept of being one of a given plurality.

5.3 Targets

Recall that the potentialist translation of a non-modal formula is the result of replacing each universal quantifier $\forall$ with $\Box\forall$ and each existential quantifier $\exists$ with $\Diamond\exists$, and translating the connectives homophonically.

Our first three targets are the necessitations of the potentialist translations of the first three Dedekind-Peano axioms:

1. $\Box\Diamond\exists y(\mathsf{NN}(y) \land \Box\forall x \neg(\mathsf{NN}(x) \land \mathsf{S}(x,y)))$
2. $\Box\forall x(\mathsf{NN}(x) \to \Diamond\exists y \Box\forall z((\mathsf{NN}(z) \land \mathsf{S}(x,z)) \leftrightarrow y = z))$
3. $\Box\forall x \Box\forall y \Box\forall z((\mathsf{NN}(x) \land \mathsf{NN}(y) \land \mathsf{NN}(z) \land \mathsf{S}(x,z) \land \mathsf{S}(y,z)) \to x = y)$

We establish those here, and turn to induction in the next subsection. The first axiom is straightforward:

Theorem 1: $\Box\Diamond\exists y(\mathsf{NN}(y) \land \Box\forall x \neg(\mathsf{NN}(x) \land \mathsf{S}(x,y)))$

Proof: Recall our axiom (Zero):

$$\Box\Diamond\exists x\#(\mathrm{Z},x),$$

and our definition of the "empty" concept Z:

$$\Box(\forall x(\mathrm{Z}x \leftrightarrow x \neq x)).$$

It follows immediately from the definition of S that no number of Z can be a successor of another number.

We turn next to the second axiom. We break it up into two parts, first showing that successors are unique:

Theorem 2 (Frege): $\Box\forall x\forall y_1\forall y_2((\mathsf{S}(x,y_1) \land \mathsf{S}(x,y_2)) \to y_1 = y_2)$

Proof: Suppose that $\mathsf{S}(x,y_1)$ and $\mathsf{S}(x,y_2)$ both hold (in a given world). Then

$$\exists G_1 \exists F_1(\#(G_1,x) \land \#(F_1,y_1) \land \exists w_1(\neg G_1 w_1 \land \forall w'(F_1 w' \leftrightarrow (w' = w_1 \lor G_1 w'))))$$

and

$$\exists G_2 \exists F_2(\#(G_2,x) \land \#(F_2,y_2) \land \exists w_2(\neg G_2 w_2 \land \forall w'(F_2 w' \leftrightarrow (w' = w_2 \lor G_2 w'))))$$

We have to show that $y_1 = y_2$. Suppose that F_1, G_1, w_1, F_2, G_2, and w_2 have the relevant features. Then we have $\#(G_1,x)$ and $\#(G_2,x)$. By (HP), $G_1 \approx G_2$. So there is a relation R such that $\forall u(G_1 u \to \exists! v(Ruv \land G_2 v))$ and $\forall v(G_2 v \to \exists! u(Ruv \land G_1 u))$. Using comprehension, let $R'uv$ hold if and only either Ruv or else $u = w_1$ and $v = w_2$. It follows that $F_1 \approx F_2$. By (HP), $y_1 = y_2$.

We turn next to the third axiom, that if two numbers have a common successor, then they are identical:

Theorem 3 (Frege): $\Box\forall x\Box\forall y\Box\forall z((\mathsf{NN}(x)\wedge\mathsf{NN}(y)\wedge\mathsf{NN}(z)\wedge\mathsf{S}(x,z)\wedge\mathsf{S}(y,z))\to x=y)$

Proof (sketch): Suppose that $\mathsf{S}(x,z)$ and $\mathsf{S}(y,z)$ both hold (in a given world). Then

$$\exists F_1\exists G_1(\#(F_1,x)\wedge\#(G_1,z)\wedge\exists w_1(\neg F_1w_1\wedge\forall w'(G_1w'\leftrightarrow(w'=w_1\vee F_1w'))))$$

and

$$\exists F_2\exists G_2(\#(F_2,y)\wedge\#(G_2,z)\wedge\exists w_2(\neg F_2w_2\wedge\forall w'(G_2w'\leftrightarrow(w'=w_2\vee F_2w'))))$$

We have to show that $x=y$. Suppose that F_1,G_1,w_1,F_2,G_2, and w_2 have the relevant features. Then we have $\#(G_1,z)$ and $\#(G_2,z)$. By (HP), $G_1\approx G_2$. Our theorem follows from what may be called the *Equinumerosity Lemma*:

Lemma: Suppose that $X\approx Y$, Xx, and Yy. Let X' be the Fregean concept defined by $\forall z(X'z\leftrightarrow(Xz\wedge z\neq x))$, and let Y' be the Fregean concept defined by $\forall z(Y'z\leftrightarrow(Yz\wedge z\neq y))$. Then $X'\approx Y'$.

For a proof, see https://plato.stanford.edu/entries/frege-theorem/proof5.htm.

Notice that in the proofs of Theorems 1-3, we have not invoked the axiom (Aristotle) that all worlds are Dedekind finite (so to speak). We do so now.

Theorem 4: No plurality is equinumerous with a proper sub-plurality:

$$\Box\forall xx\forall yy((\forall x(x\prec xx\to x\prec yy)\wedge\exists y(y\prec yy\wedge y\nprec xx))\to xx\not\approx yy).$$

Proof: Suppose that $\forall xx\forall yy(\forall x(x\prec xx\to x\prec yy)\wedge y\prec yy\wedge y\nprec xx)$. And suppose that $xx\approx yy$. Then there is a relation F that maps xx one-to-one onto yy, and a relation G that maps yy one-to-one onto xx. By first applying G and then F, we obtain a relation that maps yy one-to-one onto a proper subset of yy. This contradicts (Aristotle).

This result, together with (HP), sanctions (in this context) Euclid's Common Notion: a whole is greater than its (proper) parts.

Corollary: $\Box\forall n(\mathsf{NN}(n)\to\neg\mathsf{S}(n,n))$; no number is its own successor.

Proof: Suppose that $\mathsf{S}(x,x)$. That is,

$$\exists F\exists G(\#(F,n)\wedge\#(G,n)\wedge\exists x(\neg Fx\wedge\forall y(Gy\leftrightarrow(y=x\vee Fy))))$$

Let F, G, and x be given as in the formula:

$$\#(F,n)\wedge\#(G,n)\wedge\neg Fx\wedge\forall y(Gy\leftrightarrow(y=x\vee Fy)).$$

By (HP), $F\approx G$:

$$\exists R[\forall z(Fz\to\exists!y(Gw\wedge Rzw))\wedge\forall w(Gw\to\exists!z(Xz\wedge Rzw))]$$

But since $\forall x(Gx\to Fx)$, R is a one-to-one function on G which is not a surjection: x is not in its range. This contradicts (Aristotle).[20]

Next is a common result in the abstractionist program. It does not depend on the (Aristotle) axiom.

Theorem 5: (Frege) Suppose that 0 is a (or the) number of our "empty" concept Z (in a given world w). Consider the concept N^F of being an ancestor of 0 (in w) under the successor relation:

$$\forall x(\mathsf{N}^F(x)\leftrightarrow\mathsf{S}^*(0,x))$$

[20] Suppose we are working in a system like ours but without the (Aristotle) axiom. Let aa be Dedekind infinite, and assume $\#(aa,n)$. Then it is straightforward to show that $\mathsf{S}(n,n)$; n *is* its own successor.

Suppose $\mathsf{N}^F(n)$. Then n is either 0 or is the number of the plurality of all numbers (in w) that are less then n.

Proof: Let $A(x)$ be the Fregean concept of being either 0 or the number of numbers less than x. Of course $A(0)$. Suppose that $A(n)$ and that $\mathsf{S}(n, n')$, i.e., that n' is a successor of h and thus, from Theorem 2, n' is the successor of n. From the definition of the ancestral, we have to show that $A(n')$. We have that n is the number of numbers less than n, that $n \leq n'$, and, by the Corollary to Theorem 4, $n \neq n'$. By the definition of successor, n' is the number of numbers less than n'.

We now return to the other half of the second Dedekind-Peano axiom. We show that, necessarily, for every number n, it is possible that n has a successor:

Theorem 6: (Frege) $\Box\forall x(\mathsf{NN}(x) \to \Diamond\exists y\mathsf{S}(x, y))$

Proof: We have that for every number x, either x is a number of Z (the Fregean concept of being not self identical), or there is a plurality xx such that x is a number of xx.

Suppose that 0 is a (or by (HP), the) number of Z. Consider the plurality aa of just 0. Recall our axiom (Num Exists):

$$\Box\forall xx\Diamond\exists y\#(xx, y).$$

This entails that aa could have a number. It is straightforward that this number is a successor of 0 (namely the number one).

So now let n be the number of a plurality aa (so that $\#(aa, n)$). To invoke the heuristic, suppose that aa exists in a world w. We have to show that n could have a successor.

Case 1: There is an object c in w such that $c \not\prec aa$. Consider the plurality aa' consisting of the aa and c. By (Num Exists), aa' could have a number. So there is a world w' accessible from w and a number n' of aa. It is straightforward that n' is a successor of n.

Case 2: There is no object c in w such that $c \not\prec aa$. In other words, aa is the plurality of all objects in w. We show that this is impossible. Recall our axioms (Closure Down) and (Closure Down+), that if a number exists in a world then so do all smaller numbers. So w contains the number n of our "universal" plurality aa and all smaller numbers. By (Closure Down+), it contains 0, the number of our "empty" concept Z. Let aa be any plurality of objects in w. Then any number of aa is less than our equal to n, So that number is also in w. So w contains the number of every plurality of objects in w.

Let nn be the plurality of all ancestors of 0 in w. That is

$$\forall x(x \prec nn \leftrightarrow \mathsf{S}^*(0, x)).$$

Now if every nn had a successor in nn, then the successor relation S would be a one-to-one function on nn that is not a surjection (since 0 has no predecessor). This contradicts (Aristotle).

So there is a number m in nn that has no successor in nn. By Theorem 5, m is the number of all numbers less than m. But nn is a plurality of objects in w and so the number m' of nn is in the world w. But m' is the successor of m, and so m has a successor in w after all. This, of course, is a contradiction.

Theorem 7: $\Box\forall x((\mathsf{NN}(x) \wedge \neg\exists y\mathsf{S}(y, x)) \to \#(\mathrm{Z}, x))$. The only number that does not have a predecessor is zero.

Proof: Suppose not, that

$$\Diamond\exists x(\mathsf{NN}(x) \wedge \neg\exists y\mathsf{S}(y, x) \wedge \neg\#(\mathrm{Z}, x)).$$

Invoking the heuristic, there is a world w which contains a number a which is not a number of Z and which as no predecessor (in w). It follows that a is the number of a plurality (in w): $\exists xx\#(xx, a)$. Let aa be one such plurality (in w). So $\#(aa, a)$.

Recall our axioms (Closure Down) and (Closure Down+) that if a number exists (in a world) then so do all smaller numbers. Let A be the Fregean concept of being a number (in w) that is strictly smaller than a:

$$\forall z(Az \leftrightarrow (\#(\mathsf{Z}, z) \vee \exists zz(\#(zz, z) \wedge \forall y((y \prec zz \rightarrow y \prec aa) \wedge zz \not\approx aa))))$$

Lemma: A is closed under successor (in w): $\forall u(A(u) \rightarrow \exists v(\mathsf{S}(u, v) \wedge A(v)))$.
Proof: Suppose $A(e)$. So either e is a number of Z or e is the number of a plurality ee that maps one to one, but not onto aa.

Case 1: $\#(\mathsf{Z}, e)$, or, in words, e is a number of Z. Since pluralities are not "empty", let $c \prec aa$. Then consider the plurality cc of just c: $\forall x(x \prec cc \leftrightarrow x = c)$. Clearly, this plurality maps one to one into aa. Moreover, there is another element $d \neq c$ such that $d \prec aa$. Otherwise a would have a predecessor, namely e). So cc has a number (in w). That number is a successor of e.

Case 2: e is not a number of Z. Then there is a plurality ee such that $\#(ee, e)$ and ee is a sub-plurality of aa, but is not all of aa:

$$\forall(x \prec ee \rightarrow x \prec aa),$$

$$\exists x(x \prec aa \wedge x \not\prec ee).$$

So let $d \prec aa$ and $d \not\prec ee$. Then consider the plurality gg where $\forall x(x \prec gg \leftrightarrow (x \prec ee \vee x = d))$. Then, by (Closure Down) we have gg has a number g (in w) and so $\mathsf{S}(e, g)$. Moreover, gg is not all of aa. If it were, then we would have $\mathsf{S}(g, a)$, and so a would have a predecessor. So $A(g)$.

So we have that A holds only of numbers, and each such number has a successor that A holds of. And successors are unique. So, by (Aristotle), it follows that every member of A has a predecessor. But we also have that A holds of a number of Z, and we know that no such number can have a predecessor.

5.4 Induction

We turn to induction. Recall that Frege defines a natural number to be an ancestor of zero under the successor relation. There is what we may call a local version of induction:[21]

Theorem 8: $\Box\forall x(\mathsf{NN}(x) \leftrightarrow \exists z(\#(\mathsf{Z}, z) \wedge \mathsf{S}^*(z, x))$

In words, x is a number (in a world w) just in case x is an ancestor of a number of Z under S.

Proof: Suppose that our theorem is false. To invoke our heuristic, suppose that the theorem fails in world w. Let a be an object in w. Suppose $\exists z(\#((\mathsf{Z}, z) \wedge \mathsf{S}^*(z, a)))$. Then $\mathsf{NN}(x)$, since the relata of S are all numbers. So the right-to-left direction of the biconditional holds in w (trivially). So in w, there is a number 0 of Z and a number n which is not an ancestor of 0 under S. By our axioms (Closure Down) and (Closure Down+), every number less than or equal to n exists in w.

Let A be the Fregean concept of being a number less than our equal to n which is not an ancestor of 0 under S:

$$\forall x(A(x) \leftrightarrow (x \leq n \wedge \neg\mathsf{S}^*(0, n))).$$

Notice that $\neg A(0)$ since, of course, $\mathsf{S}^*(0, 0)$. So if $A(m)$ then $m \neq 0$. So, by Theorem 8, m has a predecessor m'. And we have $A(m')$, since $m' < m \leq n$ and m' is not an ancestor of 0 under S (since otherwise m would be).

Let nn be the plurality of all objects of which A holds. We have that the predecessor relation is a one-to-one function on nn. So, by (Aristotle) the predecessor relation is onto nn. So n has a *successor* n' such that $A(n')$. But this is absurd, since n' is not less than or equal to n.

[21] Thanks to Øystein Linnebo for suggesting this.

Recall our principle (finite) that states that, in effect, all worlds are finite:

$$\Box\exists R(\forall x\forall y_1\forall y_2((Rxy_1 \wedge Rxy_2) \rightarrow y_1 = y_2) \wedge \exists x\forall z R^*(x,z) \wedge \exists y\forall z(\neg R(y,z))) \qquad \text{(finite)}$$

We did not adopt (finite) as an axiom, opting for the ostensibly weaker (Aristotle), which only states that, in effect, all worlds are Dedekind finite.

We can now prove that all worlds are actually finite.

Corollary: (finite)

Proof: Suppose w is any world that is Dedekind finite but not finite, and let aa be the plurality of being a object in w. Let n be a number of aa, presumably in a different, accessible world. By Theorem 8, n is an ancestor of zero under s.

So we see that (Closure Down) is a powerful axiom, as is (potential closure down). If we have a number n of a Dedekind finite, but not finite Fregean concept A (or plurality aa), in a given world, then by (Closure Down) (or (potential closure down)), there is (or could be) a Dedekind infinite sequence of numbers below n, contradicting (Aristotle). In short, all worlds are actually finite.[22]

Recall the formulation of induction in a non-modal language that has no constants and successor is chararized as a relation:

$$\forall X(\forall x((\mathsf{N}(x) \wedge \forall y\neg(\mathsf{N}(y) \wedge S(x,y)) \rightarrow Xx) \wedge (\forall x\forall y((\mathsf{N}(x) \wedge Xx \wedge S(y,x)) \rightarrow Xy))) \rightarrow (\forall x(\mathsf{N}(x) \rightarrow Xx))).$$

We can simplify this a little, using present notation:

$$\forall X(((\forall x\#(\mathsf{Z},x) \rightarrow Xx) \wedge (\forall x\forall y((\mathsf{NN}(x) \wedge Xx \wedge \mathsf{S}(y,x)) \rightarrow Xy))) \rightarrow (\forall x(\mathsf{NN}(x) \rightarrow Xx))).$$

The potentialist translation of this is:

$$\Box\forall X(((\Box\forall x(\#(\mathsf{Z},x) \rightarrow Xx) \wedge (\Box\forall x\forall y((\mathsf{NN}(x) \wedge Xx \wedge \mathsf{S}(y,x)) \rightarrow Xy))) \rightarrow (\Box\forall x(\mathsf{NN}(x) \rightarrow Xx))). \quad (\text{Ind}^\Diamond)$$

In words, and using the heuristic of possible worlds, (Ind$^\Diamond$) says that IF in any world w_1, if w_1 contains a number of Z, then X holds of that number (in w_1), and IF for any world w_2, if X holds of a number n in w_2 and a successor of n exists in w_2, then X holds of that successor (in $w2$), THEN for any world w_3, X holds of all of the numbers in w_3. The consequent says that whenever a number is generated, X necessarily holds of it.

Theorem 9: (Ind$^\Diamond$)

Proof: Suppose not. So there is world w and a Fregean concept A such that necessarily, if there is number of Z then A holds of that number and, necessarily, if A holds of a number n in w and that number has a successor in w, then A holds of that successor in w, and there is a number m in w and A does not hold of m in w.

By (Closure Down) and (Closure Down+), w contains all numbers smaller than m. Define a Fregean concept B as follows:

$$\forall x(Bx \leftrightarrow (x \leq m \wedge \neg Ax))$$

Of course, we have Bm and if n is a number of Z (i.e., $\#(\mathsf{Z},n)$), then $\neg Bn$, since A necessarily holds of any number of Z. Suppose that Bq and that q is predecessor of r, so that $\mathsf{S}(q,r)$. Then Br, since otherwise we would have Aq. So every number that B applies to has a predecessor that B applies to. By (Aristotle), m must have a successor that B applies to. But, by definition, everything that has B is less than or equal to m. This is a contradiction.

[22] The usual proof, in set theory, that if a set is Dedekind finite then it is finite uses the axiom of choice. Various choice principles can be formulated in pure higher-order logic (see [21]). Our (Aristotle) and (finite) can be shown equivalent in a third-order language that has a sort for relations between pluralities and objects, and a sort for functions from pluralities to objects, plus the following choice principle:

$$\forall R(\forall xx\exists yR(xx,y) \rightarrow \exists F\forall xxRxx, F(xx)).$$

5.5 Mirroring?

Recall the potentialist translation of our non-modal language: replace all quantifiers of the form $\forall x$ and $\forall X$ with $\Box\forall x$ and $\Box\forall X$ respectively, and replace all quantifiers of the form $\exists x$ and $\exists X$ with $\Diamond\forall x$ and $\Diamond\forall X$ respectively. If φ is a formula in the non-modal language, let $\varphi^\Diamond$ be its translation.

Recall also the classical potentialist mirroring theorem (from [15] and [16]): Let $\vdash$ be the relation of classical deducibility in a non-modal *first-order* language $\mathcal{L}$. Let $\mathcal{L}^\Diamond$ be the corresponding modal language, and let $\vdash^\Diamond$ be deducibility in this language corresponding by $\vdash$, S4.2, and axioms asserting the stability of all atomic predicates of $\mathcal{L}$. Then for any formulas $\varphi_1, \ldots, \varphi_n, \psi$ of $\mathcal{L}$, we have:

$$\varphi_1, \ldots, \varphi_n \vdash \psi \quad \text{if and only if} \quad \varphi_1^\Diamond, \ldots, \varphi_n^\Diamond \vdash^\Diamond \psi^\Diamond.$$

The present non-modal language is, of course, second-order. For present purposes, we consider it to be a multi-sorted first-order language (with one sort for each kind of relation variable). There are two options for the target modal theory. One is to take the potentialist translations of the comprehension axioms as part of modal system $\vdash^\Diamond$, and the other is to include any use of comprehension as among the (non-logical) premises (the φ_i's) on the left hand side of the biconditional. Each of these follows in our modal theory (noting Lemma 5.3 of [15]).

Unfortunately, we cannot make direct use of the classical potentialist mirroring theorem here, for at least two reasons. First, the mirroring theorem requires all atomic formulas (in the modal language) to be stable, but our predication relation (between, say, monadic Fregean concepts and individual objects) is not stable. Consider the Fregean concept of being the largest number (in a given world):

$$Lx \leftrightarrow (\mathsf{NN}(x) \wedge \forall y(\mathsf{NN}(y) \rightarrow y \leq x))$$

Consider a world w that has a number n for Z, and no other numbers. Then Ln holds at w but not at any accessible world that contains more numbers. A second example is the concept of being the number of numbers (in a given world): $\#(\mathsf{NN}, x)$. A world with, say, exactly four numbers has access to a world with exactly five. A second reason why the mirroring theorem does not (directly) apply is that the present modal theory has axioms (such as (Aristotle)) that are not in the range of the potentialist translation.[23]

We can, however, make some use of the mirroring theorem. Let PA$^\Diamond$ be the modal theory consisting of S4.2, axioms stating the stability of all atomic predicates of the above Dedekind-Peano theory, including the stability of predication, as well as the potentialist translations of the instances of the comprehension scheme of the non-modal theory. Note that we do not have full comprehension in PA$^\Diamond$. Let $\vdash^\Diamond$ be deducibility in PA$^\Diamond$.

The mirroring theorem thus applies to PA$^\Diamond$, with respect to our non-modal Dedekind-Peano arithmetic: For any formulas $\varphi_1, \ldots, \varphi_n, \psi$ of the non-modal Dedekind-Peano arithmetic, we have:

$$\varphi_1, \ldots, \varphi_n \vdash \psi \text{ in the non-modal theory if and only if} \quad \varphi_1^\Diamond, \ldots, \varphi_n^\Diamond \vdash^\Diamond \psi^\Diamond.$$

It is possible to interpret PA$^\Diamond$ in our present modal (and plural) theory (which does have full modal and plural comprehension, along with axioms that are not in the range of the potentialist translation). First, if X is a concept variable, then let SX be an abbreviation of the statement that X is stable:

$$SX \leftrightarrow_{\text{def}} (\forall x(Xx \rightarrow \Box Xx) \wedge \forall x(\neg Xx \rightarrow \Box\neg Xx))$$

and similarly for relation variables. Let ϕ be any formula of our modal and plural theory, and let ϕ^S be the result of restricting all higher-order variables in ϕ to S. That is, replace each $\forall X\psi$ with $\forall X(SX \rightarrow \psi)$ and replace each $\exists X\psi$ with $\exists X(SX \wedge \psi)$. It is straightforward (but tedious) to verify that if

$$\varphi_1^\Diamond, \ldots, \varphi_n^\Diamond \vdash^\Diamond \psi^\Diamond,$$

then $\psi^{\Diamond S}$ can be derived from $\varphi_1^{\Diamond S}, \ldots, \varphi_n^{\Diamond S}$ in the present modal system.

We can do a lot better, however, since we only need one direction of the mirroring theorem.

[23] There also may be an issue with the fact that the modal language has plural terminology, but that does not seem to be problematic.

> **Theorem 10:** let $\vdash$ be deducibility in our original non-modal second-order theory of arithmetic, and let $\vdash^P$ be deducibility in our modal theory above. Then for any formulas $\varphi_1, \ldots, \varphi_n, \psi$ of the non-modal language:
>
> $$\text{if} \quad \varphi_1, \ldots, \varphi_n \vdash \psi \quad \text{then} \quad \varphi_1^\Diamond, \ldots, \varphi_n^\Diamond \vdash^P \psi^\Diamond.$$
>
> **Proof:** This is a straightforward (if tedious) induction on the length of a derivation in the non-modal theory of arithmetic (as in the proof of the mirroring theorem in [15] and [16]).

Recall that in our modal theory, we have the potentialist translations of the instances of comprehension as well as the potentialist translations of all of the Dedekind-Peano axioms. So we also have potentialist translations of every theorem of second-order Dedekind-Peano arithmetic. Putting aside the obvious anachronism, we thus have a verification of an arithmetic counterpart to one of Aristotle's claim's about actual infinity. To repeat:

> Our account does not rob the mathematicians of their study, by disproving the actual existence of the infinite in the direction of increase, in the sense of the untraversable. In point of fact they do not need the <actual> infinite . . . (207b27-30)

Of course, this applies to the mathematics of Aristotle's day, not the contemporary scene. Present focus is only on contemporary Dedekind-Peano arithmetic.

For the record, note that the following is a model for our modal theory: for each natural number n, there is a world whose domain is $\{m | m \leq n\}$, and accessibility is inclusion. So our modal theory is consistent if arithmetic is.

6 Doing without (Aristotle) and doing without (finite)

Recall our axiom (Aristotle) stating (in effect) that all worlds are Dedekind finite:

$$\Box\forall R\forall X[\forall x(((Xx \to \exists y\forall z(Xz \wedge Rxz) \leftrightarrow y = z)) \wedge \forall x_1\forall x_2\forall y((Rx_1y \wedge Rx_2y) \to x_1 = x_2)) \\ \to \forall y(Xy \to \exists x(Xx \wedge Rxy)))].$$

The plan here is to drop this. We do not want to *assert*, in the theory itself, the possible existence of, say, a plurality of all natural numbers, but we also don't want to rule out the possible existence of such a plurality. So we do not assert (here) the negation of (Aristotle), which would entail the existence of a world with a Dedekind infinite domain. Of course, we also do not wish to assert, here, the stronger principle (finite), that all worlds are actually finite, nor do we assert its negation.

Recall that, assuming (Aristotle) we defined a natural number to be the number of a Fregean concept or, equivalently, either a number of Z or the number of a plurality:

$$\mathsf{NN}(x) \leftrightarrow_{\text{def}} \exists F\#(F, x),$$

$$\mathsf{NN}(x) \leftrightarrow (\#(\mathsf{Z}, x) \vee \exists xx\#(xx, x))$$

We do not want to say that here, of course. Instead, we take the formula(s) just above to be a definition of *number*:

$$\mathsf{N}(x) \leftrightarrow_{\text{def}} \exists F\#(F, x),$$

or, equivalently:

$$\mathsf{N}(x) \leftrightarrow (\#(\mathsf{Z}, x) \vee \exists xx\#(xx, x)).$$

Now we need a definition of *natural* number, noting that there may be numbers that are not natural numbers. Recall that Frege's definition of the ancestral can be invoked in the potentialist context, provided it is formulated in terms of relations instead of functions, and does not invoke any singular terms. To deploy the heuristic, the definition works as advertised *within* each world. To remind the gentle reader, let R be a two place relation:

$$R^*xy \leftrightarrow_{\text{def}} \forall X[(Xx \wedge (\forall z \forall w((Xz \wedge Rzw) \rightarrow Xw))) \rightarrow Xy]$$

Recall also that Theorem 8 is what might be called a local version of induction:

$$\Box \forall x(\mathsf{NN}(x) \leftrightarrow \exists z(\#((\mathsf{Z}, z) \wedge \mathsf{S}^*(z, x))))$$

This assumed our prior definition of a natural number as the number of any plurality or Fregean concept, and the proof invoked our (Aristotle) axiom. Here we follow Frege and just take this formula to be a definition of natural number:

$$\mathsf{NN}(x) \leftrightarrow_{\text{def}} \exists z(\#((\mathsf{Z}, z) \wedge \mathsf{S}^*(z, x))) \qquad \text{(Frege)}$$

It is straightforward to verify that NN is stable, and that all natural numbers are numbers:

$$\Box \forall n(\mathsf{NN}(n) \rightarrow \mathsf{N}(x))$$

Recall our result that each ancestor of zero under successor is the number of all smaller numbers:

> **Theorem 5:** Suppose that 0 is a (or the) number of our "empty" concept Z (in a given world w). Consider the concept N^F of being an ancestor of 0 (in w) under the successor relation:
>
> $$\forall x(\mathsf{N}^F(x) \leftrightarrow \mathsf{S}^*(0, x))$$
>
> Suppose $\mathsf{N}^F(n)$. Then n is either 0 or is the number of the plurality of all numbers (in w) that are less then n.

Since this (or could have been) established without invoking (Aristotle), it holds here: every natural number is the number of all smaller numbers.

In the above treatment (assuming (Aristotle)), our first three targets were the necessitations of the potentialist translations of the first three Dedekind-Peano axioms:

1. $\Box \Diamond \exists x(\mathsf{NN}(x) \wedge \forall y \neg(\mathsf{NN}(y) \wedge \mathsf{S}(x, y)))$
2. $\Box \forall x(\mathsf{NN}(x) \rightarrow \Diamond \exists y \Box \forall z((\mathsf{NN}(z) \wedge \mathsf{S}(x, z)) \leftrightarrow y = z))$
3. $\Box \forall x \Box \forall y \Box \forall z((\mathsf{NN}(x) \wedge \mathsf{NN}(y) \wedge \mathsf{NN}(z) \wedge \mathsf{S}(z, x) \wedge \mathsf{S}(z, y)) \rightarrow x = y)$

In words, (1) there is a number that is not a successor of anything, (2) every number has a unique successor (i.e., the successor relation is a function), (3) successor is one to one.

In the previous section, Theorem 1 established the first of these, Theorem 2 established "half" of the second, that successors are one to one, and Theorem 3 established the third. The proofs of these did not invoke the (Aristotle) axiom, and they relied on the definition of a natural number as the number of a Fregean concept. So they hold here for *numbers*, and not just natural numbers. We have:

1. $\Box \Diamond \exists x(\mathsf{N}(x) \wedge \forall y \neg(\mathsf{N}(y) \wedge \mathsf{S}(x, y)))$
2. $\Box \forall x \forall y_1 \forall y_2((\mathsf{S}(x, y_1) \wedge \mathsf{S}(x, y_2)) \rightarrow y_1 = y_2)$
3. $\Box \forall x \Box \forall y \Box \forall z((\mathsf{N}(x) \wedge \mathsf{N}(y) \wedge \mathsf{N}(z) \wedge \mathsf{S}(z, x) \wedge \mathsf{S}(z, y)) \rightarrow x = y)$

A fortiori, since all natural numbers are numbers, these hold for natural numbers as well.

Without (Aristotle) and (finite), we would not expect Theorem 4 above, that no plurality is equinumerous with a proper sub-plurality to hold. Consider the Corollary to Theorem 4:

> $\Box \forall n(\mathsf{NN}(n) \rightarrow \neg \mathsf{S}(n, n))$; no number is its own successor.

This, too, relied on (Aristotle).

As indicated by the gloss, there are two statements to be pondered here. The first, taking the gloss literally (and out of its original context), is that no *number* is its own successor. Of course, we should not expect that to hold here, since we are not ruling out worlds with infinitely many members. As pointed out in note 20, if an interpretation has a Dedekind-infinite plurality aa in a given world, then the number of aa will be its own successor.

The other reading of Theorem 5 is that no *natural number* is its own successor. That follows from (Frege) and the definition of the ancestral:

Theorem 11: $\Box\forall x(\mathsf{NN}(x) \to \neg\mathsf{S}(x,x))$.

Proof: Let B be the Fregean concept which holds of something just in case it is a natural number that is not its own successor:

$$B(x) \leftrightarrow (\mathsf{NN}(n) \land \neg\mathsf{S}(n,n))$$

Let 0 be a (or, better, the) number of Z (the Fregean concept of being not self-identical). Since 0 has no predecessor, it is not a predecessor of itself, and so it is not a successor of itself. So $B(0)$. Now suppose that $B(x)$ and $\mathsf{S}(x,y)$. Then $\mathsf{NN}(y)$ and y has a predecessor, namely x. So $B(y)$. So we have $\mathsf{S}^*(a,y)$ So by (Frege) $\mathsf{NN}(y)$. Thus, B holds of all natural numbers.

Now we can easily establish the other "half" of the second Dedekind-Peano axiom, that every natural number could have a successor:

Theorem 12: (Frege) $\Box\forall x(\mathsf{NN}(x) \to \Diamond\exists y\mathsf{S}(x,y))$.

Proof: Suppose $\mathsf{NN}(n)$. By closure down, all numbers smaller then n exist. By Theorem 5, n is the number of the plurality of all numbers smaller than n. Let nn be the plurality of n and all numbers smaller than n. By (Num Exists) nn could have a number. This number is a successor of n.

The final axiom is induction. Recall that the potentialist translation of this is:

$$\Box\forall X(((\Box\forall x(\#(\mathsf{Z},x) \to Xx) \land (\Box\forall x\forall y((\mathsf{NN}(x) \land Xx \land \mathsf{S}(x,y)) \to Xy))) \to (\Box\forall x(\mathsf{NN}(x) \to Xx))). \quad (\text{Ind}^{\Diamond})$$

Here the proof of this is straightforward.

Theorem 13: (Ind$^{\Diamond}$)

Proof: Let X be a Fregean concept and suppose that, necessarily, X holds of every number of Z, and that, necessarily, X is closed under successors. Let w be a world. Then if w contains a number of Z, then X holds of that number, and we have that X is closed under successor in w. So, by (Frege), X holds of all numbers in w.

So, as with the previous theory based on (Aristotle), the present theory has the potentialist translations of all theorems of second-order Dedekind-Peano arithmetic.

Recall that most of the proofs in the previous section that relied on (Aristotle) used a reductio, sometimes called classical reductio or negation introduction. Here all of the relevant proofs are constructive. So if the background logic is intuitionistic, the theory proves all theorems of full second-order Heyting arithmetic (see [17]).

Finally, here is a model of the present theory, one that does satisfy the negation of (Aristotle) and thus the negation of (finite). There is one world that contains every natural number and also $\aleph_0$ (i.e., the number that Frege calls *endlos*), and for each natural number n, there is a world which contains all numbers less than or equal to n. Accessibility is inclusion.

7 Aristotelian set theory

The final project here is an Aristotelian set theory. Its intended interpretation is the hereditarily finite sets. As noted, Linnebo [15] and Linnebo [16] develop a potentialist set theory, based in a version of Frege's Basic Law V. The main idea is that, necessarily, for every plurality *aa*, there could be set whose members are the *aa*. There are axioms that make the theory equi-consistent with ZFC. The analogue of the axiom of infinity is an axiom stating that there is an transfinite stage or world, one that contains the set of all finite von Neumann ordinals, for example.

A natural attempt here would be a theory like Linnebo's, but with the aforementioned "infinity" axiom replaced with its negation, perhaps a set-theoretic analogue of our (Aristotle) axiom. Unfortunately, this will not do. It is "folklore" that ZFC^-—ZFC with the axiom of infinity replaced by its negation—is equivalent, in some sense, to Dedekind-Peano arithmetic. The fact is that the theories are mutually interpretable, but they are not definitionally equivalent. Moreover, Kaye and Wong [13] point out that ZFC^- does not prove induction for membership, nor does it prove that every set has a transitive closure. Indeed, ZFC^- does not prove that every set is a subset of a transitive set, a principle sometimes called "transitive containment". ZFC^- is a rather bizarre and unnatural theory.

Because of the Mirroring theorem, the same goes for a potentialist set theory like the envisioned one by Linnebo, but with the "infinity" axiom replaced by its negation. The plan here is to start with the non-modal "Small Set Theory" (SST) of McCarty, Shapiro, and Rathjen [forthcoming]. The only non-logical symbol is that for membership. There are four axioms:

1. **Extensionality:** $\forall x \forall y(\forall z(z \in x \leftrightarrow z \in y) \rightarrow x = y)$.

2. **Empty Set:** $\exists x \forall y . y \notin x$.

   We use "0" as a symbol for the empty set.

3. **Adjunction** $\forall x \forall y \exists z \forall u(u \in z \leftrightarrow (u \in x \vee u = y))$.

   Our unofficial (and eliminable) notation the adjunction of x and y is

   $$x \cup \{y\}.$$

   Note that, when writing $x \cup \{y\}$ here we do not presume thereby that an operation of binary union exists over the class of all sets. This is a theorem.

4. **Induction on Adjunction:** For any formula $\phi(x)$ in the language of set theory—featuring perhaps set parameters—if $\phi(0)$ and if

   $$\forall x \forall y((y \notin x \wedge \phi(x) \wedge \phi(y)) \rightarrow \phi(x \cup \{y\})),$$

   then $\forall x \phi(x)$.

The background logic for this theory is intuitionistic. All of the axioms of ZFC, except, of course, Infinity, follow from these axioms, as well as induction for membership, a theorem that every set has a transitive closure, and a theorem that every set is finite. SST is definitionally equivalent to Heyting arithmetic. If the background logic is classical, then the theory is definitionally equivalent to first-order Dedekind-Peano arithmetic.

Note that SST is first-order, with no variables or symbols for either pluralities or Fregean concepts. The plan is to develop a potentialist version $\mathsf{SST}^\Diamond$ of SST. We start with axioms assuring that sets are rigid, that they have the same members in all worlds. We add the potentialist translations of the four axioms of SST. Extensionality is:

$$\Box \forall x \forall y(\Box \forall z(z \in x \leftrightarrow z \in y) \rightarrow x = y).$$

Empty Set is

$$\Diamond \exists x \Box \forall y . y \notin x.$$

As with the Aristotelian arithmetic(s), we do not introduce any individual constants, since we do not wish to presuppose that any particular things exist. We introduce $\mathrm{E}(x)$ as an abbreviation of $\Box\forall y.y \notin x$. So our axiom is $\Diamond\exists x\mathrm{E}(x)$. Adjunction is

$$\Box\forall x\forall y\Diamond\exists z\Box\forall u(u \in z \leftrightarrow (u \in x \vee u = y)).$$

We introduce $\mathrm{A}(z, x, y)$ for z is an adjunction of x and y: $\Box\forall u(u \in z \leftrightarrow (u \in x \vee u = y))$.

Finally, induction on Adjunction: Necessarily, for any formula $\phi(x)$ in the language of the modal set theory featuring perhaps set parameters) if $\Box\forall x(\mathrm{E}(x) \to \phi(x))$ and if

$$\Box\forall x\forall y\forall z((y \notin x \wedge \phi(x) \wedge \phi(y) \wedge \mathrm{A}(z, x, y)) \to \phi(z))),$$

then $\Box\forall x\phi(x)$.

That's it. Since the language is first-order and membership is stable (indeed rigid), then the mirroring theorem entails that the potentialist translations of all of the axioms of ZFC, except, of course, Infinity, follow from these axioms, as well as induction for membership. We have that for every set x, there could be a transitive closure of x, and, crucially, it is necessary that every set is finite, and thus that every set is hereditarily finite. Our theory $\mathsf{SST}^\Diamond$ is definitionally equivalent to modalized Dedekind-Peano arithmetic[24]

ACKNOWLEDGMENTS. Thanks to Øystein Linnebo for reading previous versions of this paper and providing many helpful comments and suggestions. Thanks also to Neil Barton for insight into the philosophical and technical background. And to Tim Button for a most interesting exchange on key aspects of this paper. Thanks also to audiences in the C-FORS project at the University of Oslo for devoting several sessions to this project and providing helpful comments and suggestions. And I am indebted to two anonymous referees for helpful comments and suggestions.

References

[1] Aristotle, *The Basic Works of Aristotle*, R. McKeon, ed. Random House, New York, 1941.

[2] Button, Tim, "Level theory, Part 1: Axiomatizing the bare idea of a cumulative hierarchy of sets", *Bulletin of Symbolic Logic 27* (2021), 436–460.

[3] Button, Tim, "Level theory, Part 2: Axiomatizing the bare idea of a potential hierarchy", *Bulletin of Symbolic Logic 28* (2021), 1–29.

[4] Button, Tim, "Level theory, Part 3: a Boolean algebra of sets arranged in well-ordered levels", *Bulletin of Symbolic Logic 28* (2021), 1–29.

[5] Frege, Gottlob, *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle, Louis Nebert, 1879; translated as Begriffsschrift, a formula language, modeled upon that of arithmetic, for pure thought, in [24], 1-82.

[6] Frege, Gottlob, *Die Grundlagen der Arithmetik*, Breslau, Koebner, 1884; *The foundations of arithmetic*, translated by J. Austin, second edition, New York, Harper, 1960.

[7] Frege, Gottlob, *Grundgesetze der Arithmetik 1*, Olms, Hildescheim, 1893; translated as Gottlob Frege: *Basic laws of arithmetic*, by Philip A. Ebert and Marcus Rossberg, Oxford, Oxford University Press, 2013.

[8] Gauss, Karl Friedrich *Briefwechsel mit Schumacher*, Werke, Band 8, p. 216, 1831.

[9] Hale, Bob, *Abstract Objects*, Oxford, Basil Blackwell, 1983.

[10] Hale, Bob, and Crispin Wright, *The reason's proper study*, Oxford, Oxford University Press, 2001.

[24] A different potentialist theory of the hereditarily finite sets can be obtained from Button [2] [3] [4] by adding axioms to the effect that every set could have a singleton and that there cannot be an infinite set.

[11] Hodes, Harold [1990], "Where do the natural numbers come from?", *Synthese 84* (1990), 347-407.

[12] Hellman, Geoffrey and Stewart Shapiro, *Varieties of continua: from regions to points and back*, Oxford, Oxford University Press (2018).

[13] Kaye, Richard, and Tin Lok Wong, "On interpretations of arithmetic and set theory", *Notre Dame Journal of Formal Logic*. Volume 4 (2007), pp. 497-510.

[14] Lear, Jonathan "Aristotelian infinity", *Proceedings of the Aristotelian Society new series 80*, pp. 187-210, 1980.

[15] Linnebo, Øystein "The potential hierarchy of sets", *Review of Symbolic Logic 6* (2013), pp. 205-228.

[16] Linnebo, Øystein *Thin objects: an abstractionist account*, Oxford, Oxford University Press, 2018.

[17] Linnebo, Øystein and Stewart Shapiro, *Actual and potential infinity* Noûs 53 (2019), pp. 160-191, 2019.

[18] McCarty, Charles, Stewart Shapiro, and Michael Rathjen, "Intuitionistic sets and numbers: small set theory and Heyting Arithmetic", *Archive for Mathematical Logic*, forthcoming.

[19] Miller, Fred D. *Aristotle against the atomists*, in *Infinity and Continuity in Ancient and Medieval Thought*, edited by Normal Kretzmann, Cornell University Press, Ithaca NY, pp. 87-111, 2014.

[20] Shapiro, Stewart Foundations Without Foundationalism: *A Case for Second-Order Logic*, Oxford University Press, Oxford, 1991.

[21] Siskind, Benjamin, Paolo Mancosu, and Stewart Shapiro, "A note on choice principles in second-order logic", *Review of Symbolic Logic 16* (2023), 339-350.

[22] Sorabji, Richard *Time, Creation and the Continuum: Theories in Antiquity and the Early Middle Ages*, University of Chicago Press, Chicago, 2006.

[23] Stafford, Will, "The potential in Frege's Theorem", *Review of Symbolic Logic 19* (2023), 553–577.

[24] Van Heijenoort, Jean, *From Frege to Gödel*, Cambridge, Massachusetts, Harvard University Press, 1967.

[25] Walsh, Sean, "Comparing Peano arithmetic, Basic Law V, and Hume's Principle", *Annals of Pure and Applied Logic 163* (2012), 1679-1709.

[26] Wright, Crispin, *Frege's conception of numbers as objects*, Aberdeen, Aberdeen University Press, 1983.

Why Be a Height Potentialist?

Zeynep Soysal

Abstract

According to height potentialism, the height of the universe of sets is "potential" or "indefinitely extensible," and this is something that a (formal) theory of sets should capture. Height actualism is the rejection of height potentialism: the height of the universe of sets isn't potential or indefinitely extensible, and our standard non-modal theories of sets don't need to be supplemented with or reinterpreted in a modal language. In this paper, I examine and (mostly) criticize arguments for height potentialism. I first argue that arguments for height potentialism that appeal to its explanatory powers are unsuccessful. I then argue that the most promising argument for height potentialism involves the claim that height potentialism follows from our intuitive conception of sets. But, as I explain, on the most plausible way of developing this argument from an intuitive conception of sets, it turns out that whether height potentialism or height actualism is true is a verbal dispute, i.e., a matter of what meanings we choose to assign to our set-theoretic expressions. I explain that only pragmatic considerations can settle such a dispute and that these weigh in favor of actualism over potentialism. My discussion is also intended to serve two broader aims: to develop what I take to be the most promising line of argument for height potentialism, and to elaborate the height actualist position in greater detail than is standardly done.

1 Introduction

According to *height potentialism* (henceforth simply 'potentialism'), the height of the universe of (pure) sets is "potential" or "indefinitely extensible." This is commonly taken to imply that a (formal or informal) theory that accurately describes the universe of sets should include modal statements that capture its potential or indefinitely extensible nature.[1] For example, potentialists usually claim that a theory that accurately describes the universe of sets should include the statement that any things can form a set (usually formalized in plural logic).[2] As I will understand it in this paper, *height actualism* (henceforth 'actualism') is the rejection of potentialism: according to actualism, the height of the universe of sets isn't potential or indefinitely extensible, and our standard non-modal theories of sets (which usually include at least part of Zermelo–Fraenkel set theory with the Axiom

[1] But not always: e.g., **?** argues instead for a formalization of the extensibility of the universe of sets in higher-order logic.

[2] Those who accept or attribute that statement to potentialists include **?**, 145, **?**, 219, **?**, 699, **?**, 301, **?**, 1081, **?**, 12, **?**, 4. **?**, 82, 102 accepts similarly that we can always come to a more expansive understanding of our quantifiers, but doesn't formalize this statement in modal terms.

of Choice (henceforth 'ZFC')) don't need to be either supplemented with or reinterpreted in a modal language. Actualism, thus understood, is the default view that is at least implicit in the standard practice of set theory. Actualists can opt to go beyond rejecting potentialism and further characterize the universe of sets; one standard option for them is to characterize the universe of sets as a proper class, and another is to characterize it as a plurality that isn't singularized into an object (such as a set or a proper class).[3]

In this paper, I will examine and (mostly) criticize arguments for potentialism. My main contentions will be that the dispute between actualism and potentialism is best understood as verbal, that because of this, only pragmatic considerations can adjudicate between potentialism and actualism, and, finally, that such pragmatic considerations seem to favor actualism over potentialism. My discussion will also serve two broader aims, which are to develop what I take to be the most promising line of argument for potentialism and to elaborate the actualist position in greater detail than is standardly done.

There are two types of arguments standardly given for potentialism. The first, and most common type, are arguments that appeal to explanatory power. The idea is that potentialism is true because actualism leaves certain facts unexplained or arbitrary—such as the fact that there is no set of all sets, or that there can't be more sets than there actually are—whereas potentialism doesn't. I will argue that arguments from explanatory power for potentialism are unsuccessful. In particular, I will explain that either the potentialist's explanations of the relevant facts are no better than the actualist's, or there are only unexplained facts if potentialism is true to begin with.[4]

The second type of argument for potentialism appeals to what is part of our intuitive conception of sets. The idea is that our intuitive conception of sets has modal elements: for instance, as Cantor seems to have thought, there are multiplicities that are not "finished" or that "cannot exist together" (**?**), or, as Zermelo seems to have thought, the universe of sets is "open-ended" (**?**). The argument from an intuitive conception of sets, then, is that the universe of sets is potential because this is part of the intuitive conception of sets that we are working with. Since an accurate (formal) theory of sets is supposed to capture our intuitive conception of sets, it should thus include modal statements. As I will explain, this second type of argument for potentialism is much more promising: if our intuitive conception of sets really has modal elements, then whatever story one has for thinking that an intuitive conception of sets is true can also be used to justify these modal elements. I will argue that on the most plausible story for thinking that an intuitive conception of sets is true, the intuitive conception of sets is meaning-determining, that is, the intuitive conception of sets implicitly defines the meanings of the primitive set-theoretic expressions in it. But then, since there are also clearly non-modal intuitive conceptions of sets (such as the actualist conception),

[3] Those that opt for the former include **?**, 2, **?**, 23, **?**, **?**, 5; those that opt for the latter include **?**, **?**, or **?**. Of note is that potentialists can also accept the existence of proper classes understood as intensional entities; for discussion, see, e.g., (**?**).

[4] Some of these arguments expand upon my work in (**??**).

the argument from an intuitive conception of sets, if sound, will entail that whether potentialism or actualism is true is a verbal dispute, that is, a matter of what meanings we choose to assign to our set-theoretic expressions. I will explain that only pragmatic considerations can settle such a dispute. And I will suggest that these pragmatic considerations weigh in favor of actualism over potentialism.

I will discuss arguments from explanatory power first (§**??**) and then turn to arguments from intuitive conception (§**??**).

2 Arguments from explanatory power

It is very common for potentialists to motivate potentialism by claiming that it has some kind of explanatory superiority over actualism. More specifically, potentialists argue that their view can explain facts that actualism leaves unexplained or arbitrary. I will set aside the question of why (and whether) having explanatory superiority is evidence for the truth of a mathematical theory until §**??**; here in §**??**, I will examine the claim that potentialism has explanatory superiority over actualism.

Potentialists have described three different (though related) facts that are purportedly unexplained or arbitrary on the actualist view: the first is that the universe of sets (just as certain other collections of sets) doesn't actually form a set, the second is that the universe of sets (just as certain other collections of sets) doesn't possibly form a set, and the third is that the universe of sets (just as certain other collections of sets) isn't possibly "higher" than it actually is. I will discuss each in turn.

2.1 Unexplained actual non-existence

As is well-known, standard set theory precludes the existence of the sets of all and only the sets to which certain conditions apply, such as the condition of being non-self-membered, being an ordinal number, or being a cardinal number. Potentialists commonly remark that the actualist view leaves it unexplained why these sets don't exist; or, as James Studd puts it, it leaves unanswered the question: "*What is it about the world that allows some sets to form a set, whilst prohibiting others from doing the same?*" (**?**, 699). See, for instance, Øystein Linnebo:

> According to the actualist conception, the set-theoretic quantifers range over a definite totality of all sets. Why should the objects that make up this totality not themselves form a set? Since a set is completely specified by its elements, we can give a precise and complete specification of the set that these objects would form if they did form a set. What more could be needed for such a set to exist? (**?**, 206)[5]

[5] See also, e.g., "why does the hierarchy itself fail to be "completed" so as to constitute a set? [. . .] Why do all

Potentialists usually go on to present potentialism as filling this explanatory gap and thus as having an explanatory advantage over actualism.[6] The first argument from explanatory power for potentialism, then, is that actualism leaves it unexplained why there is no set of all non-self-membered sets (or ordinals, or cardinals, and so on), while potentialism doesn't.

Consider, first, the claim that actualism leaves these non-existence facts unexplained. An obvious reply for the actualist is to say that what explains that there is no set of all non-self-membered sets (or ordinals, or cardinals, and so on) is that it is contradictory to assume that it exists. After all, the standard textbook explanations for why there are no such sets involve mathematical arguments that show that their existence is contradictory. Take, for instance, the condition of being non-self-membered. The following reasoning, known as 'Russell's Paradox', shows that there is no set that contains all and only the non-self-membered sets: Assume that there is such a set. Call it '*r*'. Since *r* contains all and only the sets that are non-self-membered, *r* contains *r* if and only if *r* is non-self-membered. That is, *r* contains *r* if and only if *r* doesn't contain *r*. This is a contradiction. Similarly, the Burali–Forti Paradox and Cantor's Paradox yield arguments against the existence of the set of all and only the ordinals and cardinals, respectively. I will call the explanations given by these mathematical arguments 'textbook explanations'.

Potentialists commonly remark that these textbook explanations that are available to the actualists don't explain the non-existence of the relevant sets. See, for instance, James Studd:

> The derivation of Russell's paradox in Naïve Set Theory demonstrates the logical falsity of the instance of the Naïve Comprehension schema [. . .]. This provides—I am happy to grant—as good an explanation as we should expect for why this theory is inconsistent. However, the question of real interest is not why this instance of Naïve Comprehension yields a contradiction, but why certain sets—in this case, those that lack themselves as elements—are unable to form a set. And this cannot be explained merely by appeal to logical truths. (**?**, 700)

Or see Michael Dummett in this often-quoted passage:

> A mere prohibition leaves the matter a mystery. [. . .] And merely to say, "If you persist in talking about the number of all cardinal numbers, you will run into contradiction," is to wield the big stick, but not to offer an explanation. (**?**, 315f.)[7]

the urelements and sets that there actually are fail to constitute a further level that kicks off yet another series of iterations?" (**?**, 298).

[6] See, e.g., (**?**, 345f.), (**?**, 206), (**?**, 146, 152–155), or (**?**, 9f.).

[7] See also **?**, 149. Scambler also seems to think that there is something unexplained or paradoxical about the fact that certain collections don't (actually) form a set: "we are left with some things that, potentially 'paradoxically', *are not* and cannot in principle be collected together into a set" (**?**, 1081, *my emphasis*). Or see also Glanzberg: "But of course, to the extent that our notion of set is well-defined, *V* looks like a perfectly good set. We can insist it is not, but we lack a good explanation of why not" (**?**, 1493).

It is difficult to see what the problem with the textbook explanations is. It would be very implausible to deny that mathematical arguments (that is, proofs) in general can explain mathematical facts. It would be similarly implausible to deny that proofs by contradiction can explain mathematical facts; even constructivist mathematicians accept proofs by contradiction that establish the negation of some mathematical claim, including the negation of an existential claim. In a sense, the fact that the existence of a set of all non-self-membered sets is contradictory is the best explanation we could expect for why there is no such set; what more could we hope for as an explanation for why something doesn't exist than a proof that its existence is logically impossible?

As I will explain in §**??**, my view is that the textbook explanations are perfectly good explanations because they are *analytic explanations*, that is, they are arguments that show that the explanandum is entailed by *analytic truths*, or truths that follow from the implicit definitions of some relevant expressions. More specifically, my view is that the existence of the set of all sets (or ordinals, or cardinals, and so on) contradicts principles that follow from the implicit definitions of our set-theoretic expressions. As I see it, then, the textbook explanations only appear problematic if one doesn't realize why the principles that contradict the existence of the sets in question are true to begin with; in these cases, the textbook explanations appear to not be "deep" enough as explanations. Once one realizes that the principles in question are analytic, however, the textbook explanations become perfectly satisfactory.[8] As I will explain in §**??**, these (analytic) textbook explanations are available to both potentialists and actualists. But potentialists don't (currently) share my views about the textbook explanations. So, what are their complaints about the actualist's use of textbook explanations?

One can find an articulation of the potentialist's worry about the textbook explanations in (**?**, 56f.).[9] Linnebo seems to grant that proofs (by contradiction) can explain mathematical facts—even the fact that there is no set of all non-self-membered sets (or all ordinals, cardinals, and so on). However, according to him, such explanations only go through on an "intensional" specification of the sets in question, and not on an "extensional" one:

> Consider the collection of objects that are not elements of themselves. Can we form a set whose elements are the members of this collection? [. . .] [T]he answer depends on how the target collection is specified. Suppose the target is specified intensionally, say, by means of the condition $x \notin x$. Then the characterization of the desired set is logically incoherent. We want a set r such that $\forall x(x \in r \leftrightarrow x \notin x)$, but a simple theorem of first-order logic tell us that there is no such thing. Suppose instead that the target collection is specified in a purely extensional way, say, as a (perhaps infinite) list that includes each and every non-self-membered set in our domain. Then there is no logical or conceptual obstacle to the formation of a set whose elements are precisely the items

[8] I have also argued for this in (**??**).

[9] See also (**?**, 179).

> on this list. All that the paradoxical reasoning shows is that the set cannot itself be on the list and in this sense has to be "new". But this conclusion is unproblematic. There is no logical or conceptual obstacle to the formation of "new" objects beyond those that figure on some given list. (**?**, 56f.)

According to Linnebo, a domain can be "extensionally specificed" when it can be "specified as a plurality of objects"; and if a domain can only be specified "intensionally" then it "doesn't admit of a specification as a plurality" (**?**, 53). Linnebo further specifies what an "extensional specification" is by saying that a list of the members of a collection is an extensional specification of the collection, but the order of the list doesn't matter:

> The linear order of the list is immaterial. Suppose we wish to keep track of some web pages to which a new web page is intended to link. Suppose we write each of the target URLs on a separate piece of paper. It does not matter if the pieces are rearranged. Together, the pieces still provide a purely extensional specification of the target collection. (**?**, 57)

On this reading, then, the argument from explanatory power is that the actualist can't explain why the collection of all sets when it is specified extensionally doesn't form a set, while the potentialist can.

Consider once again the claim that the actualist leaves some non-existence fact unexplained. Whenever some collection of sets is "extensionally specified" via a list (either ordered or not), the actualist should say that this collection already (actually) forms a set. Indeed, any list of sets (certainly any finite list, but also any list of some cardinality) will already provably form a set for the actualist, for instance, via applications of the Axiom of Replacement.[10] Now, if lists (ordered or not) are the only way to specify a collection "extensionally," then the actualist should say that there is no extensional specification of the collection or domain of all sets on the actualist conception; the only way to specify or refer to the totality of all actually existing sets is by using expressions such as "all sets," or "the totality of all actually existing sets," or "everything to which my expression 'set' applies," and so on. The actualist can then invoke a textbook explanation to explain why these sets thus specified don't form a set. For instance, the actualist can say that no actually existing set in the totality of "all sets" is self-membered (a consequence of the Axiom of Foundation and the Axiom of Separation); therefore, there is no set that contains "all sets" in this totality of "all sets," since otherwise it would have to be self-membered. (This is my favorite version of the textbook explanation, which I will flesh out further in §**??**.) Linnebo should find this explanation satisfactory, given that he grants that textbook explanations are explanatory when they concern intensionally specified collections.

[10] Presumably, a specification of all sets using some formula $\phi(x_1, ..., x_n)$ of the language of set theory wouldn't count as "extensional" for Linnebo, but otherwise the actualist could say that again provably by an instance of the (first-order) Reflection Principle, any collection thus specified already forms a set.

Can a collection of sets be specified extensionally in a way other than via a list (ordered or not)? Linnebo seems to claim that, more generally, if one specifies a collection as a plurality, then one has specified it extensionally (**?**, 57).[11] If that is the case, then one option for the actualist would be to reply that the totality of all sets can be specified extensionally because it is a (mere) plurality; as I mentioned in §**??**, it is common for actualists who want to further specify the universe of sets to specify it as a plurality that doesn't form a set. The actualist can then run the same textbook explanation as above, but applied to the plurality of all sets: there is no set of the plurality of all sets because such a set would have to be self-membered, which contradicts the fact (which follows from the Axiom of Foundation and the Axiom of Separation) that no set in the plurality of all sets is self-membered. The actualist thus doesn't leave it unexplained why there is no set of all the actually existing sets—whether specified extensionally or intensionally. Moreover, her textbook explanation is, on the face of it, just as explanatory in either case (I will defend that they are explanatory further in §**??**).

Potentialists might take issue with this last textbook explanation, because they usually deny that the totality of all sets forms a plurality (at any world).[12] On the potentialist view, whenever there are some sets, there "can" be a set of these things. So, whenever there are some sets at a world, that isn't the "totality" of all sets, since there are still sets that can be formed from the ones in that world. For the potentialist, the "totality" of all sets in question is thus the totality of all actual and merely possible sets. But these potentialist claims don't give the actualist reason to reject that the collection of all sets is a mere plurality, for the actualist simply rejects the potentialist's claim that there are merely possible sets. These potentialist claims thus also don't raise any new explanatory problem for the actualist: The potentialist hasn't pointed to a fact that is unexplained on the actualist view by denying that the totality of all actual and possible sets forms a plurality. At best, the potentialist could object that the actualist leaves it unexplained why there aren't merely possible sets, which is the explanatory challenge I consider in §**??**.

To conclude my discussion of the first argument from explanatory power, let me turn to its second claim, namely, that potentialism can explain the actual non-existence claims purportedly left unexplained on the actualist view. Here, some care is needed to specify what exactly is the explanandum that the potentialist is concerned with; as Sam **?**, fn. 20 has also pointed out, potentialists seem to subtly shift the question at issue. Recall that the actualist is accused of leaving unexplained why there is no set of all sets on her actualist picture. For the actualist, there are no merely possible sets, so the "universe" or "totality" of all sets refers to all the actually existing

[11] More specifically, (**?**, 57) says that a "simple but useful way to talk about extensionally specified collections is provided by the logic of plurals." Arguably, this doesn't give an account of when a collection can be specified extensionally, but only entails that we can refer to extensionally specified collections using plural language. But I consider this option here because it seems plausible that there should be an alternative to the account of extensionally specified collections as lists (ordered or unordered), and I take the account that says that extensionally specified collections are pluralities to be a plausible alternative that is also in the spirit of the text.

[12] See, e.g., (**?**, 156–158), (**?**, 219f.).

sets. The corresponding question that the potentialist should be answering, then, is: Why is there actually no set of all actually existing set on the potentialist picture?[13] For the potentialist, the set of all actually existing sets isn't actual but merely possible. So, to understand why, on the potentialist view, such a set is merely possible and not actual, we need to get a fix on the notion of "possibility" that the potentialist employs.

Potentialists have proposed different interpretations of their modality: one on which the modality is primitive and idiosyncratic to mathematics or set theory;[14] a second, "constructivist" interpretation on which sets don't exist timelessly and metaphysically necessarily, but rather come into existence via some kind of social process;[15] and a third, broadly "linguistic" interpretation on which the extensions of our expressions become more inclusive.[16]

It is clear that the potentialist doesn't have a satisfactory explanation on the first interpretive option: saying that the universe of sets doesn't actually form a set because it merely possibly forms one on a primitive sense of 'possibly' isn't an illuminating explanation, and certainly not for those who reject potentialism to begin with. But this point will become clearer in §**??** where I explain the most plausible way of making sense of a primitive notion of modality for the potentialist, so I won't say more about it here.

On the constructivist interpretation, the existence of (pure) sets is metaphysically contingent; some type of social process brings about their existence at a point in time. On this interpretation, then, the potentialist's explanation for why there is no set of all actually existing sets is that this set hasn't yet been constructed or brought into existence by the relevant social process. The problem with this explanation is that it is extremely controversial to deny that abstract mathematical objects exist timelessly and necessarily (if they exist); even potentialists usually don't want to deny this.[17] It is arguably part of our intuitive notion of metaphysical possibility and of abstract mathematical objects that they exist timelessly and necessarily. Denying this is, at the very least, an important intuitive cost. Moreover, the view that sets are constructed by a social process in time is standardly taken to entail a weaker mathematical theory than ZFC.[18] For these reasons, and for anyone who endorses the standard view that mathematical objects exist timelessly and necessarily if they exist—and, certainly, for the actualist—the potentialist explanation on the constructivist interpretation will be highly unsatisfactory.

On the linguistic interpretation, the potentialist points out that we can (in the sense that it is physically possible for us to) change the extensions of our set-theoretic expressions (perhaps

[13] **?**, fn. 20 calls this 'the problem of actuality', and argues that potentialists don't have a clear response to it. This problem also seems implicit in Berry's (**?**, ch. 5) critical discussion of Parsonian potentialism and is related to a criticism of potentialism raised by **?**, 303f.. As I explain next, I think that the potentialist does have an explanation, but it is a bad one.

[14] Those who propose an idiosyncratic modality include **?** and **?**.

[15] E.g., **?**, 285 and **?**, 60 discuss that interpretation of the potentialist's modality (though they don't endorse it).

[16] See, e.g., **?**, **?**, **?**, **?**, **?**.

[17] See, e.g., (**?**), (**?**, 158), (**?**, 706), (**?**, 167), (**?**, 288), (**?**, 146), (**?**, 113).

[18] For discussion, see, e.g., (**?**, 285). For development of the argument, see, e.g., (**?**).

together with the range of the quantifiers) so that they apply to more things than they did before the change. As Timothy Williamson puts it:

> For given any reasonable assignment of meaning to the word 'set' we can assign it a more inclusive meaning while feeling that we are going on in the same way. (**?**, 20)

Similarly, **?** explains that the modal operators "describe how the interpretation of the language can be shifted—and the domain expanded—as a result of abstraction," where "the intended meaning of '$\Box\varphi$' is 'no matter how we abstract and thereby shift the meaning of the language, φ" (**?**, 205). On this third interpretation, then, the potentialist's explanation for why there is no set of all actually existing sets seems to be that this plurality isn't a set because we haven't yet shifted the meaning of our set-theoretic expressions (for instance, via abstraction) so as to call this mere plurality a 'set'. The problem with this explanation is that it doesn't address why there is no set of all actually existing sets on the actual usage of 'set'. The actualist can certainly grant the potentialist that we can change the meaning of 'set' so that it also applies to what the actualist used to call 'proper classes' or 'mere pluralities'. But this doesn't address why those things the actualist calls 'proper classes' or 'mere pluralities' in the current, actual language don't form sets on the current, actual meaning of 'set'—which is precisely the why-question the potentialist was supposed to be answering. Like the actualist, the potentialist could respond that the reason why such collections aren't sets on the current, actual meaning of 'set' is given by some textbook explanation: their being a set would be contradictory, or perhaps logically incompatible with some basic facts about the application conditions of 'set', such as the fact that 'set' doesn't apply to anything that is self-membered (as I will further explain in §**??**). But the linguistic interpretation doesn't provide the potentialist with any alternative explanation for why there is no set of all actually existing sets—holding fixed the current meaning of 'set'.

I don't see how else potentialists can explain the actual non-existence claims.[19] Some potentialists are explicit that potentialism doesn't explain why there is no set of all non-self-membered sets (or of all actually existing sets, ordinals, cardinals, and so on), and even seem to grant that textbook explanations are satisfactory after all:

> [A]lthough the Russell reasoning shows there necessarily are some things that *aren't* the members of a set, nevertheless any possible things *can be* the members of a set: there are no 'special' things (plural) that somehow by their very nature cannot be the members of a set. (**?**, 1081)

[19] Perhaps potentialists could say that there is no set of all non-self-membered sets because the Axiom of Foundation is true (that is, every nonempty set is disjoint from one of its elements) and that they can explain why the Axiom of Foundation is true better than competitors; **?**, 216f., e.g., gives an argument for the Axiom of Foundation from a potentialist perspective. The problem with this argument is that there is no reason to think the potentialist's explanation of the Axiom of Foundation is any better than the actualist's, since the actualist can appeal to the iterative conception of sets, and it is well-known that the latter motivates the Axiom of Foundation perfectly well (the Axiom of Foundation is even considered essential to the iterative conception of sets). See also discussion in §**??**.

> Recall that RP [(Russell's Paradox)] shows that, necessarily, the plurality of all sets does not *actually* form a set. This leaves open, however, whether that plurality *could* form a set. [. . .] In response to RP, the modal strategy maintains that *within* any given set-theoretic structure certain pluralities will not form sets, but every such plurality *could* form a set in some larger set-theoretic structure. (**?**, 12f.)[20]

When it comes to explaining the actual non-existence claims, then, the potentialist doesn't have an advantage over the actualist. At best, if the potentialist avails herself of some textbook explanation (as I argue in §**??** that she should), then the actualist and the potentialist explanations are on a par.

As I mentioned above, at times it seems that the potentialist is concerned with a subtly different explanandum, namely, why the universe of sets *understood in the potentialist's sense* doesn't actually (or possibly) form a set. Charles Parsons, for instance, motivates the explanatory advantage of potentialism by explaining that when some multiplicity doesn't determine a set, "this is due to the fact that in a certain sense the multiplicity does not exist" (**?**, 345):

> I suggest interpreting Cantor by means of a modal language with quantifiers, where within a modal operator a quantifier always ranges over a set [. . .]. Then it is not possible that all elements of, say, Russell's class exist, although for any element, it is possible that *it* exists. (**?**, 346)

Linnebo motivates potentialism's explanatory advantage similarly in (**?**):

> On [the potentialist's] [. . .] conception, the hierarchy is potential in character and thus intrinsically different from sets, each of which is completed and thus actual rather than potential. This intrinsic difference affords potentialists [. . .] a reason to disallow the disputed set formation. (**?**, 206)

On one natural readings, Parsons and Linnebo here aren't concerned with explaining why there is no set of all actually existing sets, but, rather, with explaining why the totality of sets doesn't (and couldn't) form a set, where the "totality" of sets is understood as consisting of all actual and merely possible sets. The latter explanation, which we discussed earlier in §**??**, is that whenever there are some sets at a world, that isn't "the totality of sets," because there are still sets that can be formed from the ones in that world; since the totality of sets doesn't exist at any given world, it also can't determine a set at any given world.

[20] Menzel says something similar: 'Note the question is not: Why is there no universal set, that is, no set containing all the urelements and all the sets? As we've just seen, the iterative conception of set provides a cogent answer to that question: only those pluralities that "run out" by some level of the cumulative hierarchy constitute sets at the next level and, obviously, the entire hierarchy is not such a plurality; there is no level at which the members of all the levels form a set' (**?**, 298). After the quote I gave above, Studd says: "The derivation of a contradiction from the relevant instance of Naïve Comprehension shows that these sets do not form a set, but fails to explain why they *cannot*" (**?**, 700), which suggests that it might be better to interpret him also as granting that Russell's Paradox explains why there is no set of all sets, but as failing to explain why there couldn't be a set of all sets, as I discuss in §**??**.

As I already explained, on this reading, the potentialist isn't raising an explanatory problem for the actualist. The actualist can explain why the universe of sets isn't actually a set, where "the universe of sets" is understood as the totality of all actually existing sets. Given that she doesn't recognize the need to introduce modal notions into set theory, she doesn't need to explain why "the universe of sets" understood as consisting of actual and merely possible sets doesn't form a set. In other words, the fact that the actualist doesn't explain why there is no set of all actual and possible sets doesn't show that there is something unexplained or arbitrary *on the actualist view* of the universe of sets, and thus it doesn't show that potentialism has an explanatory advantage. To take a somewhat facetious analogy, this would be like arguing for the existence of God by saying that it best explains why angels can transcend physical boundaries; an atheist isn't going to accept that there are angels to begin with. That being said, and again as I mentioned earlier, the potentialist could object that the actualist leaves it unexplained why there are no merely possible sets (why are there no angels?). This is the explanatory challenge to which I turn next.

2.2 Unexplained impossible existence

On the face of it, actualists reject that there can be more sets than there actually are. On a second reading, the argument from explanatory power for potentialism is that this is an unexplained fact on the actualist view: actualism leaves it unexplained why the universe of sets (as well as some other collections) *can't* form a set. This argument is implicit in the penultimate set of quotes from §**??** that grant that standard mathematical arguments can explain why the problematic collections don't actually form a set, but point out that these arguments don't explain why these collections *couldn't* form sets. See, also, for instance:

> [W]ith the combinatorial/iterative conception in mind, why *can't* we "collect together" or "lasso" all the sets in the ZF hierarchy, and form the collection of them all? (**?**, 111, *my emphasis*)
>
> [T]he Actualist takes there to be some plurality of objects (the sets) forming an iterative hierarchy structure [. . .]. But the following modal intuition seems appealing: for any plurality of objects satisfying the conception of an iterative hierarchy [. . .] *it would be in some sense (e.g., conceptually, logically or combinatorically if not metaphysically) possible* for there to be a strictly larger model of [a full-width iterative hierarchy] [. . .] which, in effect, adds a new stage above all the ordinals within the original structure together with a corresponding layer of classes. (**?**, 15, *my emphasis*)[21]

On the potentialist view, in contrast, these collections and pluralities *can* form sets; on the fact it is, then, the potentialist view doesn't have unexplained facts about impossible set existence.

[21] See also, e.g., Uzquiano who claims that we need "some independent reason to doubt that we *can* collect the non-self-membered sets in the totality into a set" (**?**, 147, *my emphasis*).

The key question for this second reading of the potentialist argument from explanatory power is: Does the actualist need to recognize that there are such modal facts in need of explanation? Once again, for the actualist, the universe of sets doesn't have a potential or modal nature. She follows the standard mathematical practice of interpreting 'can'-statements in terms of (actual) existence.[22] So, for instance, the actualist will say that on her view, there "can't" be a set of all sets in the sense that there *isn't* one, and there isn't one because the assumption that there is one entails a contradiction.

What the potentialist needs is to point to certain modal facts that the actualist needs to recognize, and ask why these facts don't hold on the actualist picture (if they don't). To do this, the potentialist needs to give an interpretation of her modality—and thus of the sense in which sets "can" or "can't" form sets—that is comprehensible to the actualist. Consider again the three standard potentialist interpretations of their modality.

It should seem clear that the actualist doesn't need to accept or explain that their universe of sets "can't" form a set in a primitive sense proper to the potentialist's explication of their potential hierarchy. But this point will become clearer after my discussion in §**??**, so I won't say more about it for now.

Recall that on the constructive interpretation, there "can" be more sets than there actually are in the sense that it is metaphysically possible for these (actually non-existing) sets to exist, as a result of some kind of social process of set construction. The challenge here, then, is that the actualist leaves it unexplained why it isn't metaphysically possible for there to be more sets than there actually are as a result of some process of social construction. The actualist can accept that there can't be more sets than there actually are in this sense, and she can give the standard explanation of that fact: there can't be more sets than there actually are in the constructivist sense because abstract mathematical objects exist timelessly and necessarily (if they exist), as is almost universally accepted (as we discussed in §**??**). On this second interpretation, there is as little unexplained on the actualist view as there is on any view which entails that abstract mathematical objects exist timelessly and necessarily, if they exist.

On the linguistic interpretation, the potentialist's challenge to actualism is that the actualist view leaves it unexplained why we (physically) can't assign more inclusive meanings or extensions to our set-theoretic expressions or quantifiers. The problem with this challenge is that the actualist can simply accept that we can assign more inclusive meanings and extensions to our set-theoretic expressions and quantifiers; she can even give the metasemantic explanation of that fact as the potentialists do.[23] So, it isn't an unexplained fact (or even a fact) on the actualist view that we can't

[22] **?**, 282 discuss (and concede) this standard practice to interpret 'can'-statements in mathematics non-modally. An alternative interpretation of 'can' is as epistemic possibility, i.e., there can't be a set of all non-self-membered sets in the sense that we know that there isn't one (because, e.g., we have a proof).

[23] As will become clear in §**??**, actualists and potentialists actually accept a similar metasemantic theory; they just disagree on the actual use of the set-theoretic expressions and quantifiers (see footnote **??**).

assign more inclusive meanings or extensions to our set-theoretic expressions and quantifiers. The actualist can accept that we can assign more inclusive meanings, but she will just refuse to do so: For the actualist, 'set' refers to all and only the things that appear at some stage of the iterative hierarchy (which is given by the familiar intuitive story, perhaps together with the axioms of ZFC). If she wants, the actualist can further say that the "totality" of all sets, called 'V', is a proper class or a mere plurality. Either way, V isn't a set, on pain of contradiction. The actualist can grant that one could decide to use 'set' in a new and different way so as to call everything the actualist calls 'sets' *and* V a 'set'. But that term 'set' has a different extension (and thus also a different meaning), than the actualist's word 'set'; and this is something that the potentialist accepts. So, the actualist can agree that 'V can be a set' is true in the sense that, if we were to change the meaning of 'set' to refer to both sets and to V, then V would be a set. But the actualist doesn't want to change the meanings of her expression 'set'. For her, it would be best to disambiguate and use, for instance, 'shmet', to refer to the more inclusive extension. Then, when the potentialist asks: "But why can't the plurality you call 'V' form a set?" the actualist will say that V is a shmet, proper class, or mere plurality, but that it isn't a set given what she means by 'set'. Shmets, proper classes, and mere pluralities are similar to sets in some respects, but they aren't sets. And whoever adopts the more inclusive meaning of 'set' herself agrees with this: she agrees that there are distinct meanings to 'set' before and after the shift in meaning, and so that "new sets" and "old sets" aren't the same things.[24] The potentialist just insists on using 'set' with a new meaning instead of using 'shmet' or 'proper class' or 'mere plurality'. On this third interpretation, then, is no fact—let along an unexplained fact—that V can't form a set on the actualist picture; there are just different decisions about how to use set-theoretic expressions. In §**??**, I will explain that there are pragmatic reasons to prefer the actualist's use of 'set'. My point here is simply that there is no unexplained fact that sets can't form more sets on the actualist picture.[25]

2.3 Arbitrary height

Finally, potentialists also commonly object that the actualist is committed to a theory on which the "height" of the universe of sets is arbitrary. See, for instance:

> Why is the hierarchy only as "high" as it is? Why do all the urelements and sets that there actually are fail to constitute a further level that kicks off yet another series of iterations? (**?**, 298)
>
> The main challenge will be to motivate and defend the threshold cardinality beginning

[24] See, e.g., the way **?**, 55–59 explains this view.

[25] **?**, 214–216 and **?**, 711 argue that the potentialist can also explain which collections or pluralities can form sets and which cannot. From my arguments in this section (§**??**), it follows that, depending on the interpretation of the modality at issue, the actualist should either reject that these collections or pluralities can form sets or give the same explanation as the potentialist for why these (and only these) can form sets.

> at which pluralities are too large to form sets. Why should this particular cardinality mark the threshold? Why not some other cardinality? (**?**, 152)[26]

There is a reading of these quotes on which the challenge for the actualist is to explain why it isn't *possible* for there to be more sets, leading back to the argument discussed in §**??**.[27] I take the challenge that is more specifically about the "height" of the universe of sets to go as follows. The actualist (just like the potentialist) is committed to the universe of sets being in some sense "as high as possible": this idea is standardly included as part of the intuitive iterative conception of sets—for instance, in the Cantorian idea of "absolute infinity" (**?**)—and in set theorists' practice of accepting the existence of larger and larger cardinals.[28] Standard set theory entails that x isn't a set if it can be put in one-to-one correspondence with V or the ordinals Ω.[29] But this suggests that there is some "threshold height"—the "height" of V or Ω—beyond which collections don't form sets. According to the potentialist, this doesn't fit with the intuitive idea that the universe of sets is as high as possible; see, for instance, Linnebo in the passage following the above:

> Wherever it has been possible to go on to define larger sets, set theorists have in fact done so. So it remains arbitrary that there should be no sets of this [threshold] cardinality or some even larger one. (**?**, 153)

The third explanatory challenge for the actualist, then, is to explain why the universe of sets has this threshold height given that it should be as high as possible.

As I have previously argued (**??**), I think the actualist can straightforwardly maintain both that the collection of all sets (or of all ordinals, or all cardinals) has a threshold "height" and that it is as "high" as possible.[30] Consider, first, that there is a specific notion of "height" (or "size") given in set theory by the notion of an ordinal (or cardinal): In set theory, ordinals (and cardinals) are defined as *sets* of a certain type.[31] On the standard definition, an ordinal β is *larger* than an ordinal α just in case $\alpha \in \beta$. With these specific explications of the informal notion of "height," "size," and "larger than," the actualist can then note that standard set theory entails that for every ordinal (cardinal) there is a larger one. Furthermore, the actualist can adopt the set-theoretic practice of accepting the existence of stronger and stronger theories that postulate the existence of larger and

[26] See also, e.g., **?**, 152f., 155, **?**, 23, **?**, 152f., **?**, 206, **?**, 700, **?**, 14–17, **?**, 6f..

[27] For instance, this is a natural reading of the way **?**, 14–17 puts the challenge about height: " [T]he Actualist takes there to be some plurality of objects (the sets) forming an iterative hierarchy structure [. . .]. But the following modal intuition seems appealing: for any plurality of objects satisfying the conception of an iterative hierarchy [. . .] it would be in some sense (e.g., conceptually, logically or combinatorically if not metaphysically) possible for there to be a strictly larger model of [a full-width iterative hierarchy]" **?**, 15.

[28] See, e.g., (**?**) for discussion of the role of such "maximality principles" in set theory.

[29] This is the Limitation of Size Principle, usually proved in a set theory with classes such as Neumann–Bernays–Gödel set theory (NBG). **?**, 162 also derives it in set theory with pluralities.

[30] **?** also briefly makes similar points against the potentialist's arbitrary height argument.

[31] Specifically, ordinals are sets that are transitive and well-ordered by $\in$, and cardinals are ordinals that aren't in one-to-one correspondence with any smaller ordinal.

larger cardinals in this sense.[32] For the actualist, these facts suffice to capture the intuitive idea that the universe of sets is as "high" as possible.

The actualist will then think of the "height" or "size" of V differently. For one, the "height" or "size" of V isn't an ordinal or a cardinal (on pain of contradiction). The actualist could define a notion of "size" that applies to V; for instance, by saying that two classes have the same *shardinality* if and only if they can be put in one-to-one correspondence. Shardinality and cardinality are then both "sizes" in an informal sense, but they aren't the same type of "size," on pain of contradiction. For the actualist, not every type of "size" is a cardinality, again on pain of contradiction.

These claims nicely parallel what the actualist says about "collections," as we saw at the end of §**??**: The universe V isn't a set (on pain of contradiction). The actualist can define a notion of "collection" that applies to it, for instance, by calling it a 'proper class' and defining these in some standard way. Then proper classes and sets are both "collections" in a loose sense, but they aren't the same type of "collection," on pain of contradiction. For the actualist, not every type of "collection" is a set, again on pain of contradiction.[33]

The actualist thus maintains both that the universe of sets is a "high" as possible (that is, more specifically, there is no largest cardinal/ordinal) and that there is a threshold "height" (that is, more specifically, a shardinal, or shordinal) above which "collections" (that is, more specifically, proper classes and sets, or mere pluralities and sets) don't form sets. This threshold isn't arbitrary: it is the first "height" at which the assumption that x has that height entails that x isn't a set. If she wants, the actualist could also accept that just as there are larger and larger ordinals and cardinals there are shlarger and shlarger shardinals and shordinals, by accepting the existence of collections that are neither sets nor proper classes, such as Super-Classes, Super-Super-Classes, and so on.[34] This would allow her to maintain the intuitive idea that there is no largest "size" in the loose sense, either. But, importantly, the idea that the universe of *sets* is as "large" as possible is captured by a more specific notion of "size" that applies to sets and that is defined in set theory. And once again, not all things that intuitively or informally count as "sizes" or "collections" are set-theoretic entities, on pain of contradiction; this is simply one of the fundamental commitments of the actualist view.

I will flesh out the actualist position and specifically the textbook "on pain of contradiction" explanation further in §**??**. But we can already conclude that the actualist can coherently maintain both that there is no arbitrary "stopping point" of the universe of sets and that there is a "size" in the loose sense above which collections are too "large" (in the loose sense) to form sets.

In the end, I don't think there is any hope for arguing for potentialism on the basis of its explanatory powers or advantages. This undermines a lot of what is standardly said in motivating

[32] The practice described, e.g., in (**?**).

[33] Of note is that the potentialist who accepts classes as intensional entities can also accepts this (see footnote **??**).

[34] These types of collections are discussed, e.g., in (**?**).

potentialism. I do think that there is an alternative and more promising way to argue for potentialism. But, as we will see, this will also reveal that the debate over potentialism versus actualism isn't as substantial as one might have thought.

3 Arguments from an intuitive conception of sets

As I see it, a much more promising argument for potentialism starts from the observation that our intuitive conception of sets has some modal component. For some, the intuitive conception in question is the "iterative" conception of sets, which is sometimes taken to be implicit in the writings of Cantor and Zermelo. For instance, in the quote we saw earlier, Parsons suggests "interpreting Cantor by means of a modal language" (**?**, 346), and, as the title of that paper suggests, he takes himself to be answering the question "What is the iterative conception of sets?" So, Parsons seems to think that the iterative conception of sets, implicit in the writings of Cantor, is best understood as having some modal component. Similarly, **?** starts by explaining that on the "familiar iterative conception" there "seems to be something inherently potential about the set theoretic hierarchy" (**?**, 205). Along the same lines, in their papers on potential infinity, Linnebo and Stewart Shapiro discuss how Cantor and Zermelo left room in their writings for some "limited form" of potential infinity (**?**, 286). It is also common for potentialists to focus on the naïve conception of sets and to suggest, for instance, that the modal statement that any sets can form a set "retains the intuitive plausibility of Naive Comprehension" (**?**, 12).[35] The idea here seems to be that our intuitive conception of sets includes something like the claim that the extension of every predicate is a set, and that this is best understood as having a modal component. In both cases, potentialism is then introduced as a theory that makes these modal conceptions of sets precise.

As I see it, this approach can yield a strong argument for potentialism, provided that the intuitive conception of sets in question is itself true: potentialism is true because it is entailed by our intuitive conception of sets, which is itself true. This is the general form of what I call the 'argument from an intuitive conception' for potentialism. To be successful, the argument from an intuitive conception thus has to be supplemented with an argument for why the relevant intuitive conception of sets is true.

An "intuitive conception of sets" is standardly understood to consist of basic or even axiomatic statements about sets couched in an at least partly informal language. For instance, the iterative conception of sets is usually taken to include the statements that sets are formed in stages, that at the first stage, there is an empty set, that whenever there are two sets the set that contains them is formed at the next stage, and so on. These basic statements about sets usually can't be justified inferentially or proved from other basic statements. This leaves broadly two options for arguing

[35] See, e.g., **?**, 82f., or references in (**?**, 71f.), and perhaps **?**, 150. **?**, 15 in the quote above also suggests that the modal version of the Axiom of Naïve Comprehension has strong intuitive appeal.

for their truth: via abductive arguments, on which an implicit conception of sets is true because it best serves certain theoretical and explanatory roles in mathematics or the broader sciences,[36] or via arguments from implicit definitions, on which an implicit conception of sets is true because it implicitly defines the primitive mathematical expressions in it (such as 'set' and 'membership').[37]

In my view, potentialists and actualists alike should opt for the latter option. Firstly, abductive arguments for the truth of an intuitive conception of sets often invoke the conception's ability to explain set-theoretic paradoxes;[38] but this would reduce the argument from an intuitive conception of sets to the argument from explanatory power, which, as I argued in §**??**, is unsuccessful (though, as I will explain in §**??**, explanatory considerations can still be relevant to the argument from an intuitive conception via an argument from implicit definitions). Secondly, potentialists and actualists alike need a metasemantic theory; that is, they need an account of what determines the meanings or extensions of their set-theoretic expressions. But it is widely accepted that something like an implicit definition theory is the only viable metasemantic theory for the language of mathematics.[39] Thirdly, as I will explain in §**??**, an argument from implicit definitions can help potentialists and actualists further motivate and develop the textbook explanations for the non-existence of contradictory sets that I have been invoking in §**??** by revealing that they are analytic explanations. Finally, as I will explain in §**??**, an argument from implicit definitions can help potentialists explain their interpretations of their modality, and, in particular, it can help potentialists who want to use a primitive notion of modality defend and explain their interpretation. In sum, I think the potentialist's argument from an intuitive conception of sets is strongest and most fruitful when supplemented with an argument from implicit definitions for the truth of the intuitive conception in question. In the following §**??**, I will flesh out one such argument from implicit conception and explain its consequences for the debate between potentialism and actualism (in §§**??**–**??**). But note that even if one rejects the implicit definitions strategy in favor of the abductive strategy, my conclusion that the debate over actualism and potentialism can only be settled on the basis of pragmatic considerations will still stand. Indeed, if we set aside the virtue of explaining set-theoretic paradoxes (because of the arguments from §**??**), an abductive argument for one conception over another will invoke pragmatic virtues of simplicity, mathematical strength, usefulness, and so on, which are precisely the types of considerations I take to bear on the question of which meanings to assign to our set-theoretic expressions, as I will explain in §**??**.[40]

[36] See, e.g., (**??**), and more recently (**?**). The "anti-exceptionalist" views about logic and mathematics originating in (**??**) could be included in category.

[37] See, e.g., recently (**?**) or (**??**).

[38] E.g., **?**, 64–69 does this.

[39] See also my arguments in (**?**). Potentialists also generally accept this type of metasemantics; see, e.g., (**?**), (**?**, 33ff., 135ff.), or (**?**, 155ff.).

[40] See again, e.g., (**?**, 44–69), (**??**), (**?**).

3.1 A descriptivist metasemantics

An *argument from implicit definitions* for the truth of some intuitive conception of sets says that this intuitive conception of sets is true because it is "intrinsically justified," or "analytically" true, or "implicitly defines" the primitive set-theoretic expressions (that is, the expressions that aren't explicitly defined). The details of the argument from implicit definitions won't make a difference to my overall claims about the debate between actualism and potentialism, but, for concreteness, let me outline a specific (and my preferred) way to run this argument. We start with a metasemantic view of how set-theoretic expressions get their meanings and extensions from their use, known variously as 'descriptivism', 'the method of implicit definitions', 'the Hilbertian Strategy', or 'the Ramsey–Carnap–Lewis method'.[41] On this metasemantic view, speakers use certain sentences in a privileged way, and their privileged use determines that these sentences are true (or, perhaps, that they are true if the use is consistent). The sentences that are used in this privileged way are thus in some sense "analytic" or "true in virtue of meaning." So, if the sentences that are part of some intuitive conception are used in the relevant privileged way, they are also thereby true in virtue of meaning.

Here is one way to spell out this kind of metasemantic view in a bit more detail.[42] On a descriptivist metasemantics, speakers associate primitive set-theoretic expressions with a certain theory or description, and these set-theoretic expressions thereby have meanings or extensions that make this associated theory true (if the theory is consistent).[43] The "speaker associations" here are intended to capture the speakers' use of the expressions: to associate some expressions $E_1, \ldots, E_n$ with a theory $T(E_1, \ldots, E_n)$ is, roughly, to be disposed to accept sentences in $T(E_1, \ldots, E_n)$ "come what may," that is, no matter what evidence one were to have.[44] "Accepting" some $\phi \in T(E_1, \ldots, E_n)$ can manifest itself in various ways; for instance, one can be disposed to utter ϕ, to say 'Sentence ϕ is true', to base one's reasoning on the truth of ϕ, and so on.[45] Two further clarifications will be relevant for the discussion that follows. Firstly, associated theories don't need to be in a formal, uninterpreted language; they can have meaningful expressions of ordinary language.[46] Expressions that occur in an associated theory but that aren't implicitly defined by that particular associated theory are what **?** calls the 'old' terms of an associated theory or what I like to call 'anchors' (**?**). So, in other words, a theory associated with our set-theoretic expressions can be "anchored" in

[41] This is also a version of inferentialism, as explained in (**?**) and (**?**).

[42] This is my preferred view, which I develop in (**?????**).

[43] For (**?**), for instance, consistency isn't even a requirement, **?** discusses that Σ_1-soundness might be needed; see (**?**) and (**?**) for discussion. It is standardly thought that the sentences in the theory conditionalized on the existence of entities that satisfy the theory are themselves automatically true (see, e.g., (**?**) and (**?**)).

[44] More precisely, it is to be disposed to accept them conditional on the existence of $E_1, \ldots, E_n$. For more discussion and explanation of the descriptivist view, see, e.g., (**??**).

[45] As explained in, e.g., (**?**) or (**?**, 33ff.).

[46] This is unlike on the "formalist" Hilbertian or Carnapian readings of the descriptivist metasemantics, which is very common; see, e.g., (**?**). As I argue in (**??**), realizing that associated theories needn't be in a formal language where only the logical expressions are assumed to have meanings resolves numerous philosophical puzzles about mathematics.

ordinary language. (Moreover, the primitive set-theoretic expressions themselves don't need to be restricted to the formal symbol '$\in$'; they can be, say, 'set' and 'membership'.) Secondly, the descriptivist view leaves open that not all set theorists associate the same theory with set-theoretic expressions. Distinct associated theories T_1('set', 'membership') ≠ T_2('set', 'membership') could yield the same meanings and extensions for the primitive set-theoretic expressions in case these theories are "equivalent" in some strong sense (call this 'translational equivalence'). For instance, we would like to say in general for this type of metasemantic view that expressions that have mere typographical or grammatical differences can still have the same "privileged use," and, thus, the same meaning.[47] In such cases, we would thus like the respective associated theories to come out as translationally equivalent.[48] But, of course, distinct associated theories don't need to be translationally equivalent. Descriptivism thus also leaves open that different set theorists use the primitive set-theoretic expressions with different meanings (as well as, perhaps, different extensions).

With this descriptivist metasemantic view on the table, potentialists can say that among the statements that are part of the theory we associate with our set-theoretic expressions are modal statements, such as the statement that any things can form a set. These modal statements—and potentialism itself—are thus true because they are part of the theory we associate with our set-theoretic expressions. In other words, the potentialist can argue that their modal intuitive conception of sets is part of the theory associated with set-theoretic expressions, and, thus, that it is true (in virtue of meaning). This is the argument for potentialism from an intuitive conception of sets via the implicit definitions strategy. I think this type of argument is at least implicit in the remarks that potentialism uncovers something that is part of our intuitive conception of sets or that it captures our naïve intuitions about sets. And I think that this type of argument is very strong, because something like this metasemantic story must be correct for mathematical expressions.[49]

The key premise potentialists need to defend is that modal statements really are part of the theory we associate with our set-theoretic expressions. As I see it, potentialists have two options for making this case. First, they can say that ordinary modal expressions such as 'can', 'possible', or 'necessary', occur in the theory we associate with our primitive set-theoretic expressions and they retain their standard meanings; in other words, the potentialist can say that the theory associated with set-theoretic expressions has modal anchors. The second option is to say that modal expressions are part of the primitive expressions implicitly defined by the theory associated with set-theoretic expressions; in other words, the potentialist can say that we associate a theory T('set', '$\in$', 'can') with our set-theoretic expressions together with an expression of set-theoretic or mathematical modality, 'can'. Here, the word 'can' thus gets a new set-theoretic or mathematical sense in virtue of this association. As I see it, this second option yields the best way to understand

[47] See, e.g., the notion of "stylistic variance" discussed by (**?**, 139ff.), or notions of notational variance discussed by, e.g., (**??**).

[48] See (**?**) for the development of such a notion of translational equivalence.

[49] As potentialists themselves agree, see footnote **??**.

the potentialist's claim that her modality is primitive or idiosyncratic to set theory or mathematics, or the claim that the potentialist can only specify "structural features" that any interpretation of the modality should have (**?**, 171): for, on this option, the story of the potential hierarchy of sets itself determines the (structural features of) the sense of 'can' at issue. The first option, in turn, can capture the other interpretations of the modality: for instance, one can anchor one's associated theory in the constructive interpretation of 'can' on which a sentence such as "there can be more sets" in the associated theory is assumed to mean that there is a metaphysically possible world in which some social process has yielded further sets than there are in the actual world. (As I will argue in §**??**, the linguistic interpretation of 'can' is harder to make sense of on this picture and this counts against it.)

On either option, the potentialist needs to argue that certain modal statements about sets are part of the theory we associate with set-theoretic expressions, that is, that we hold certain modal statements about sets true come what may, that certain modal statements about sets are part of the basic story of the hierarchy of sets we accept non-inferentially and intuitively. But the problem is that some people—and, in particular, actualists—will reject that modal statements are part of the theory that *they* associate with their set-theoretic expressions. For actualists, any modal talk in the description of the iterative hierarchy of sets is metaphorical, and it is eliminated in the ultimate story that is held true come what may; after all, actualism as I understand it here is precisely the rejection of potentialism (§**??**). Whatever intuitive conception actualists associate with set-theoretic expressions, modal expressions are neither going to be anchors nor further primitives in their associated theory. On a more positive characterization, the theory actualists associate with sets could simply be any standard non-modal formulation of the iterative conception of sets.[50] It would likely include statements such as 'There is an empty set', 'Every set is well-founded', and so on—so, statements that can be formalized or formulated as axioms of set theory.[51] As I mentioned in §**??**, actualists can also further specify the universe of sets as either a proper class or a mere plurality. So, actualists's associated theory might include statements such as 'the universe of sets is a mere plurality'; it could thus have the notion of a plurality as either an anchor or a primitive in their associated theory. Then, the statement 'the plurality of all sets isn't a set' would be entailed by the theory that the actualist associates with her set-theoretic expressions. Alternatively,

[50] Some options include part of the conception outlined by **?**, or even a stage theory, e.g., articulated by **?**, or **?**, 61–64's "minimal" conception of sets.

[51] Note that axioms themselves can be part of (or be entailed by) intuitive conceptions. In that case, an intuitive conception can "justify" an axiom if the intuitive conception is part of an associated theory: this would show that the axiom (or something that entails the axiom) is itself analytic. If axioms aren't themselves part of or entailed by some intuitive conception, then it is harder to account for how an intuitive conception can justify the truth of the axiom from the current perspective: if some intuitive conception C is part of an associated theory, and $C+A$ is a distinct, translationally non-equivalent theory, then strictly speaking the primitive set-theoretic expressions in C and in $C+A$ have distinct meanings; facts about C-sets won't justify facts about $C+A$-sets, and axiom A isn't true of C-sets. (This is the familiar point that Carnapian explications "change the subject.") As I will explain in §**??**, an alternative would be that the intuitive conception gives *pragmatic* motivation for an axiom A: one pragmatic reason to accept $C+A$ rather than $C+\neg A$ might be that the former is more similar to or captures the spirit of C. For further discussion, see, e.g., (**?**).

actualists could invoke the notion of a proper class—and perhaps even super-classes, and super-super classes, and so on—and hold true come what may statements such as 'The totality of all sets forms a proper-class', 'Proper classes aren't sets', 'Super-classes are neither sets nor proper classes', and so on. Either way, there is no modal language in the theory that the actualist associates with her set-theoretic expressions; from the current metasemantic perspective, we can even say that not having modal statements in the theory one associates with set-theoretic expressions is just what it is to be an actualist.

Actualists and potentialists thus associate different theories with their set-theoretic expressions. Unless their theories end up being translationally equivalent, this entails that actualists and potentialists mean different things by their set-theoretic expressions, and, thus, that the dispute between actualists and potentialists is merely verbal. If the potentialist and actualist theories are translationally equivalent, then there is no disagreement between the actualist and potentialist, for they are merely choosing different words to express the same things; for instance, the potentialist uses 'there can be' for what the actualist expresses as 'there is'. There is a current debate over whether the potentialist and actualist theories are translationally (or "notationally") equivalent; for instance, **?**, §7 argue that their theories aren't mere notational variants of actualist ones if higher-order resources are involved, while Tim **?** discusses this and the "near-synonymy" of many versions of the potentialist theories and actualist theories, suggesting that near-synonymy might suffice for translational equivalence and thus for sameness of meaning. I won't take a stand on this debate here.[52] I will only note that if potentialists are right, then this would entail, from the current perspective, that they mean something different with their set-theoretic expressions than the actualists, because the potentialist and actualist theories in question are associated, meaning-determining theories. Indeed, the argument from an intuitive conception for potentialism works precisely because the potentialist theory is (or is entailed by, or is a precisification of) an associated, meaning-determining theory. So, in either case, the dispute between the actualist and the potentialist isn't as substantial as one might have thought. I will explain how the debate over potentialism and actualism can move forward from here in §**??**. But first, since we now have the actualist view supplemented with the implicit definitions strategy, let me clarify two things left open in our discussion in §**??**.

3.2 Textbook explanations as analytic explanations

First, we can illuminate the actualist's textbook explanation for why there is no set of all non-self-membered sets (or of all cardinals, ordinals, and so on). The actualist can say that the fact that there is no such set follows from at least one principle that is part of the theory she associates with her set-theoretic expressions. For instance, surely, statements such as the Axiom of Foundation and

[52] Although I argue that neither near-synonymy nor synonymy (or definitional equivalence) suffice for translational equivalence on a descriptivist metasemantics (**?**).

the Axiom of Separation—or some other statements that entail ‘No set is a member of itself’—are going to be part of the theory actualists associate with their set-theoretic expressions. But this means that, whatever ‘set’ and ’membership’ mean, they mean something such that ‘No set is self-membered’ is (analytically) true. But then, it is a logical consequence of analytic truths that there is no set of all non-self-membered sets. This analytic explanation is as good an explanation as we can hope to get for why there is no set of all non-self-membered sets.[53]

To take an analogy, say (plausibly) that we associate the term ’bachelor’ with some theory that includes ‘every bachelor is unmarried’, that is, we hold ‘every bachelor is unmarried’ to be true come what may. Then, if someone asks “But why isn’t there a happily married bachelor?” the explanation we should give is: “There isn’t a happily married bachelor because it is analytically true that every bachelor is unmarried, and it is a simply consequence of this analytic truth that there isn’t a happily unmarried bachelor.” This is as good as explanations ever get. And I contend that the actualist has just as good an explanation for why there is no set of all non-self-membered sets: there is no set of all non-self-membered sets because it is an analytic truth that no set is self-membered (given what we associate with our set-theoretic expressions), and it is a simple consequence of this that there is no set of all non-self-membered sets. This is as good as explanations get; it would be just as absurd to keep asking “Ok, but *why* is there no set of all sets?” as it would to keep asking “Ok, but *why* is there no happily married bachelor?”

The lingering worry articulated by Linnebo that we discussed in §**??** should no longer have force. Consider the plurality of all sets. It follows from how we define ‘set’ that everything in that plurality is something that isn’t self-membered. That is, there are no self-membered sets in this plurality, and thus there is also no set of all the sets in this plurality. The same thing could be said about bachelors: Consider the plurality of all bachelors. There is no married person in this plurality, simply because of how we define ‘bachelor’.

It can be hard to dispel the idea that there is still something unexplained on the picture that I am proposing. Here is an attempt to capture this idea. Say we consider the plurality of all sets; but now we imagine that they are all “laid out” in front of us. On the face of it, one can then imagine that it is perfectly possible for there to be one more set that contains all of these sets, laid out in front of us; we imagine “lassoing” all the sets laid out in front of us. But that is an illusion. If what is laid out in front of us really is the plurality of all sets, then all the sets are already out there in front of us. If one imagines, say, a (finite) list of sets, then one can genuinely imagine the set that contains all of the sets on this finite list. But that is very different from imagining *all sets* laid out in front of us. Think of another analogy. One can imagine the set of all natural numbers, “laid out” in front of us. On the face of it, one can then imagine that there is a natural number that is larger than any number in this set. But of course, there can’t be. If what we are really imagining is the

[53] This cashes out the proposal I make in (**??**) to understand conception-based explanations as analytic or conceptual explanations.

infinite series of the natural numbers, then there are no more natural numbers to be added to the series! On the actualist perspective, these are both instances of how our intuitions about infinities can be radically misleading.

Finally, note that potentialists have the same type of explanation available to them, using their own associated theory. In fact, the (non-modal) Axiom of Foundation itself is usually accepted as part of the potentialist conception of sets,[54] so the exact same explanation just discussed is also available to the potentialist. So, when it comes to explaining why there is no set of all non-self-membered sets—or no set of all ordinals, or all cardinals, and so on—actualists and potentialists can be exactly on a par.

3.3 Unexplained impossible existence on a primitive modality

The second point that we can now clarify is why actualists don't need to explain why there can't be a set of all sets on a sense of 'can' that is primitive to the potentialist. As I explained in §**??**, the best way to make sense of a primitive potentialist modality is to say that 'can' gets its meaning together with primitive set-theoretic expressions by being associated with a theory that involves them all. What this means is that the word 'can' has this primitive meaning only for speakers who are disposed to accept the potentialist theory come what may. Actualists aren't such speakers. This doesn't necessarily mean that the potentialist's sense of 'can' can't be translated into the actualist's language. But to avoid ambiguity, when the potentialist asks 'Why can't the plurality of all sets form a set?', the actualist should first read this as something like: 'Why shcan't the plurality of all shmets form a shmet?' One answer the actualist can then give to this question is that shmets are simply outside the scope of her theory. And if there is a translation of the potentialist's expressions into the actualist's, then the sentence 'The plurality of all shmactually existing shmets shcan form a shmet', which is (presumably) entailed by the potentialist's associated theory, should be translated into a sentence that is true given the actualist's associated theory, since an adequate translation should at least preserve analytic truths or truths entailed by the associated (meaning-giving) theory. But then, the actualist would also be able to explain why this fact holds about sets, given her associated theory. So, once again, the potentialist hasn't shown that there is an unexplained fact on the actualist picture.

3.4 A cost–benefit comparison of actualism and potentialism

Let us take stock. I argued that the best way to argue for potentialism is to say that it follows from an implicit definition of set-theoretic expressions. But then, given that there are other ways to implicitly define set-theoretic expressions—such as, in particular, the actualist's—the dispute between potentialism and actualism should be understood as merely verbal: Actualists

[54] See, e.g., (**?**, 216f.)

and potentialists (plausibly) mean different things with their set-theoretic expressions, and, thus, they don't really disagree with each other over whether the universe of sets is potential. Rather, the universe of sets is potential on the potentialist's sense of 'set', while it isn't on the actualist's sense of 'set'. As with all verbal disputes, the natural next question is whether there is any reason to pick one meaning assignment over another. That is, is there reason to associate the potentialist theory with 'set' and 'membership' as opposed to the actualist theory? Or, if we disambiguate and associate, say, 'p-set' and 'p-membership' with the potentialist's associated theory and 'a-set' and 'a-membership' with the actualist's associated theory, is there reason to speak a language with 'p-set' and 'p-membership' as opposed to a language with 'a-set' and 'a-membership'? These kinds of questions can only be resolved by considering pragmatic (or perhaps normative) reasons. Indeed, there is no question of which choice gets at the truth, since both options will yield their respective (analytic) truths; that is, these aren't pragmatic reasons for or against the *truth* of potentialism or actualism. For simplicity (and without loss of generality), let me only consider the question of which theory to associate with (and thus which meaning to assign to) the expressions 'set' and 'membership'. I will conclude by considering some pragmatic reasons that bear on this question and suggest that these weigh against associating the potentialist theory with the primitive set-theoretic expressions; or, as I will say, "against being a potentialist." As I explained in §§**??**–**??**, I take the most plausible versions of potentialism to be the ones with the primitive and the linguistic interpretations of the modality. I will thus only consider these two options in my cost–benefit comparison.

There are many different types of pragmatic reasons one can give for or against using certain mathematical expressions with certain meanings. I won't aim to be exhaustive here, and my discussion can thus also be seen as an invitation to consider further pragmatic costs and benefits on behalf of potentialism and actualism. But here are what I take to be the broad categories of pragmatic criteria to consider: simplicity and ease of use, fruitfulness (mathematical, scientific, or perhaps even philosophical), faithfulness to some historical or pre-theoretical conception, and explanatory power (mathematical or otherwise). The arguments in §**??** support that the potentialist language doesn't have any explanatory advantage over the actualist language. As far as I know, no version of the potentialist theory promises to have useful consequences for the natural sciences, so I will set considerations of scientific fruitfulness aside. Most versions of the potentialist formal theories are at least mutually interpretable with standard ZFC (as I mentioned in §**??**, **?** argues that many are even near-synonymous), and thus arguably don't provide any new mathematical results (including ease of proof) over ZFC.[55] This suggests that actualism and potentialism are on a par in terms of mathematical fruitfulness (though simplicity and ease of expression can also help mathematically, and, as I will explain, those considerations favor actualism). The fact that non-modal ZFC is still the standard choice in the mathematical community might also suggest that poten-

[55] See, e.g., (**?**) and (**?**).

tialism isn't more fruitful mathematically than actualism. In any case, the standard motivations for potentialism are philosophical, so more would need to be said to motivate the mathematical fruitfulness of potentialism. I thus also set considerations of mathematical fruitfulness aside here.

3.4.1 Pragmatic comparison of actualism vs. potentialism with a primitive modality

Consider, first, the option where the potentialist's modality is primitive. I think one can make the case that it is a pragmatic cost to have one more primitive expression in one's associated theory, especially if that expression already has a standard meaning in the background language. Having an extra such primitive adds some complication to the associated theory and ambiguity in the overall language, both of which constitute at least some pragmatic cost.

Take, next, the criteria of faithfulness to some historical or pre-theoretical notion. As I mentioned at the beginning of §**??**, some potentialists motivate their potentialist conception by saying that it captures the intuitive part of the naïve (and inconsistent) conception of sets, because the potentialist conception makes true the modal version of Naïve Comprehension: for instance, on the version concerning pluralities, this is the statement that every "plurality could form a set," as **?**, 12f. put it. Even granting that it is desirable to retain a version of Naïve Comprehension in a theory of sets,[56] I don't see how a primitive notion of "can" could help capture any intuitive version of Naïve Comprehension. Arguably, 'can' in the potentialist's associated theory is a modality in name only: For one, it isn't any of the intuitive notions involved in any characterization of the intuitive idea of Naïve Comprehension. Moreover, as Andrew **?** has argued in a slightly different context (concerning "width" potentialism), the potentialist's primitive mathematical "modality" seems to flout some of the core orthodoxies about modality in general.[57] Potentialists would thus have to argue that there is nonetheless something of the intuitive idea that things "can" form sets that is retained in their statements such as 'every plurality can form a set', which seems to be a major challenge.

To summarize, there are some pragmatic costs to potentialism with a primitive modality, and there are no clear pragmatic benefits to it. The balance of pragmatic reasons seems to favor actualism.

3.4.2 Pragmatic comparison of actualism vs. potentialism with a linguistic modality

Second, consider the option of anchoring the potentialist's associated theory with a linguistic kind of modality. Some have argued that potentialists adopt some modal principles that conflict

[56] I doubt this; indeed, it has also been argued that Cantor's conception of sets is fundamentally different from the logicists' "naïve" conception (see, e.g., (**?**)), and so one might argue that our concept of set should be faithful to Cantor's notion rather than to the naïve notion.

[57] **?** shows that this "modality" radically flouts Brouwer's principle in that it is possible that there are truths that are possibly impossible and Leibniz biconditionals (i.e., what is possible, in the broadest sense of possible, is what is true in some broadly possible world) (**?**, 132).

with the intuitive linguistic understanding of the modality at issue.[58] I think these concerns are serious, and, in our framework, they put pressure on whether the potentialist's modality really is anchored in a linguistic type of modality. A related worry is that, on a natural understanding of the potentialist's linguistic interpretation, the potentialist theory is merely describing how one could use the non-logical expressions of the language of set theory (that is, 'set' and '∈', or just '∈') with different (more and more expansive) meanings, without actually (at least explicitly) talking about sets.[59] On this understanding, the argument from an intuitive conception for potentialism as I have developed it wouldn't apply: the potentialist isn't implicitly defining a modal conception of sets, but, rather, she is theorizing about different kinds of possible semantic changes involving 'set' (that is, their theory is about 'set', and not *sets*).[60] On this interpretation, it is clear that actualists and potentialists are talking past each other: As I discussed in §§**??**–**??**, the actualist can grant that one could change the meaning of 'set', but she doesn't want to use 'set' with different meanings. Instead, she is theorizing about sets on her (fixed) use of 'set'. The actualist can capture the more expansive meanings that the potentialist is concerned with, simply with different words than 'set'—for instance, 'proper class', 'super class', and so on. So, what would be the pragmatic benefit of adopting a theory of semantic change as opposed to the actualist's theory of sets?

?, 12 argues that the potentialist theories on this linguistic interpretation don't commit one to the existence of sets (since they only theories about how one could use language), and thus they could accommodate a nominalist metaphysics. Some might find this to be a philosophical advantage. But note that actualism, too, is consistent with various ontological views about the nature of sets (structuralism, thin realism, robust realism, and so on), depending on what else one decides to include in one's associated theory. Actualism could even be combined with the claim that sets don't exist, if it is understood, for instance, that an associated theory merely delineates analytic truths about sets that are conditional on the existence of sets (roughly such as "Carnap-sentences" in (**?**)).[61]

A second reason to adopt the potentialist theory on this linguistic understanding might be that it adequately captures a real phenomenon of semantic change. But it is unclear whether anyone actually uses 'set' in this ever-changing way, and there seem to be reasons against speaking in this way. In general, there is some pragmatic cost to using a word with multiple different

[58] E.g., **?**, 62f. discusses how the "maximality" principle "[a]t every stage, all the entities that can be introduced are in fact introduced" that Linnebo adopts conflicts with the intuitive understanding of his linguistic modality, since we don't introduce all abstraction principles at once (see also Berry's similar points against Studd's linguistic modality (**?**, 63–65). Similarly, **?**, 114f. argues that the intuitive linguistic understanding makes little sense of iterated modalities or quantifying into modal contexts.

[59] See, e.g., the views of (**?**) and (**?**).

[60] **?** might be giving an alternative view on which the meanings of the set-theoretic expressions don't change but only the meanings of the quantifiers do. On this view, then, we could say that sets are implicitly defined in terms of different quantifiers, e.g., that Warren's NC+: $\exists^+ y \forall x(x \in y \leftrightarrow \phi(x))$ (or perhaps a version of it with a 'set' predicate) is part of the theory associated with set-theoretic expressions (**?**, 104).

[61] See (**?**) for discussion of the related problem of existence for descriptivism about the language of set theory. See also clarifications about this point in §**??**.

meanings, and to changing the meaning of a word midway through some piece of reasoning. At the very least, this complicates the semantic theory for that expression, and it can lead to miscommunication among speakers. But what the potentialist theory is capturing, on this view, is precisely the practice of changing the meanings of one's words, including right in the middle of a reasoning process. For instance, the idea is that in running through the reasoning of Russell's paradox, we shift the extension of our quantifiers and/or set-theoretic expressions (**?**, 102–110), which is why it is "possible" that there is a set of all those (actual) non-self-membered sets. If there is an alternative way of speaking without switching the meanings of our expressions, there would be some pragmatic reason to prefer it. But that is precisely what the actualist associated theory is giving us: we can use 'proper class' or 'mere plurality' for the collections that are sets on the expanded understanding of 'set', without changing the meaning of any of our expressions.[62]

Consider, finally, the criterion of faithfulness. Once again, I think it is unclear that potentialism here captures the intuition that there can be more sets. Surely, the intuition wasn't that one can have more sets because we can change the meaning of 'set' (or of the quantifiers). Rather, the intuition was that there can be more things *like those things*, the sets, so, more "collections" in some intuitive sense of 'collection' (and under the standard sense of the quantifiers). In contrast, it is available to the actualist to say that, although there aren't more sets on pain of contradiction, there are indeed more "collections": namely, proper classes, super-classes, and so on. Logic tells us that for no collection of type X, there is an X of all Xs. But for any collection of some type, there is some collection of some other type of all the collections of the former type. As I see it, this story captures the original intuition even better than the potentialist story on the linguistic interpretation (or on the interpretation with a primitive modality).

Once again, we saw some pragmatic costs to potentialism this time with a linguistic modality, and no clear pragmatic benefits. The balance of pragmatic reasons seems once again to favor actualism over potentialism.

4 Conclusion

This paper examined arguments for being a potentialist. The most common type of argument given for potentialism are arguments from explanatory power. I argued that these arguments are unsuccessful: either the potentialist is on a par with the actualist when it comes to explaining facts about sets, or the facts that are purportedly unexplained on the actualist picture should only be accepted if one is a potentialist to begin with. I then outlined what I take to be the most promising line of argument for potentialism: potentialism is true because it is part of the meaning-determining, analytic part of a theory of sets. The problem with this argument is that not everyone adopts (nor should adopt) this potentialist conception of sets. Moreover, as I argued, there are

[62] These are also reasons against adopting NC+ as part of our associated theory (see footnote **??**).

numerous pragmatic reasons against adopting the potentialist conception of sets over the default, actualist, conception of sets that underlies our standard mathematical practice. Why not, then, simply be an actualist?[63]

[63] For very helpful comments and discussion, I am grateful to Sharon Berry, Tom Donaldson, Jens Kipper, Daniel Rowe, Daniel Waxman, and audiences at the University of Konstanz and the University of Oxford. I am also grateful to two anonymous reviewers whose comments substantially improved the paper.

On the limits of comparing subset sizes within $\mathbb{N}$

Sylvia Wenmackers[*]

July 22, 2024

Abstract

We review and compare five ways of assigning totally ordered sizes to subsets of the natural numbers: cardinality, infinite lottery logic with mirror cardinalities, natural density, generalised density, and α-numerosity. Generalised densities and α-numerosities lack uniqueness, which can be traced to intangibles: objects that can be proven to exist in ZFC while no explicit example of them can be given. As a sixth and final formalism, we consider a recent proposal by Trlifajová (2024), which we call c-numerosity. It is fully constructive and uniquely determined, but assigns merely partially ordered numerosity values. By relating all six formalisms to each other in terms of the underlying limit operations, we get a better sense of the intrinsic limitations in determining the sizes of subsets of $\mathbb{N}$.

1 Introduction

The size of subsets of natural numbers played a central role in historical debates on the vexing notion of infinity. Already in the ninth century, Thabit ibn Qurra defended the existence of different sizes of infinite collections of natural numbers (see, e.g., Mancosu, 2009, §2). Galilei (1638) famously compared the set of perfect squares to the full set of natural numbers. The finding that the former is both a proper subset of the latter (which suggests a smaller size) and can be put into one-to-one correspondence with it (which suggests equal size) is known as Galileo's paradox. Galileo concluded that sizes of infinite collections cannot be meaningfully compared.[1]

[*] Centre for Logic and Philosophy of Science, Institute of Philosophy, KU Leuven, Belgium. E-mail: sylvia.wenmackers@kuleuven.be, URL: https://www.sylviawenmackers.be/

[1] On the first day of the dialogue, Salviati says that "we cannot speak of infinite quantities as being the one greater or less than or equal to another" (Galilei, 1638, p. 31); he argues for this based on a consideration of the 'number' of all (natural) numbers and of (perfect) squares and their roots, from which he concludes that "the attributes 'equal,' 'greater,' and 'less,' are not applicable to infinite, but only to finite, quantities" (Galilei, 1638, pp. 32–33).

While many early mathematicians similarly rejected the idea that the infinite can be represented by an extended number system,[2] later mathematicians have found different ways of expressing sizes of infinite sets. Here, we take 'size' to be a pretheoretical term, that can be formalized in different ways: not only as cardinality, but also in terms of natural density and other approaches. In this article, we take the set of natural numbers $\mathbb{N} \overset{\text{def}}{=} \{1, 2, 3, \ldots\}$ as the canonical example of a countably infinite set. We focus on its collection of subsets, the power set $\mathcal{P}(\mathbb{N})$, to discuss and compare different notions of size.

We start by reviewing and comparing five formalisms for assigning totally ordered (also called linearly ordered) sizes to subsets of $\mathbb{N}$: cardinality, infinite lottery logic, natural density, generalised density, and α-numerosity.[3] In all cases, the notion of 'size' is altered compared to its initial meaning that only applied to finite collections. This has been investigated before (see, e.g., Mancosu, 2009, Parker, 2013), but a new theory proposed by Trlifajová (2024) invites us to examine it once again. This paper highlights the non-uniqueness and non-constructive aspects of sizes assigned to infinite co-infinite sets on some accounts. In contrast, Trlifajová's (2024) theory is constructive but also requires us to drop yet another property of earlier notions of size.

The main goal of the present paper is not to defend a particular theory as providing us with the best notion of size, but rather to study the connections between various formalisms and to figure out which insights they give us collectively into size-related properties of subsets of $\mathbb{N}$. To achieve this, we will reconstruct the six theories in terms of an underlying limit operation that acts on the sequence of partial sums of the characteristic sequence of a given subset of $\mathbb{N}$ or the corresponding density, relative to $\mathbb{N}$.

We explore this issue in the context of standard set theory: Zermelo–Fraenkel set theory with the Axiom of Choice (ZFC), which takes the notion of arbitrary subsets of $\mathbb{N}$ (and other infinite sets) for granted. Authors like Feferman (1999, p. 102) objected to this approach, which raises a potential concern for our present project: does it really make sense to assign sizes to arbitrary subsets of natural numbers? Cardinality does so without a problem, which is unsurprising given that ZFC was designed with Cantor's approach in mind. Natural density fails to assign a size to some subsets of the set of natural numbers. Natural density can be extended to $\mathcal{P}(\mathbb{N})$, but

[2] But see, e.g., Mancosu (2009) for notable exceptions, such as Grosseteste, Maignan, and Bolzano. We review a recent construction of Bolzano's approach in section 3.

[3] These five formalisms suffice for our present purposes, without constituting an exhaustive overview. Other methods for assigning sizes to subsets of $\mathbb{N}$ exist. For instance, ideals and filters give a different notion of 'small' and 'large' sets (see, e.g., Schechter, 1997, p. 101). Although we will encounter ideals and filters in relation to α-numerosities, and there may be interesting connections to other parts of this paper, we will not discuss this approach here.

not in unique way: the underdetermination shows up for some infinite co-infinite sets. As we will see, α-numerosity is defined for all subsets of natural numbers but also lack uniqueness, which affects *all* infinite co-infinite sets.

We discuss multiple ways of dealing with the underdetermination of α-numerosity. In particular, Trlifajová's (2024) formalism trades the non-uniqueness of totally ordered α-numerosity for a partially ordered notion of numerosity.

Throughout this article, we will pay special attention to *intangibles*, in the terminology of Schechter (1997, §14.77): objects that can be proven to exist in ZFC but for which it can also be proven that no explicit example of them can be constructed, because their existence cannot be proven in ZF with a weaker choice principle, such as dependent choice (DC). Schechter (1997, §6.2 and §14.77) describes ZF + DC as 'quasiconstructive', in contrast to Bishop's constructivism, which rejects the law of the excluded middle. In this article, we similarly focus on the effects of invoking a nonconstructive choice axiom, without other constructivist considerations.[4] Like Schechter (1997, §14.77), we assume the view that intangibles are "*created by* our acceptance of the Axiom of Choice".

Further on, it will be helpful to be able to compare how fine-grained different measures are. We say that a measure m is *more fine-grained* than a measure M (or, equivalently, that M is *more coarse-grained* than m) if $m(S) = m(T)$ implies $M(S) = M(T)$ for all S and T in the domain where both measures are defined, but there exist S and T such that $M(S) = M(T)$ while $m(S) \neq m(T)$. In other words, there can be no M-difference without an m-difference, but the inverse does occur. This phrasing, familiar from the literature on emergence, already suggests that M may be multiply realizable by various more fine-grained measures; this is indeed what we shall find.

The paper is structured as follows. Section 2 reviews the five formalisms and compares them in terms of fine-grainedness and underdetermination. All formalisms are consistent on their own and have specific applications. Moreover, their results can be combined fruitfully, provided that we do not conflate their different notions of 'size'. If we violate this condition, we run into inconsistencies such as Galileo's paradox. Section 3 introduces the sixth approach by Trlifajová (2024) and reconstructs all theories in terms of different limit operations. Section 4 summarizes our findings.

2 Formalisms for assigning totally ordered sizes to subsets of $\mathbb{N}$

Dedekind (1888) defined infinite sets as those sets that have a proper subset

[4] See Maschio (2020) for a discussion of natural density in the context of Bishop's constructivism.

that can be put into one-to-one correspondence with the full set.[5] This shows that a version of Galileo's paradox will arise for any infinite set and thus complicates any assignment of size to such sets. Set sizes can be compared by the subset relation (between a set and another one contained in it) and by one-to-one correspondence (between the elements of two sets of equal size). Both methods give compatible verdicts when comparing sizes of finite sets, but Galileo's paradox and Dedekind's definition both show that this is not the case for infinite sets. This observation can be reconstructed as a dilemma between these two comparison methods (Mancosu, 2009): to extend the notion of size to infinite sets, one can either respect the ordering induced by the subset relation (called the part–whole principle or the Euclidean principle[6]) or define equality based on the existence of a bijection (called Hume's principle[7]), not both.

Historically, the second horn of the dilemma was chosen first: Cantor (1895) used it to develop cardinality theory. More recently, Benci and Di Nasso (2003) chose the first horn to develop α-numerosity theory. However, Bolzano's *Paradoxes of the Infinite* from 1848 can be regarded as a precursor to the latter idea: Trlifajová (2024) has given a reconstruction, which we review in section 3.

As we will see, natural density retains neither: in number theory, it is usually interpreted as a probability (see, e.g., Tenenbaum, 2015, Ch. III.1). If the measure does not express set size, then this classical dilemma does not apply to it. In this paper, however, we do consider natural density as a notion of (relative) size, showing that the paradox can be reconstructed as a trilemma instead.

Below, we review five formalisms that allow us to assign totally ordered sizes to subsets of $\mathbb{N}$: cardinality (§2.1), infinite lottery logic which includes mirror cardinalities (§2.2), natural density and generalised density (§2.3), and α-numerosity (§2.4).

2.1 Cardinality

Cardinality theory starts from Hume's principle, which takes the existence of a bijection between sets as indicative of their equal 'size' expressed as a cardinal number. So, cardinality is defined via equivalence classes on sets between which there exists a one-to-one mapping.[8]

[5] Bolzano had given a similar characterization earlier; see, e.g., Mancosu (2009).

[6] Named after the fifth and final 'Common Notion' in Book I of Euclid's *Elements*, which says that the whole is greater than the part.

[7] Named after Hume (1739–1740, Book I, Part III, §I): "When two numbers [i.e., collections] are so combined, as that the one has always a unit answering to every unit of the other, we pronounce them equal", as quoted (in German) by Frege (1884, Ch. IV, §63).

[8] For a brief contemporary presentation of cardinality theory, see, e.g., Jech (2003, Ch. 3).

The cardinality of a (well-orderable) set S is written as $|S|$. Assuming ZFC, all sets are well-orderable, so all sets have a well-defined cardinality. In particular, since $\mathbb{N}$ is well-ordered by its standard canonical order, cardinal numbers of subsets of $\mathbb{N}$ are fully determined: they are equal to the natural number of elements for finite subsets and they are equal to the first infinite (or 'transfinite,' on Cantor's terminology) cardinal number, $\aleph_0$, for all infinite subsets of $\mathbb{N}$. In other words, while they are sensitive to singleton differences between finite sets (i.e., maximally fine-grained), they maximally coarse-grain the sizes of infinite sets by mapping co-finite as well co-infinite infinite sets to the same value.

2.2 Mirror cardinality

Rather than applying Hume's principle to sets themselves, one can apply it to their complements instead; doing so requires a fixed superset, here $\mathbb{N}$. This approach yields "mirror images of the cardinals" for co-finite sets, as discussed by Mancosu (2015, p. 386).

A related proposal has been raised in the context of probability. Norton (2021) considered full label-invariance on countable sets as a notion of uniformity stronger than mere singleton-uniformity: he required that the measure assigned to a subset of a countable set should be invariant under permutations of the labels. Since permutations are one-to-one mappings, we can expect the proposal to be close to cardinality.

Indeed, the approach of Norton (2021) distinguishes the probability of finite subsets by their finite cardinality (maximally fine-grained) and similarly distinguishes the probability of co-finite subsets. All infinite co-infinite sets get the same probability rank, so this remains coarse-grained. In particular, Norton (2021, §8) introduced infinite lottery logic in terms of a chance function, Ch. To finite non-empty sets with n elements, Ch assigns the valuation V_n, which may be read as 'unlikely'. To co-finite sets that are complementary to sets with n elements, Ch assigns the valuation V_{-n}, which may be read as 'likely'. The values of the empty set and the full set are V_0, read as 'certain not to happen' and V_{-0}, 'certain to happen', respectively. Finally, Ch maps all infinite co-infinite sets to the same value, V_∞: this is read as 'as likely as not'. The readings of the values are informal interpretations, motivated by their ordering. This is given by an antisymmetric, transitive and irreflexive relation, $<$, on them (Norton, 2021, p. S3866):

$$V_0 < V_1 < V_2 < V_3 < \ldots < V_\infty < \ldots < V_{-3} < V_{-2} < V_{-1} < V_{-0}.$$

Here, we observe that this approach could be considered as an alternative conception of size as well, not based on the subset relation, but on the existence of bijections augmented with the relation of complements (relative to $\mathbb{N}$). Like cardinality, it is total on $\mathcal{P}(\mathbb{N})$ and uniquely determined, but it is more fine-grained. It still coarse-grains all infinite co-infinite sets.

2.3 Natural density

Number theory offers natural density as a different notion of size that is able to make distinctions among infinite co-infinite subsets of the natural numbers. While natural density is more fine-grained than approaches based on bijection, we will review in this section that it remains coarse-grained to some extent and that the measure is not total on $\mathcal{P}(\mathbb{N})$.

The natural density (or asymptotic density) d of a subset of natural numbers is defined as

$$d(S) \overset{\text{def}}{=} \lim_{n \to \infty} \frac{|(S \cap \{1, 2, 3, \ldots, n\})|}{n}, \tag{1}$$

with $S \in \mathcal{P}(\mathbb{N})$ such that this limit is indeed defined. In words, the natural density of a subset of $\mathbb{N}$ is the limit (if it exist at all) of the (finite) cardinality of the intersection of this set with an initial segment of $\mathbb{N}$, as this initial segment goes to infinity.

The natural density can also be understood in terms of characteristic sequences. The characteristic sequence of a set $S \in \mathcal{P}(\mathbb{N})$ is defined as follows, for all natural numbers n:[9]

$$\chi_n(S) \overset{\text{def}}{=} |S \cap \{n\}|.$$

This is a binary sequence that indicates whether or not n is in the set, respectively by value 1 or 0. Now we define the sequence of partial sums of the characteristic sequence of S, which is a non-decreasing integer sequence:

$$f_n(S) \overset{\text{def}}{=} \sum_{i=1}^{n} \chi_i(S).$$

In what follows, it will be helpful to also define the non-decreasing sequence of finite intersections of a given set, $S \subset \mathbb{N}$, with initial segments of $\mathbb{N}$ of length n:

$$S_n \overset{\text{def}}{=} (S \cap \{1, 2, 3, \ldots, n\}).$$

The sequence of finite sizes of S_n is equal to the aforementioned sequence of partial sums of the characteristic sequence: $f_n(S) = |S_n|$.

Using this notation, we may rewrite the natural density as:

$$d(S) = \lim_{n \to \infty} f_n(S)/n,$$

where $1/n$ is a normalization factor.[10]

[9] Throughout this article, we will indicate a sequence $a : \mathbb{N} \to X$ by its value at a generic position n, a_n. χ_n is a first example of this gloss.

[10] Observe that the natural density of a subset S can be regarded as the Césaro limit of its characteristic sequence:

$$C - \lim \chi_n(S) \overset{\text{def}}{=} \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \chi_i(S) = \lim_{n \to \infty} \frac{f_n(S)}{n} = d(S).$$

For subsets of $\mathbb{N}$, $\lim_{n\to\infty} f_n(S)$ suffices to indicate the cardinality: it results in the finite cardinality for finite sets and it diverges for sets of cardinality $\aleph_0$. Below, we will encounter the non-normalized sequence in other approaches as well.

The above sequences all take the canonical order of $\mathbb{N}$ for granted. Observe that, if we considered a sequence f'_n that maps n to the finite intersections in a different order, the resulting limit of $f'_n(S)/n$ might be different.

2.3.1 Examples of natural densities

The natural density is zero for all finite subsets and equal to one for all infinite co-finite subsets. Moreover, the natural density is also zero for some infinite co-infinite sets, such as the set of perfect squares, cubes, and higher powers, as well as the set of primes. So, the natural density coarse-grains finite sets and some infinite co-infinite sets. Sets with the same natural density can be distinguished in terms of the rate of convergence. For instance, the natural density of the primes is zero, but its rate of convergence is well-studied, too: the prime number theorem implies that the natural density of primes converges to zero as $1/ln(n)$ (see, e.g., Fine and Rosenberger, 2016, Ch. 4).

The complements of zero-density sets have natural density one. The natural density does distinguish between other infinite co-infinite sets, such as the even numbers, the numbers divisible by three, four, etc., and other arithmetic progressions (i.e., translations of the former). In general, for any set

$$\mathbb{M}_{a,i} \overset{\text{def}}{=} \{n \mod a = i \quad | \quad n \in \mathbb{N}\},$$

the natural density is:

$$d(\mathbb{M}_{a,i}) = 1/a.$$

Throughout the paper, we pay special attention to the case of the set of even numbers, $\mathbb{E} \overset{\text{def}}{=} \mathbb{M}_{2,0}$, and the set of odd numbers, $\mathbb{O} \overset{\text{def}}{=} \mathbb{M}_{2,1}$, both of which have natural density 1/2.

Like the cases with a natural density of zero or one, also all intermediate assignments are coarse-grained, in the sense that there are infinitely many other sets with the same natural density. To see this, it suffices to consider the countably infinite collection of sets that make a finite difference with the given set. Of course, taking the union or intersection of a given set with an infinite set of zero density also illustrates this, provided that the resulting set has a defined natural density. This is not guaranteed, since the sets for which d is defined do not form a σ-algebra citep[pp. 23–24]Kubilius:1964.

2.3.2 Subsets without a natural density

An example of a set that does not have a well-defined natural density is the set of natural numbers whose binary expansion has an odd length, for which the 'lower density' (corresponding to the lower limit in eq. 1) is 1/3 and the upper density (upper limit thereof) is 2/3. In general, a limit point of $f_n(S)/n$ is the limit of a converging subsequence (see, e.g. Tao, 2017), so the lower and upper density are the extremal values of the collection of limit points of $f_n(S)/n$. To illustrate that the interval spanned by the lower and upper density need not be symmetric around 1/2, as the previous example might suggest, consider the natural numbers that start with decimal 1, which has a lower density of 1/9 and an upper density of 5/9 (Tenenbaum, 2015, p. 415).

Moreover, it is possible to construct sets with lower density zero and upper density one. This can be achieved by alternating between increasingly long intervals of natural numbers that are excluded or included, such that there are subsequences of $f_n(S)/n$, indexed by the ends of either the excluded or the included intervals, that converge to zero or one, respectively.

Throughout this paper, we will focus on an example of this kind: the set $\mathbb{S}$, with super-exponential stretches. We define $\mathbb{S}$ recursively as follows: $\mathbb{S}_1 \overset{\text{def}}{=} \{\}$ and for all $n > 1$, $\mathbb{S}_n \overset{\text{def}}{=} \mathbb{S}_{n-1}$ if $k \overset{\text{def}}{=} \lceil \log_2(\log_2 n) \rceil$ is even and $\mathbb{S}_n \overset{\text{def}}{=} \mathbb{S}_{n-1} \cup \{n\}$ if k is odd.[11]

The extreme values of $f_n(\mathbb{S})/n$ occur at $n = 2^{(2^k)}$. If k is odd, we have:

$$f_n^{\mathrm{o}}(\mathbb{S})/n = \sum_{l=0}^{k} (-1)^{l+1} 2^{(2^l - 2^k)}.$$

Each term is smaller than the next and the signs alternate. The largest term, which occurs for $l = k$, equals $+1$; the second largest term, for $l = k - 1$, is $-\frac{1}{n^{1/2}}$; and the third largest, for $l = k - 2$, is $+\frac{1}{n^{3/4}}$. Hence we see that:

$$1 - \frac{1}{n^{1/2}} < f_n^{\mathrm{o}}(\mathbb{S})/n < 1 - \frac{1}{n^{1/2}} + \frac{1}{n^{3/4}}.$$

Since $\frac{1}{n^{1/2}}$ and $\frac{1}{n^{3/4}}$ tend to zero in the limit of n to infinity, this shows that the limit of $f_n(\mathbb{S})/n$ tends to 1 for the subsequence where $n = 2^{(2^k)}$ and k is odd.[12] So, the upper density of $\mathbb{S}$ is 1.

If $n = 2^{(2^k)}$ and k is even, we have:

$$f_n^{\mathrm{e}}(\mathbb{S})/n = \sum_{l=0}^{k-1} (-1)^{l+1} 2^{(2^l - 2^k)}.$$

There is one fewer term than in the odd case: this corresponds to the fact that now we are considering an initial fragment that ends on a segment of

[11] $\lceil \cdot \rceil$ indicates the ceiling function. An initial fragment of the set is given explicitly by $\{3, 4, 17, 18, 19, 20 \ldots\}$.

[12] Numerical values of $f_n^{\mathrm{o}}(\mathbb{S})$ at $n = 2^{(2^k)}$ with k odd can be found in the OEIS as the odd-numbered values of sequence A325912 (Manyama, 2019).

which the elements are not included. As before, each term is smaller than the next and the signs alternate. This time, the largest term, for $l = k - 1$, equals $\frac{1}{n^{1/2}}$, while the second largest, for $l = k - 2$, is $-\frac{1}{n^{3/4}}$. Hence, we obtain:

$$\frac{1}{n^{1/2}} - \frac{1}{n^{3/4}} < f_n^{\mathrm{e}}(\mathbb{S})/n < \frac{1}{n^{1/2}}.$$

This shows that the limit of $f_n(\mathbb{S})/n$ tends to 0 for the subsequence where $n = 2^{(2^k)}$ and k is even. So, the lower density of $\mathbb{S}$ is 0. Since it differs from the upper density, this proves that $\mathbb{S}$ has no natural density. We will return to this set below, in the context of numerosity theory.

Natural density is ideally suited for sets with a characteristic sequence that is periodic eventually: for those, it equals the relative fraction of included elements within the period. The natural density may be undefined for other sets, even if they are recursively enumerable (i.e., Turing-recognizable), as the above examples show.

2.3.3 Extension to $\mathbb{N}$ and relation to probability

As already mentioned, the domain of d does not exhaust $\mathcal{P}(\mathbb{N})$ and it is not an algebra.[13] However, it is possible to extend the natural density to a measure that is total on $\mathcal{P}(\mathbb{N})$ (as discussed, e.g., in Wenmackers and Horsten, 2013, §3.2). This is achieved by replacing the standard limit in the definition (eq. 1) by a generalised limit, which is a different algebra homeomorphism: a real-valued, free ultrafilter limit. For sets for which the standard natural density is undefined, this extension yields a non-unique result in the interval spanned by the lower and upper density (for details, see Schurz and Leitgeb, 2008). Its construction crucially relies on the Hahn–Banach theorem, which requires the Boolean prime ideal theorem (see, e.g., Schechter, 1997, Ch. 12). So, although the full Axiom of Choice is not required, this extension does require a non-constructive axiom, which is the source of its non-uniqueness.[14] For $\mathbb{S}$, this means that its generalized limit could be 0, 1, or some intermediate number, depending on the free ultrafilter used for the limit.

Remark that the underdetermination that shows up in the generalised density for infinite co-infinite sets such as $\mathbb{S}$ does not indicate a pathology of these sets. Also, there is nothing arbitrary or underdetermined about these subsets in themselves. What is arbitrary is to choose one limit point above all others, a choice that is achieved by a non-constructive object, such that all such assignments become arbitrary but fixed at once.

[13] Kerkvliet and Meester (2016) made a proposal to extend natural density uniquely to a larger subset of $\mathcal{P}(\mathbb{N})$. While we do not consider it here, for our purposes it is important that their extension is not total.

[14] The Hahn–Banach theorem is sufficient to prove the Banach–Tarski decomposition of a three-dimensional sphere, which is widely regarded as pathological (Pawlikowski, 1991).

Generalised density is related to the standard notion of probability (without countable additivity; see, e.g., Kubilius, 1964, p. 24): it is real-valued, non-negative, finitely additive, and normalized on $\mathbb{N}$. Indeed, it is an instance of the general notion of a limiting relative frequency, which has become a textbook definition of probability. As such, it can be considered as a finitely additive probability measure that is singleton-uniform on $\mathbb{N}$ (i.e., $d(F) = 0$ for all finite $F \subset \mathbb{N}$).

Like all other merely finitely additive probability measures on infinite domains, generalised density is an intangible. In section 2.4, we will encounter more intangibles: free ultrafilters and the α-numerosities that depend on them.

At this point, despite the explicit example of $\mathbb{S}$, perhaps it could be speculated that other sets that fail to have a natural density are random in some sense and thereby connected to intangibles. The next section shows that this speculation is misguided.

2.3.4 Almost all subsets of $\mathbb{N}$ are intangibles with density $1/2$

Despite the previous focus on extending the measure via a generalised limit that lacks uniqueness, almost all subsets of $\mathbb{N}$ do have a unique natural density. To see this, first observe that almost all binary sequences are algorithmically random (Martin-Löf, 1966), where 'almost all' means that they have measure one in the Lebesgue measure on the Cantor space $\{0, 1\}^{\mathbb{N}}$ (i.e., the fair Bernoulli measure on the set of infinite binary sequences).[15] Let us call subsets of $\mathbb{N}$ with a characteristic sequence that is algorithmically random 'random subsets' (cf. Axon, 2010, §3.1). Then, almost all subsets of $\mathbb{N}$ are random sets. Algorithmically random sequences are normal, which implies in the case of binary sequences that they have a limiting relative frequency of $1/2$. Hence, random sets have a natural density of $1/2$. This proves our initial statement: almost all subsets of $\mathbb{N}$ have a unique natural density, namely $1/2$.

When defining randomness in terms of *Kollektivs*, von Mises's (1928) guiding concept was precisely that no admissible place selection rule should pick out a subsequence with a lower or higher limiting relative frequency than $1/2$. If those existed, then it would be possible to bet on random outcomes with a higher than $1/2$ probability of winning, which goes against the pretheoretical concept of random outcomes. Place selection rules have later been replaced by more precise conditions on randomness, including algorithmic randomness (also known as Martin-Löf randomness; Martin-Löf, 1966).

[15] By this canonical measure, every element of $\mathcal{P}(\mathbb{N})$ has measure zero: it treats each subset of $\mathbb{N}$ on a par (unlike d, for instance), regardless of its properties, so it is not an alternative measure of size.

Looking at the definition of natural density (eq. 1), however, it may appear strange that random subsets have a well-defined natural density. How would one compute it without an algorithmic description of the set? Indeed, an explicit computation is not possible, because random subsets are intangibles: we can demonstrate that random subsets exist in ZFC, in abundance, without being able to give a single explicit example.[16] In this sense, the problem of determining the natural density of a fully specified random subset cannot arise. If we consider R to be a random set, then $\chi_n(R)$ is not computable, so the limit in equation 1 for the natural density cannot be computed explicitly either, yet it must be $1/2$ because $\chi_n(R)$ of a random set R is normal.

Moreover, it may appear strange that natural density is measured on a continuous scale ($[0, 1]$), yet almost all sets have the same value ($1/2$). This may be understood as follows. First consider a finite set, F, with a number of elements, $|F| = n$, that we assume to be even for simplicity. For each subset X of F, we can define its F-density as: $d_F(X) = |X|/n$, which takes values in $V = \{0, 1/n, \ldots, 1/2 - 1/n, 1/2, 1/2 + 1/n, \ldots, 1 - 1/n, 1\}$. Now consider the distribution that expresses how many subsets of F have an F-density equal to each of these values, which can be visualized as a histogram: see Figure 1 for some numerical examples. For each n, the distribution has a maximum that occurs at $d_F(X) = 1/2$, corresponding to the collection of subsets with $n/2$ elements. Their number is given by the central binomial coefficient, $n!/((n/2)!)^2$, while the total number of subsets of F is 2^n and the number of bins equals $|V| = n + 1$. So, their relative frequency evolves as $n!/((n/2)!)^2 \times (n + 1)/2^n$, which diverges in the limit of n to infinity. In the limit of n to infinity, the full distribution goes to a delta function with peak at density $1/2$ (cf. Figure 1). The limiting case shows once more that almost all subsets of $\mathbb{N}$, with infinitely long characteristic sequences, have a natural density of $1/2$. These include $\mathbb{E}$ and $\mathbb{O}$, but almost all others are random sets.

2.4 α-Numerosity

If we return to the initial trilemma, cardinality chose the Humean principle and natural density retained neither the Humean nor the Euclidean principle. So, we have yet to explore a theory that chooses the latter. α-Numerosity theory indeed takes the Euclidean principle as its starting point. α-Numerosities of subsets of $\mathbb{N}$ (canonically ordered) can be defined axiomatically as a function num from $\mathcal{P}(\mathbb{N})$ to some set of values $\mathcal{N} \supset \mathbb{N}$ on which the total order relation $<$ and the operation $+$ are defined, as follows:[17]

[16] This may be another reason for rejecting the notion of arbitrary subsets of $\mathbb{N}$, as Feferman (1999) did.

[17] This definition is a simplified version of the more general definition of α-numerosity on a class of labelled sets presented by Benci and Di Nasso (2019, Ch. 15); see also Benci

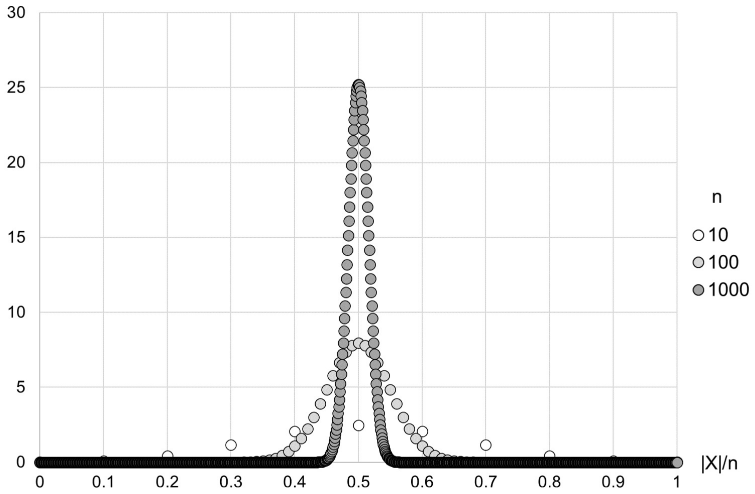


Figure 1: Histogram of the number of subsets of a set F with a given fraction of the total number of elements, $|F| = n$, for n equal to 10, 100, and 1000. This shows that the subsets, $X \subset F$, with $|X|/n = 1/2$ dominate in the limit of n to infinity.

Unit $\forall n \in \mathbb{N},\ \mathrm{num}(\{n\}) = 1$.

Additivity $\forall S, T \in \mathcal{P}(\mathbb{N})$, if $S \cap T = \emptyset$ then $\mathrm{num}(S \cup T) = \mathrm{num}(S) + \mathrm{num}(T)$.

Finite approximation $\forall S, T \in \mathcal{P}(\mathbb{N})$, if $f_n(S) \leq f_n(T)$ for all $n \in \mathbb{N}$ then $\mathrm{num}(S) \leq \mathrm{num}(T)$.

Euclidean principle $\forall S, T \in \mathcal{P}(\mathbb{N})$, if $S \subset T$ then $\mathrm{num}(S) < \mathrm{num}(T)$.

Agreement with finite cardinality follows from the first two axioms. They also ensure the Euclidean principle for finite subsets. The third axiom implies that the order on the numerosities of sets agrees with that of the finite cardinalities of their intersections with initial segments of $\mathbb{N}$; this ensures a certain harmony with natural density, as we will see below. The final axiom is needed to extend the Euclidean principle to infinite sets. In doing so, Hume's principle no longer holds, though a weaker statement applies to α-numerosities: if $\mathrm{num}(S) = \mathrm{num}(T)$, then there exists a bijection between S and T (so $|S| = |T|$). The inverse direction is not guaranteed; consider, for example, $S = \mathbb{N}$ and T any infinite strict subset. Then both have cardinality $\aleph_0$, yet $\mathrm{num}(S) > \mathrm{num}(T)$. So, α-numerosity is more fine-

and Di Nasso (2003) for an earlier version, as well as Benci et al. (2006) for a related approach, in which the Euclidean principle follows from the axioms.

grained than cardinality.[18] Cardinality is compatible with infinitely many finer grained notions of size. Although one might have hoped that the axioms for α-numerosity pick out one particular fine-graining, this turns out not to be the case, as we will see in section 2.4.2. Before showing that, we first examine a model of the theory.

2.4.1 The numerosity of $\mathbb{N}$ and the non-Archimedean α-limit

It follows from the axioms that num($\mathbb{N}$) cannot be finite and that num($\mathbb{N}$) is the largest numerosity value among all sets in $\mathcal{P}(\mathbb{N})$. It is customary to set

$$\mathrm{num}(\mathbb{N}) \overset{\mathrm{def}}{=} \alpha;$$

hence the name α-numerosities. A model for these axioms is obtained by setting the codomain $\mathcal{N}$ equal to an infinite hyperfinite set:

$$\mathcal{N}_\alpha \overset{\mathrm{def}}{=} \{0, 1, 2, \ldots, \alpha - 2, \alpha - 1, \alpha\}.$$

This is an infinite initial segment of a set of hypernatural numbers[19] extended with zero, ${}^*\mathbb{N} \cup \{0\}$, on which the relation $<$ (total order) and the operation $+$ are defined by transfer from the corresponding operations on $\mathbb{N}$. The values in $\mathcal{N}_\alpha$ suffice to assign α-numerosities to all elements of $\mathcal{P}(\mathbb{N})$. For each finite subset, the α-numerosity is a finite number and, for each co-finite set, it is α minus a finite number. The α-numerosity of an infinite co-infinite set is an infinite number that is infinitely smaller than α. For example, for the set of even numbers, the α-numerosity is usually taken to be $\alpha/2$ (although $\alpha/2 - 1$ is admissable, too, as we will discuss below).

α-Numerosity can be understood in terms of a non-Archimedean limit operation, called the α-limit (written as $\lim_{n\uparrow\alpha}$). The α-limit of a sequence can be thought of as follows: first, the sequence with indices in $\mathbb{N}$ is extended to a hypersequence with indices in ${}^*\mathbb{N}$; then, the extended sequence is evaluated at α. Using this notation, we can write (for any $S \in \mathcal{P}(\mathbb{N})$):

$$\mathrm{num}(S) = \lim_{n\uparrow\alpha} f_n(S).$$

We return to this in section 3.1.

[18] Lynch and Mackey (2023) recently proposed a similar formalism called 'magnum theory', which maps $\mathcal{P}(\mathbb{N})$ "to a subclass of the surreals, the surnatural numbers **Nn** (the non-negative omnific integers)". Their theory is also designed to adhere to the Euclidean principle and one way to define the magnum is via the natural density.

[19] A set of hypernatural numbers, ${}^*\mathbb{N}$, indicates a non-standard model of Peano arithmetic, first discovered by Skolem (1934).

2.4.2 α-Numerosity is totally ordered and highly non-unique

In the preamble to their axioms, Benci and Di Nasso (2003) assumed that the order on $\mathcal{N}$ is total; this indeed applies to $\langle \mathcal{N}_\alpha, < \rangle$. We will investigate the effects of taking $\mathcal{N}$ to be a merely partially ordered set in section 3.

So, α-numerosities are elements of a hyperfinite set that is totally ordered. They are constrained by the partial order of the subset relation on $\mathcal{P}(\mathbb{N})$ (via the final axiom: the Euclidean principle), as well as by the other axioms, but these are not sufficient to specify a total order. Any partial order can be extended to a total order (by the Szpilrajn extension theorem; also mentioned by Mancosu, 2009), but doing so requires a non-constructive choice principle (namely, the axiom of finite choice, weaker than the Axiom of Choice but also independent of ZF). For our case, this means that the axioms do not suffice to determine a total order on all of $\mathcal{P}(\mathbb{N})$ uniquely.[20]

To bridge the gap from a partial to a total order, the construction of α-numerosities crucially relies on a non-constructive object: a free ultrafilter on $\mathbb{N}$, which is an intangible (Schechter, 1997, §6.33). There are uncountably many free ultrafilters,[21] so this completion of the order is highly non-unique. In fact, using the reflections from section 2.3.4, we see that almost all members of a free ultrafilter on $\mathbb{N}$ are random subsets of $\mathbb{N}$ (which are intangibles in their own right). They have density 1/2, as do their complements, which are also random sets. Therefore, for almost every member of a free ultrafilter, it is indeed completely arbitrary to include it rather than its complement.

The axioms of Benci and Di Nasso (2003) require a specific type of free ultrafilter, called a *selective* ultrafilter, the existence of which is independent of ZFC, but we do not go into that here. For our purposes, it is merely important to note that this restriction does not lead to a unique assignment.

To illustrate the non-uniqueness, let us return to the infinite co-finite set $\mathbb{S}$ that was recursively defined in section 2.3.2. The α-numerosity of this set depends on the ultrafilter. In particular, the most extreme values are obtained when α is of the form $2^{(2^\kappa)}$. Here, κ is an infinite hypernatural, infinitely smaller than α. There are two ways in which this can happen: either the set $\{2^{(2^k)} \mid k \in 2\mathbb{N} - 1\}$ is in the ultrafilter, such that κ is odd, or $\{2^{(2^k)} \mid k \in 2\mathbb{N}\}$ is in the ultrafilter and κ is even. (Observe that these sets indicate the indices of the subsequences corresponding with respectively the upper and lower density of $\mathbb{S}$.)

If κ is odd, then num($\mathbb{S}$) is maximal:

$$\mathrm{num}(\mathbb{S}) = \lim_{n\uparrow\alpha} f_n(\mathbb{S}) = \lim_{n\uparrow\alpha} \sum_{l=0}^{k} (-1)^{l+1} 2^{(2^l)}.$$

[20] In the context of probability theory, a similar point has been discussed by Easwaran (2014) and Hofweber (2014).

[21] In fact, the set of free ultrafilters on $\mathbb{N}$ has the same cardinality as $\mathcal{P}(\mathcal{P}(\mathbb{N}))$; see Schechter (1997, §6.33) for references to proofs.

The result of this is a hyperfinite sum with $\alpha + 1$ terms.[22] It equals a particular value in $\mathcal{N}_\alpha$, but I am not aware of a closed formula to express it with.[23] Still, we can give an indication of the value by considering the three largest terms of the sum: α, $-\sqrt{\alpha}$, and $\sqrt{\sqrt{\alpha}}$. Hence,

$$\alpha - \sqrt{\alpha} < \mathrm{num}(\mathbb{S}) < \alpha - \sqrt{\alpha} + \sqrt{\sqrt{\alpha}}.$$

This shows that the largest possible α-numerosity that can be assigned to $\mathbb{S}$ is smaller than that of any co-finite set, yet bigger than some other infinite co-infinite sets, such as that of non-squares (which can be chosen to have numerosity $\alpha - \sqrt{\alpha}$).

If α is of the form $2^{(2^\kappa)}$ with even κ, then $\mathrm{num}(\mathbb{S})$ is minimal:

$$\mathrm{num}(\mathbb{S}) = \lim_{n \uparrow \alpha} f_n(\mathbb{S}) = \lim_{n \uparrow \alpha} \sum_{l=0}^{k-1} (-1)^{l+1} 2^{(2^l)}.$$

The result is a hyperfinite sum with α terms, the two largest of which are $\sqrt{\alpha}$ and $-\sqrt{\sqrt{\alpha}}$. Hence, in this case:

$$\sqrt{\alpha} - \sqrt{\sqrt{\alpha}} < \mathrm{num}(\mathbb{S}) < \sqrt{\alpha}.$$

This shows that the smallest possible α-numerosity that can be assigned to $\mathbb{S}$ is larger than that of any finite set, yet smaller than that of some infinite co-infinite sets, such as that of the perfect squares (which can be chosen to have numerosity $\sqrt{\alpha}$).

To recap, while the α-numerosity assignment to finite and co-finite sets is fully specified by the axioms, α-numerosities of infinite co-infinite sets depend on the properties of the free ultrafilter used to construct them. This may result in infinite differences in the α-numerosity of infinite co-infinite sets that do not have a natural density and finite differences for those that do (as shown in section 2.4.4).

2.4.3 Critical responses to α-numerosity

So far, we have seen that α-numerosity theory requires an intangible that cannot be proven to exist in ZFC. Moreover, if this object exists, it is highly non-unique and particular assignments crucially depend on its properties. These aspects play a major role in the critical reception of Euclidean theories of set size.

Gödel (1947) argued that Cantor's definition of cardinality is uniquely well-motivated and the only correct notion of set size. For decades, this was

[22] Hyperfinite sums generalize the standard notion of a finite sum to a sum with a hypernatural number of terms; for details see, e.g., Goldblatt (1998, §12.7).

[23] This is not due to the α-limit, but to the fact that a closed formula for $\sum_{l=0}^{k}(-1)^{l+1}2^{(2^l)}$ is unlikely to be found; in any case, Manyama (2019) reported none.

the received view that nearly stopped the search for alternative conceptions of set size in its tracks. However, Mancosu (2009) reviewed historical as well as contemporary theories on which sizes of subsets of natural numbers can be compared meaningfully, thereby contesting Gödel's view that Cantor's cardinality theory was inevitable. In particular, α-numerosity theory was developed by Benci and Di Nasso (2003).

Parker (2013) admitted that the numerosity theory of Benci and Di Nasso (2003) is logically consistent (provided that ZFC is), but he offered some new counterarguments. He argued that (total) Euclidean theories of set size are too arbitrary, because they violate weak invariance principles (such as translation or rotation invariance). This argument mainly relied on point sets in metric spaces, but for number sets (like the one under consideration here) Parker (2013, §5) similarly concluded that they are epistemically not very useful, because not all specific size assignments are uniquely well-motivated. As a result, Parker (2013) argued that numerosity assignments are uninformative, and even misleading: since at least some of them could just as well be different, they do not reveal a stable property intrinsic to those sets. This seems to violate a deeply rooted assumption about our size conception, perhaps its very essence. So, this objection is closely linked to the non-uniqueness of the free ultrafilter, and its consequences on α-numerosity assignments, which cannot even be fully specified because free ultrafilters are intangibles.

So, while the preamble stipulates that the order must be total, the axioms do not fully determine which infinite co-infinite sets are bigger than which. This depends on the particular selective ultrafilter.

The fact that selective ultrafilters are independent of ZFC should be flagged as a separate issue. Depending on one's philosophical views on mathematics, this may elicit different reactions, akin to those raised in response to the Continuum Hypothesis, which has a similar status.[24] Indeed, the existence of selective ultrafilters is implied by the Continuum Hypothesis (as also remarked by Parker, 2013, §6). Mathematical pluralists might take the independence of ZFC to mean that it is admissible to investigate the implications of α-numerosity theory, while it is equally admissible to explore the consequences of the negation of this possibility. Mathematical realists, on the other hand, might demand stronger, independently motivated axioms to settle the truth or falsehood of the existence of selective ultrafilters and thus of the (in)correctness of the α-numerosity approach to set sizes. So, it is unclear what a realist in the sense of Gödel (1947) should conclude: on the one hand, they view Cantor's cardinality as the ultimate theory of set size; yet, if they complete Gödel's program and find independently justified axioms that settle the Continuum Hypothesis—and if it is found true—that implies the type of object that fuels an alternative theory of set size.

[24] I am grateful to a reviewer, who suggested making this analogy.

In section 3, we discuss a related approach by Trlifajová (2024) based on the Fréchet filter that may escape these criticisms.

2.4.4 Relation to natural density and additional constraints

As mentioned, the third axiom of α-numerosity theory guarantees a certain agreement with natural density. In particular, for any set S such that $d(S)$ is defined, it holds that $d(S) = st(num(S)/\alpha)$, where $num(S)/\alpha$ is a hyperrational number (i.e., an element of ${}^*\mathbb{Q}$) and st is the standard part function that maps a finite hyperrational number to the unique nearest standard real number (for details, see Benci and Di Nasso, 2019, §16.9). This implies that sets which have the same natural density, cannot have numerosities that differ by more than a finite number.[25] In particular, considering random subsets, there are infinitely more subsets of $\mathbb{N}$ with an α-numerosity within a finite interval around $\alpha/2$ than anywhere else in the hyperfinite set $\mathcal{N}_\alpha$.

In addition to the axioms, some further size comparisons can be motivated and achieved by restricting the ultrafilter accordingly. For instance, the axioms do not fix whether various sets $\mathbb{M}_{a,i}$ for given a have exactly the same α-numerosity; they could also differ by a finite amount. However, one may further stipulate (cf. Benci and Di Nasso, 2003, pp. 63–64):

$$\begin{aligned} \forall a \in \mathbb{N}, \forall i \in \{0, \ldots, a-1\} \quad & \mathrm{num}(\mathbb{M}_{a,i}) = \alpha/a; \\ \forall p \in \mathbb{N} \quad & \mathrm{num}(\{n^p \mid n \in \mathbb{N}\}) = \sqrt[p]{\alpha}. \end{aligned}$$

We could add further stipulations to fix, for instance, the α-numerosity of the set of natural numbers whose binary expansion has an odd length. Since this set does not have a natural density (recall that its lower and upper natural density were resp. 1/3 and 2/3), its α-numerosity may differ by an infinite amount, depending on the ultrafilter. Still, uncountably many α-numerosity assignments are left to the specific properties of the (arbitrary but fixed) free ultrafilter.

We may call $num(S)/\alpha$ the α-density of S. Like natural density, α-density can be given a probabilistic interpretation. Next, we comment on a related formalism.

2.4.5 Relation to probability

α-Numerosity theory is closely related to non-Archimedean probability (NAP) theory, as discussed by Benci et al. (2013, §5) and Benci et al. (2018, §3.6). The non-uniqueness of NAP functions has been discussed by Benci et al. (2018, §6.1). For the present purposes, we limit ourselves to the case of a NAP function that represents a singleton-uniform probability measure on $\mathbb{N}$,

[25] This can be seen as follows. For any sets S and T such that $d(S) = d(T)$, it holds that $st(num(S)/\alpha) = st(num(T)/\alpha)$, hence $st(num(S)/num(T)) = 1$. If their ratio is infinitely close to 1, the numerosities cannot differ by more than a finite number.

which is equal to a α-numerosity assignment up to a normalization factor $1/\alpha$. So, this leads to normalized measure of sizes of all elements of $\mathcal{P}(\mathbb{N})$.

The non-uniqueness of NAP assignments has been discussed in detail for the subset of even numbers, $\mathbb{E}$, and that of odd numbers, $\mathbb{O}$, by Benci et al. (2013, pp. 141–142). In terms of α-numerosity, this boils down to two possible assignments: either $\text{num}(\mathbb{E}) = \alpha/2 = \text{num}(\mathbb{O})$, or $\text{num}(\mathbb{E}) = \alpha/2 - 1 < \text{num}(\mathbb{O}) = \alpha/2 + 1$. This difference depends on whether $\mathbb{E}$ is in the free ultrafilter used to generate the hypernatural numbers, or $\mathbb{O}$ is; or, equivalently, whether α is even or odd (Benci and Di Nasso, 2003, p. 63). The possibility of ending up with $\text{num}(\mathbb{E}) < \text{num}(\mathbb{O})$, with a difference of 1, can be understood by considering that we defined $\mathbb{N}$ as starting from 1, so the odd numbers are one step ahead at the odd positions in the canonical order on $\mathbb{N}$; this constant difference remains visible in the non-Archimedean limit at α of $f_n(\mathbb{E})$ and $f_n(\mathbb{O})$ if α is odd. Similar considerations hold for other sets of the form $\mathbb{M}_{a,i}$. We return to this at the end of section 2.5.

2.5 Composite picture

It is time to take stock, by giving an overview of the formalisms reviewed so far. To compare various size assignments across the formalisms and their relation to the subset lattice on $\mathcal{P}(\mathbb{N})$, they are depicted in Figure 2. Since α-numerosities are partially underdetermined by the axioms, this picture shows just one possibility among many: one where α is a multiple of two, three, four, etc., as well as a power of two, three, etc. We have not included $\mathbb{S}$ in the figure, but observe that it could be placed nearly anywhere: from a point between the sets of cubes and squares until a point between the sets of non-squares and non-cubes. We have also not included random sets: they would be at the levels of $\mathbb{E}$ and its finite difference sets.

Let us now summarize the most relevant properties of the formalisms reviewed in this section in Table 1. Of these alternatives, α-numerosity provides the most fine-grained measure of the sizes of subsets of natural numbers: it is total on $\mathcal{P}(\mathbb{N})$, is maximally fine-grained on finite and co-finite subsets, and not only distinguishes among infinite co-finite sets but makes similarly fine-grained distinctions among them as it does for (co-)finite ones. However, unlike the other alternatives (except for generalised density), α-numerosity is not uniquely determined by its axioms. The Euclidean principle requires that α-numerosity assignments respect the ordering due to the subset relation on $\langle \mathcal{P}(\mathbb{N}), \subset \rangle$. Although the third axiom restricts the possible assignments further (by demanding a form of correspondence with natural density), this still only specifies a strict partial order, while α-numerosity values are part of the totally ordered set of hypernatural numbers, $\langle {}^*\mathbb{N}, \subset \rangle$. As before with generalised density, totality is provided by an intangible.

Table 1: Comparison of four formalisms.

	Cardinality	Infinite lottery logic	α-Numerosity	Natural density
Total on $\mathcal{P}(\mathbb{N})$	Yes	Yes	Yes	No, but can be extended
Unique assignments	Yes	Yes	No	Yes, but not the extension
Normalized (size($\mathbb{N}$) = 1)	No, not normalizable	No, not normalizable	No, but normalizable	Yes
Distinguishes between:				
Finite sets	Yes	Yes	Yes	No
Co-finite sets	No	Yes	Yes	No
Infinite co-infinite sets	No	No	Yes	Yes

As reviewed in section 2.4.3, in response to the results of α-theory, Parker (2013) concluded that (total) Euclidean assignments of set size contain arbitrary aspects that are therefore misleading. He did see some hope for a different Euclidean theory of size limited to partial assignments (e.g., "**Even** might be regarded as neither smaller than **Odd**, nor larger, nor equal" Parker, 2013, p. 609), but we will see in section 3 that partially ordered numerosity can do better than this (e.g., it excludes the possibility that $\mathbb{E}$ might be larger than $\mathbb{O}$ in a non-arbitrary way).

A similar tension between totality and arbitrariness has been discussed in the context of probability theory by Easwaran (2014). As summarized by Benci et al. (2018, p. 543): "Easwaran observes that real-valued probability functions leave out part of the structure of the partial order, whereas hyperreal-valued probability functions add structure in an arbitrary way." Likewise, when assigning sizes to subsets of $\mathbb{N}$, it seems that we are forced to choose between options with too little or too much structure: whereas cardinality, infinite lottery logic, and natural density do not preserve the partial order of the subset lattice, α-numerosity comes with excess structure added by the specifics of some free ultrafilter.

Easwaran (2014) suggested using standard probability functions (which miss some structure on the partial order of the events) together with the subset structure of the algebra of events: in the case of equal probabilities, one may still treat the larger set as more probable (e.g., when deciding betting preferences). He argued that this is better than accepting the arbitrary excess structure baked into a non-standard probability function. If we apply Easwaran's proposal to our current case, we should reject α-numerosity in favour of natural density, while keeping an eye on the subset relation when ranking set sizes. For example, $d(\mathbb{E} \setminus \{2\}) = d(\mathbb{E})$ but $\mathbb{E} \setminus \{2\} \subset \mathbb{E}$, so we may treat the former as 'smaller', in some sense. However, this proposal does not seem sufficient, since $d(\{1\}) = d(\{2,3\})$, but $\{1\}$ and $\{2,3\}$ are not comparable by the subset relation, so they are still 'equal', in the same sense of size.

An alternative approach has been suggested, but not developed, for non-Archimedean probabilities; Benci et al. (2018, p. 544): "one can always consider the entire family of NAP functions modelling a given situation, rather than an arbitrary representative of it (see also Wenmackers and Horsten [2013]). [...] As a whole, the family shows us how much the probabilities of a given event, and the order of probabilities of multiple events, can vary (dependent on the choice of ultrafilter)." This proposal is related to imprecise or interval-based probabilities. Likewise, we could consider the collection of α-numerosity functions ranging over all selective ultrafilters.

In the next section, we review a very recent approach that manages to trace the contours of such a collection, without listing its members.

3 Partially ordered c-numerosity of subsets of $\mathbb{N}$

It is clear that all α-numerosity assignments agree on finite and co-finite subsets of $\mathbb{N}$, and that they differ amongst each other by at least 1 on infinite co-infinite sets. Now, one may wonder about the collection of subsets that all free ultrafilters have in common. This is the co-finite filter or Fréchet filter on $\mathbb{N}$:

$$\mathcal{F} \overset{\text{def}}{=} \{S \subseteq \mathbb{N} \mid \mathbb{N} \setminus S \text{ is finite}\}.$$

So, $\mathcal{F}$ consists of all co-finite sets of $\mathbb{N}$; it is free but not an ultrafilter itself (see, e.g., Jech, 2003, p. 73). It can be defined in ZF, so it does not require any non-constructive choice principle. This suggests the possibility of developing a fully (quasi-)constructive theory akin to α-numerosity.[26]

Recently, Trlifajová (2024) has made a reconstruction of Bolzano's work and indeed arrived at such a formalism based on the co-finite filter rather than a free ultrafilter. Like α-numerosity theory, Trlifajová (2024)'s theory applies to all countable sets, but we limit our review to its results for subsets of $\mathbb{N}$. Again, elements of such sets are only considered in their canonical order.

Trlifajová (2024) reconstructed the work of Bolzano as being about sequences of the form $f_n(S)$ (i.e., $|S_n|$; $\sigma(S)$ in her notation), where S is a subset of natural numbers. First, she showed that the set $\{f_n(S) \mid S \in \mathcal{P}(\mathbb{N})\}$

[26] The potential of developing non-standard analysis using the constructive co-finite filter instead of free ultrafilters has also been considered, for instance, by Palmgren (1998).

exhausts the set of non-decreasing sequences of natural numbers, which we call $\mathcal{S}$. Then, she equipped this set with addition (and multiplication, which we do not consider here) defined component-wise, and with equality and ordering defined as equality or ordering (resp.) of terms eventually.

The notion of equality or ordering of terms 'eventually' means the relation holds between the terms for a co-finite subset of indices of the sequences; *i.e.*, such that the set of indices is an element of the co-finite filter $\mathcal{F}$. This can be viewed as a type of limit operation, along the co-finite filter. We will call it the *$\mathcal{F}$-limit* ($\lim_{\mathcal{F}}$).

With natural density, different sequences $f_n(S)/n$ may have the same limit: these sequences form an equivalence class under the standard limit operation. Likewise, what really matters for size attributions in the sense of c-numerosity is not an individual size sequence function, $f_n(S)$, either, but rather its equivalence class under the Fréchet filter. Hence, it is convenient to quotient the Fréchet filter out of the space of non-decreasing sequences: $\mathcal{S}/\mathcal{F}$.[27] With addition, equality and order defined as before, Trlifajová (2024, p.96) showed $\mathcal{S}/\mathcal{F}$ to be a "partially ordered non-Archimedean commutative semiring". This means that it contains infinite elements (as before) and retains the same arithmetical properties as $\langle \mathbb{N}, +, \times \rangle$ (which is a totally ordered Archimedean commutative semiring).

Now, equivalence classes of f_n are elements of $\mathcal{S}/\mathcal{F}$. They play a role very similar to the α-numerosity function, num, but with the co-finite filter, rather than a free (and selective) ultrafilter. Therefore, we will call this the *c-numerosity*, where the 'c' stands for co-finite as well as for constructive:

$$\forall S \in \mathcal{P}(\mathbb{N}), \quad \text{c-num}(S) \overset{\text{def}}{=} \lim_{\mathcal{F}} f_n(S) \overset{\text{def}}{=} [f(S)]_{\mathcal{F}} \in \mathcal{S}/\mathcal{F}.$$

Another precursor to this approach is found in Peano (1910):[28] he proposed to associate with each sequence an 'end' value ('*fine*' in Italian), which can be understood as a type of limit different from the standard one. In general, Peano (1910, p. 780) took two sequences to have the same end value if they are equal from a certain index onward. For a sequence that is constant eventually, its end value is defined to be equal to that constant. (Hence, all real values are among the ends.) For sequences such as $a_n = 1/n$, however, Peano argued that their ends are actual infinitesimals and, for sequences such as $a_n = n$, actual infinities. He defined the sum and product of end values as the end value of the sum or product of the corresponding sequences. Peano (1910) did not use his construction to discuss set sizes, but applying it to sequences of the form $f_n(S)$ yields results equivalent to those of

[27] See Trlifajová (2024, p. 112; emphasis added): "The codomain of σ is the set of non-decreasing sequences of natural numbers *modulo the Fréchet filter* which is just partially and not linearly ordered."

[28] I am grateful to a referee for providing me with a copy of this source.

Trlifajová (2024).[29]

Observe that c-numerosities *also* provide a model for the simplified axioms for α-numerosities that we presented in section 2.4, provided that we drop the requirement from the preamble that $\mathcal{N}$ is totally ordered. This amounts to setting the codomain of the new numerosity function to $\mathcal{N}_c \overset{\text{def}}{=} \mathcal{S}/\mathcal{F}$. In section section 2.4, we used a totally ordered hyperfinite set as the codomain, $\mathcal{N}_\alpha = \{0, 1, 2, \ldots, \alpha - 2, \alpha - 1, \alpha\}$, which results in the core difference with merely partially ordered c-numerosities covered here.

C-numerosities agree with α-numerosities on finite sets. Choosing

$$\text{c-num}(\mathbb{N}) \overset{\text{def}}{=} \alpha,$$

the assignments also agree on co-finite sets. In other words, if we restrict $\mathcal{N}_c$ to the values assigned to finite and cofinite sets, that subset is totally ordered. So, what about infinite co-infinite sets? In general, their c-numerosity (in $\mathcal{N}_c$) cannot be mapped to a unique value in $\mathcal{N}_\alpha$. Returning to the example of even and odd numbers, Trlifajová (2024, p. 96) wrote: "While there are one fewer or as many even numbers as odd numbers, their sum is equal to α." Adapted to our notation, she obtained:

$$\text{c-num}(\mathbb{E}) \leq \text{c-num}(\mathbb{O}) \leq \text{c-num}(\mathbb{E}) + 1;$$

$$\text{c-num}(\mathbb{O}) + \text{c-num}(\mathbb{E}) = \alpha.$$

Although $\mathbb{E}$ and $\mathbb{O}$ are comparable to each other on this approach, neither of these sets is comparable to the infinite co-infinite set $\mathbb{S}$ that was recursively defined in section 2.3.2. This example shows that c-numerosities are only partially ordered. Before we elaborate on this example, it is instructive to examine the limit operations underlying the various approaches.

3.1 Comparing the limit operations

Any limit operation on sequences requires an equivalence relation that consists of two aspects: a notion of tolerance and a notion of qualified index sets, given by a filter.[30] For the standard limit operation on sequences, $\lim_{n\to\infty}$, the qualified index sets are those with a sufficiently large index and the tolerance is given by arbitrarily small differences along those indices (e.g., $|a_n - a_{n-1}| < 1/n$ 'eventually', i.e., for all $n > N$ for some N). In this case, the qualified index sets are given by the co-finite sets in $\mathbb{N}$. The family of such sets forms a filter: the Fréchet filter, $\mathcal{F}$, on $\mathbb{N}$.

[29] To my knowledge, Peano's (1910) paper has not been translated. Bottazzi and Katz (2021, pp. 477–478) described the gist of it, rephrased in contemporary terms, as the construction of a "partially ordered non-Archimedean ring" consisting of the set of ends of real-valued sequences "with zero divisors that extend $\mathbb{R}$".

[30] For a helpful overview, see Tao (2017). For a comparison that focuses on the difference between the standard limit and the α-limit, see Wenmackers (2019, §6.1).

A non-Archimedean limit operation, such as the α-limit, differs from the standard limit on both accounts: the qualified index sets are given by a different type of filter, a free ultrafilter on $\mathbb{N}$ (on which additional conditions may be imposed), and the tolerance is given by exact equality (zero tolerance).

This presentation suggests two additional types of limit operations: one that combines a free ultrafilter with arbitrarily small differences, which was the generalised ultrafilter limit we used to extend the natural density to $\mathcal{P}(\mathbb{N})$ in section 2.3.3, and another one that combines the Fréchet filter with zero tolerance, which was $\lim_{\mathcal{F}}$ introduced in section 3.

Hence, natural density, generalised density, α-numerosity, and c-numerosity can all be understood as a type of limit operation applied to $f_n(S)$ or $f_n(S)/n$: Table 2 gives an overview.[31]

Table 2: Different formalisms in terms of the type of limit operation (rows) and sequence (columns).

tolerance	filter		f_n	f_n/n
$< \epsilon(n)$	$\mathcal{F}$	standard limit	finite cardinality or ∞	natural density
	$\mathcal{U}_\alpha$	generalised limit		generalised density
$= 0$	$\mathcal{F}$	$\mathcal{F}$-limit	c-numerosity	c-density
	$\mathcal{U}_\alpha$	α-limit	α-numerosity	α-density

We could also introduce mirror cardinalities by considering $f_n(\mathbb{N} \setminus S)$ as well. Moreover, observe that finite and co-finite sets give rise to sequences that are *extremely* well-behaved in the standard limit. As remarked in footnote 10, the natural density of a set can be regarded as the Césaro limit of its characteristic function. But for any finite set, F, $\lim_{n\to\infty} \chi_n(F) = 0$ and $\lim_{n\to\infty} \chi_n(\mathbb{N} \setminus F) = 1$, so the additional averaging provided by the Césaro limit is not needed to get convergence in this case.

Table 2 helps us to understand the relations of fine-grainedness between the measures, as well as the source of partial underdetermination present in some of them. The $\mathcal{F}$-limit of sequences is more fine-grained than the standard limit, because it requires exact equality of terms rather than equality up to an arbitrarily small difference (while using the same notion of eventuality); and it is more coarse-grained than the α-limit because it uses a smaller filter (while using the same notion of equivalence). In Figure 3, we depicted the partial order on the formalisms induced by their fine-grainedness.

Both the standard limit and the $\mathcal{F}$-limit are constructive, the other two are not. Hence, natural density and c-numerosity are free of arbitrary structure introduced by intangibles, whereas both generalised density and α-numerosity show signs of them.

[31] Recall from section 2.3 that the standard limit of $f_n(S)$ *is* defined for all sets S, even when that of $f_n(S)/n$ is not.

3.2 Another look at set $\mathbb{S}$

Now that we have a better grasp of the connection between the different formalisms via the underlying limit operations, let us illustrate it with the example the set $\mathbb{S}$ (as defined in section 2.3.2).

While the standard limit of $f_n(\mathbb{S})$ is well-defined ($+\infty$), that of $f_n(\mathbb{S})/n$ is not. In other words, the natural density of $\mathbb{S}$ is undefined. As we discussed in section 2.3.2, the reason is that $f_n(\mathbb{S})/n$ has different subsequences, each with a different limit (ranging from 0 to 1). The natural density is expressed by real numbers, which are totally ordered, and the measure does not depend on arbitrary structure, but this is only possible by excluding sets like $\mathbb{S}$ from its domain.

A generalised density does assign a limit-value to $f_n(\mathbb{S})/n$, equal to that of some arbitrary but fixed subsequence. So, this measure is total on $\mathcal{P}(\mathbb{N})$ and assigns values from a totally ordered range, but with a partially arbitrary character.

Likewise, as we have seen in section 2.4.2, the α-numerosity of $\mathbb{S}$, defined as the α-limit of $f_n(\mathbb{S})$, can be any of a wide range of infinite numbers (depending on which subsequence is selected by the underlying ultrafilter). The fact that all admissible α-numerosities are infinite corresponds with the divergence of the standard limit. The fact that they differ by more than a finite amount reflects the lack of a unique natural density. In particular, the α-limit of $f_n(\mathbb{S})/n$ varies from an infinitesimal less than $1/\sqrt{\alpha}$ (with standard part 0) to more than $1 - 1/\sqrt{\alpha}$, which is infinitesimally close to 1 (i.e., has standard part 1). These two assignments correspond to the two subsequences we considered before to argue that $f_n(\mathbb{S})/n$ has subsequences with limits of 0 and 1. So, α-numerosity is similar to general density in terms of totality, arbitrary, and lack of uniqueness. Moreover, its range is non-Archimedean to allow for the Euclidean principle.

Finally, the c-numerosity of $\mathbb{S}$ is comparable to some sets (for instance, it is larger than any finite set and also larger than the set of cubes), but not to others (such as the set of squares, $\mathbb{E}$, $\mathbb{O}$—or any of the $\mathbb{M}_{a,i}$, for that matter). This measure is also non-Archimedean, to allow for the Euclidean principle. This time, the measure is unique and total on $\mathcal{P}(\mathbb{N})$ but now the order is partial, again to allow for sets like $\mathbb{S}$.

Recall that, if one accepts infinite co-infinite sets as unproblematic in general, there is nothing intrinsically pathological or unknowable about $\mathbb{S}$. It is not an intangible itself; its characteristic function just keeps alternating over stretches that grow at a super-exponential rate. As a result, its growth rate ($f_n(\mathbb{S})$) does not stabilize relative to that of initial fragments of $\mathbb{N}$ (of length n).

Most of the formalisms that we reviewed are dealing with this fact by leaving something undetermined: either by not assigning a value to some sets at all or by not totally ordering the values (natural density and c-numerosity, respectively), or by assigning them totally ordered values in some globally consistent but arbitrary way (general density and α-numerosity). In all these cases, something has to go: totality (whether totality on the domain or total ordering of the co-domain) or uniqueness of assignment, respectively. The cardinality-based approaches do not face this dilemma, but they also leave unexpressed much non-arbitrary structure that the other theories agree upon regarding the sizes of infinite co-finite sets.

3.3 The relative strength of c-numerosity

Since the c-numerosity approach does not invoke intangibles, it escapes the criticism that Parker (2013) directed at total Euclidean theories (reviewed in section 2.4.3). However, Parker (2013, §9) also anticipated Euclidean theories with "partial size assignments". While total theories were too arbitrary on his view, he expected partial theories to be too weak and narrow: unable to decide the relative sizes of some simple sets. In particular, Parker (2013, p. 609) expected that such theories would not make any comparison between the size of the sets of even and odd numbers, $\mathbb{E}$ and $\mathbb{O}$, ranking them as "neither smaller [...], nor larger, nor equal". However, we have seen that c-numerosity is a partial Euclidean theory that does exclude that $\mathbb{O}$ could be larger than $\mathbb{E}$. It constrains the order as follows: $\text{c-num}(\mathbb{E}) \leq \text{c-num}(\mathbb{O}) \leq \text{c-num}(\mathbb{E}) + 1$ with $\text{c-num}(\mathbb{E}) + \text{c-num}(\mathbb{O}) = \alpha$. So, c-numerosity is much less weak than anticipated by the criticism of Parker (2013, §9). (See also the response in Trlifajová, 2024, §6.)

The axioms of numerosity theory merely specify a partial order. For numerosity values to be totally ordered, this has to be postulated separately, as Benci and Di Nasso (2003) did. Doing so comes at the cost of introducing arbitrary and unknowable structure due to an intangible. C-numerosity theory, in contrast, honours the Euclidean principle and the other axioms of numerosity theory, while remaining fully constructive. It is as fine-grained as possible and as coarse-grained as needed. So, c-numerosity seems to hit the right balance between expressiveness and leaving out comparisons that are not stable. This leads to a merely partially ordered notion of size, which means that its assignments are further removed from the usual notion of number. Moreover, the inequalities of c-numerosity assignments exactly trace the hull of all possible α-numerosity assignments. We explore this departure from our familiar size concept and the relation between the two theories in the next section.

3.4 Epistemicism or supervaluation

Taking our inspiration from formal methods for dealing with vagueness, we may consider a particular α-numerosity function (generated by a specific free ultrafilter) as a 'precisification' of c-numerosity.[32] Likewise, we may consider each generalised density as a precisifation of the natural density. We could also say that a total extension of c-numerosity (and natural density) is multiply realizable.

In analogy to epistemicists about vagueness (such as Williamson, 1994), we might assume that there is one true Euclidean notion of size—given by one correct precisification, but which is unknowable to us. In this case, it is not the Euclidean notion of size itself but our knowledge of it that is incomplete. In principle, such a position should be acceptable to a mathematical realist. On the one hand, the fact that α-numerosity functions—like the free ultrafilters that generate them—are intangible is congenial to this interpretation: the unknowability is guaranteed. Yet, it seems hard to defend that any specific one could be privileged, since these objects do indeed represent much arbitrary structure.

Alternatively, we may view the Euclidean notion of size embodied in the axioms reviewed at the start of section 2.4 as partially underdetermined. Then, like supervaluationists, we may treat all precisifications on a par and consider the set of such precisifications. In the study of vagueness, the notions of 'supertrue' and 'superfalse' are defined as truth values that agree on the set of all precisifications. Analogously, here, we may consider subsets of $\mathbb{N}$ that have a 'super-α-numerosity' as those that have the same α-numerosity on each precisification:[33] these are exactly the ones that have a unique c-numerosity, i.e., just the finite and co-finite sets.

One complication, mentioned before, is that we may *define* the α-numerosity function to assign $1/a$ to each subset $\mathbb{M}_{a,0}$ etc.; so, we have to evaluate the issue relative to such potential additional constraints at hand. Observe that, although Trlifajová (2024) did not suggest this (and, in fact, argued against doing so), additional constraints may be added to partially ordered numerosities, too. (After all, adding, e.g., $\mathbb{E}$ to $\mathcal{F}$ and extending it such that it forms a filter, does not turn it into a free ultrafilter.)

All infinite co-infinite sets fail to have a super-α-numerosity. We may wish to distinguish cases where the α-numerosities of different precisifications only differ by a finite amount from those where they differ by a larger margin; the former are the sets with a natural density. Mutatis mutandis, we may also consider a dual approach with subvaluations: a sub-α-numerosity is an α-numerosity that is assigned by at least one precisification.

Admittedly, the terms 'underdetermination', 'precisification', and 'sub-

[32] Both α- and c-numerosity quotient a filter out of the space of non-decreasing sequences: the Fréchet filter used in c-numerosity is contained in any free ultrafilter used in α-numerosity, so the former assignments are compatible with the former, but not vice versa.

[33] Equivalently, we may say that it is supertrue that a certain set has a particular α-numerosity. This suggestion has been made in the context of NAP functions, by Benci et al. (2018, p. 544).

valuation' all take total measures with totally ordered values as the standard. This is indeed a standard in the sense that earlier theories that assign sizes to subsets of $\mathbb{N}$ adhere to it. Moreover, according to Forti (2022), general comparability is an essential property of the classical notion of magnitude or size.

Forti (2022) motivated the assumption that Euclidean set sizes are totally ordered as follows: "Following the ancient praxis of comparing magnitudes of homogeneous objects, a very general notion of size of sets, whose essential property is general comparability of sizes, can be given through a total preordering $\preceq$ of sets according to their sizes [. . .]" Here, 'homogeneous' means that the objects are of the same kind, with the same dimensions, though this condition was relaxed by later mathematicians. For instance, the cardinality of $\mathbb{N}^2$ can be compared to that of $\mathbb{N}$ (and they are equal). Although this concerns a preorder (or quasiorder: i.e., a binary relation that is reflexive and transitive), rather than an order (which is additionally required to be antisymmetric), it is a *total* preorder, on which any two elements are comparable (as Forti, 2022, specified in his footnote 3).

Nevertheless, once this classical assumption has been made explicit, it becomes possible to question and perhaps abandon it. A possible lesson that we could draw from studying numerosity theories is exactly that Euclidean set sizes are merely partially ordered. Our review shows that the notion of size has already undergone drastic changes; this may be the next, comparatively conservative, extension of it. Reluctance to accept the possibility of a merely partially ordered notion of set size may also resemble the epistemicist approach to vagueness: by positing crisp thresholds, it replaces vagueness by unknowable and non-constructive elements (cf. Sorensen, 2022, §3). In both cases, this trade is possible, but by no means necessary.

3.5 A final look at the size of $\mathbb{E}$ versus $\mathbb{O}$

It may be instructive to return to our running example of $\mathbb{E}$ versus $\mathbb{O}$ one final time. Let us consider their even and numbered subsequences of f_n/n:

$$\begin{aligned} f_{n\in\mathbb{E}}(\mathbb{E}) &= 1/2; & f_{n\in\mathbb{E}}(\mathbb{O}) &= 1/2; \\ f_{n\in\mathbb{O}}(\mathbb{E}) &= 1/2 - 1/(2n); & f_{n\in\mathbb{O}}(\mathbb{O}) &= 1/2 + 1/(2n). \end{aligned}$$

These sequences have the same standard limit point, 1/2; hence, the natural density of $\mathbb{E}$ and $\mathbb{O}$ is equal. But $f_{n\in\mathbb{O}}(\mathbb{E}) < f_{n\in\mathbb{O}}(\mathbb{O})$ for all n, so $f_n(\mathbb{E})/n$ has a strictly smaller 'α-limit point' than $f_n(\mathbb{O})/n$ if and only if $\mathbb{O}$ is in the ultrafilter. Hence, the α-density of $\mathbb{E}$ may be smaller than that of $\mathbb{O}$, but since both $\mathbb{E}$ and $\mathbb{O}$ have natural density 1/2, it seems arbitrary which one is in the ultrafilter to decide this matter. We might as well choose neither, as in the Fréchet filter, which yields the inequalities of the c-density. Since the natural density for $\mathbb{E}$ and $\mathbb{O}$ is defined, the generalised limit must equal this value, too.

Although the generalised limit involves a free ultrafilter, and thus the result may in general depend on whether or not $\mathbb{O}$ is included, the difference is glossed over in this case due to the tolerance built into this limit operation. Only for sets S such that $f_n(S)/n$ has multiple standard limit points, which thus lack a natural density, the tolerance built into the generalised limit is not enough to absorb these differences.

Using the terminology of section 3.4, we may say that $1/2$ and $1/2(1 - 1/(2\alpha - 1))$ are both sub-α-numerosities of $\mathbb{E}$ while the set lacks a super-α-numerosity. Put differently, it is subtrue that the set of even numbers has the same numerosity as the set of odd numbers, but this equality is not supertrue.

We have to admit that debating the α-numerosity values of infinite co-infinite sets may skirt dangerously close to discussing how many angels can dance on the head of a pin. Mancosu (2009) quoted Descartes (*Principes de la Philosophie*, I.26) on the issue at hand: "We will not bother to reply to those who ask if the infinite number is even or odd or similar things since it is only those who deem that their mind is infinite who seem to have to tackle such difficulties." Using sophisticated tools, it seems that the answer 'it depends' can be refined to 'it depends on which ultrafilter you use to investigate the matter,' but these answers may reveal more about the tools that were used (intangibles from $\mathcal{P}((\mathcal{P}(\mathbb{N})))$ than about the objects under study (elements of $\mathcal{P}(\mathbb{N})$) themselves. In contrast, c-numerosity refuses in a Cartesian way to answer whether the size of $\mathbb{N}$ is even or odd, or any other questions that go beyond the restrictions implied by its axioms.

4 Conclusion and outlook

We have reviewed six formalisms for assigning sizes to subsets of $\mathbb{N}$: cardinality, infinite lottery logic with mirror cardinalities, natural density, generalised density, α-numerosity and c-numerosity. By evaluating these formalisms together and by studying the connections between them, we gained a better understanding of the essential trade-offs between different approaches to assigning sizes to subsets of $\mathbb{N}$.

We paid special attention to α-numerosity and c-numerosity, which both respect the Euclidean principle, while providing a measure that is total on $\mathbb{N}$ and fine-graining the natural density. α-Numerosity is the most fine-grained of the two and yields a total order of its values. However, this comes at the cost of arbitrary structure associated with intangibles. C-numerosity does not rely on intangibles, and hence does not introduce arbitrary structure in its notion of size. As such, it overcomes earlier criticisms raised against Euclidean theories of size (in particular from Parker, 2013). However, it does require us to drop another long-held assumption about set sizes, i.e., that they are totally ordered. This demand comes on top of dropping both the

Humean principle and the assumption that sizes are Archimedean quantities, which all Euclidean theories of size applicable to infinite sets have to give up.

In other words, if we start from the classical dilemma between the Humean and the Euclidean principle (previously discussed by Mancosu, 2009) and walk the Euclidean path, we encounter another fork down the road. The additional dilemma asks us to choose between accepting the consequences of invoking an intangible (α-numerosity) or dropping the total order of sizes (c-numerosity). Rather than arguing for either of the options, I suggest that α-numerosity and c-numerosity are best viewed as two sides of the same coin: the inequalities of c-numerosity trace the intervals in which all possible α-numerosity values lie. Together, they show us that there is an inherent limitation on how closely a Euclidean notion of size, applicable to all subsets of $\mathbb{N}$, can resemble its counterpart for finite sets.

Density-based approaches, which were intended to measure relative size, escape the first dilemma and show a third path (retaining neither principle). Still, they similarly face the second dilemma: a choice between invoking an intangible (generalised density) or dropping totality (natural density).

Moreover, we reconstructed the six theories as different types of limit operations on sequences of partial sums of the characteristic sequence of a given set, $f_n(S)$, or the corresponding density, $f_n(S)/n$. As a result, we could trace the differences in fine-grainedness of these measures as well as the non-constructive aspects present in two of them to the underlying limit operation.

Although the overview might suggest that all four combinations of two properties (considering two types of filters on $\mathbb{N}$ and two tolerance principles) have now been fully developed, there is room for exploring new variants. After all, natural density and c-numerosity are both based on the Fréchet filter with a different tolerance principle (recall Table 2). The fact that natural density is an Archimedean measure, while c-numerosity is not, is a direct consequence of the stricter tolerance principle of the latter theory. Yet, they *also* differ in the way they give up totality: while natural density gives up totality on the domain, c-numerosity merely gives up totality of the ordering on the co-domain.

This suggests two new approaches, which have not been developed (as far as I know). First, we could consider a ‘gappy’ numerosity theory, that only assigns values to subsets of $\mathbb{N}$ that have a super-α-numerosity (i.e., only to finite and co-finite sets). This would be a constructive theory, weaker than c-numerosity. In fact, it would be similar to infinite lottery logic, with V_∞ removed by a big gap. Second, and more interestingly, we could consider a constructive extension of natural density, which is total on $\mathcal{P}(\mathbb{N})$, by giving up the total order on its values. Hence, this new measure cannot take real values, since they form a totally ordered field. Instead, it takes values on a merely partially ordered field. However, this means that this field cannot be complete either (DeMarr, 1967), so we would have to give up the least upper bound property or one of the field axioms. To the best of my knowledge, this avenue has not been explored as an alternative for measuring set size.

Finally, there is room for Euclidean theories that do not aim for correspondence with natural density. Dropping the third axiom gives rise to a total, non-Archimedean, partially ordered measure, which does not compare infinite co-infinite sets unless they are related by the subset relation. It is stronger than infinite lottery logic but weaker than c-numerosity.

Acknowledgments

I am grateful to Tomasz Placek and Klaas Landsman for their feedback on an earlier version and to two anonymous reviewers for their constructive reports, which helped me to improve this paper.

References

L. M. Axon. *Algorithmically Random Closed Sets and Probability.* PhD thesis, University of Notre Dame, Graduate Program in Mathematics, Notre Dame, Indiana, 2010.

V. Benci and M. Di Nasso. Numerosities of labelled sets: A new way of counting. *Advances in Mathematics*, 173:50–67, 2003.

V. Benci and M. Di Nasso. *Alpha-Theory: Mathematics with Infinite and Infinitesimal Numbers.* World Scientific, Singapore, 2019.

V. Benci, M. Di Nasso, and M. Forti. An Aristotelian notion of size. *Annals of Pure and Applied Logic*, 143:43–53, 2006.

V. Benci, L. Horsten, and S. Wenmackers. Non-Archimedean probability. *Milan Journal of Mathematics*, 81:121–151, 2013. doi: 10.1007/s00032-012-0191-x.

V. Benci, L. Horsten, and S. Wenmackers. Infinitesimal probabilities. *British Journal for the Philosophy of Science*, 69:509–552, 2018. doi: 10.1093/bjps/axw013.

E. Bottazzi and M. G. Katz. Infinitesimals via Cauchy sequences: Refining the classical equivalence. *Open Mathematics*, 19:477–482, 2021.

G. Cantor. Beiträge zur Begründung der transfiniten Mengenlehre. *Mathematische Annalen*, 46:481–512, 1895.

R. Dedekind. *Was sind und was sollen die Zahlen?* Vieweg, Braunschweig, Germany, 1888.

R. DeMarr. Partially ordered fields. *The American Mathematical Monthly*, 74:418–420, 1967.

K. Easwaran. Regularity and hyperreal credences. *Philosophical Review*, 123:1–41, 2014.

S. Feferman. Does mathematics need new axioms? *The American Mathematical Monthly*, 106:99–111, 1999.

B. Fine and G. Rosenberger. *Number Theory. An Introduction via the Density of Primes*. Birkhäuser, Cham, Switzerland, 2016. Second edition.

M. Forti. A Euclidean comparison theory for the size of sets. `http://arxiv.org/abs/arXiv:2212.05527`, 2022.

G. Frege. *Die Grundlagen der Arithmetik*. Verlag von Wilhelm Koebner, Breslau, 1884.

G. Galilei. *Discorsi e dimostrazioni matematiche, intorno à due nuove scienze*. Elsevier, Leiden, 1638. Translated by H. Crew and A. de Salvio, introduced by A. Favaro "Dialogues Concerning Two New Sciences". Macmillan New York, 1914.

K. Gödel. What is Cantor's continuum problem? *American Mathematical Monthly*, 54:515–525, 1947.

R. Goldblatt. *Lectures on the Hyperreals; An Introduction to Nonstandard Analysis*, volume 188 of *Graduate Texts in Mathematics*. Springer, New York, NY, 1998.

T. Hofweber. Infinitesimal chances. *Philosophers' Imprint*, 14:1–14, 2014.

D. Hume. *A Treatise of Human Nature*. London, UK, 1739–1740. Reprinted by Oxford University Press, Oxford, UK, 2000.

T. J. Jech. *Set Theory*. Springer Monographs in Mathematics. Springer, Berlin, Germany, 2003. Reprint of the Third Millennium Edition, Revised and Expanded.

T. Kerkvliet and R. Meester. Uniquely determined uniform probability on the natural numbers. *Journal of Theoretical Probability*, 29:797–825, 2016.

J. Kubilius. *Probabilistic Methods in the Theory of Numbers*, volume 11 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, Rhode Island, 1964. Translated by G. Burgie and S. Schuur.

P. Lynch and M. Mackey. Counting sets with surreals. Part I: Sets of natural numbers. arXiv preprint arXiv:2311.09951, 2023.

P. Mancosu. Measuring the size of infinite collections of natural numbers: Was Cantor's theory of infinite number inevitable? *The Review of Symbolic Logic*, 2:612–646, 2009.

P. Mancosu. In good company? On Hume's principle and the assignment of numbers to infinite concepts. *The Review of Symbolic Logic*, 8:370–410, 2015.

S. Manyama, 2019. `https://oeis.org/A325912`; retrieved May 6, 2024.

P. Martin-Löf. The definition of random sequences. *Information and Control*, 9:602–619, 1966.

S. Maschio. Natural density and probability, constructively. *Reports in Mathematical Logic*, 55:41–59, 2020.

J. D. Norton. Eternal inflation: When probabilities fail. *Synthese*, 198: 3853–3875, 2021.

E. Palmgren. Developments in constructive nonstandard analysis. *The Bulletin of Symbolic Logic*, 4:233–272, 1998.

M. W. Parker. Set size and the part–whole principle. *Review of Symbolic Logic*, 6:589–612, 2013.

J. Pawlikowski. The Hahn–Banach theorem implies the Banach–Tarski paradox. *Fundamenta Mathematicae*, 138:21–22, 1991.

G. Peano. Sugli ordini degli infiniti. *Rendiconti della Reale Accademia dei Lincei*, 19:778–781, 1910. Reprinted in: G. Peano, *Opere Scelte*, vol. 1, Edizioni Cremonese, Roma (1957), pp. 359–362.

E. Schechter. *Handbook of Analysis and its Foundations.* Academic Press, San Diego, CA, 1997.

G. Schurz and H. Leitgeb. Finitistic and frequentistic approximation of probability measures with or without σ-additivity. *Studia Logica*, 89:257–283, 2008.

Th. A. Skolem. Über die Nicht-charakterisierbarkeit der Zahlenreihe mittels endlich oder abzählbar unendlich vieler Aussagen mit ausschliesslich Zahlenvariablen. *Fundamenta Mathematicae*, 23:150–161, 1934.

R. Sorensen. Vagueness. In E. N. Zalta and U. Nodelman, editors, *Stanford Encyclopedia of Philosophy.* 2022. `https://plato.stanford.edu/archives/win2023/entries/vagueness/`.

T. Tao. Generalisations of the limit functional, 2017. URL `https://terrytao.wordpress.com/2017/05/11/generalisations-of-the-limit-functiona`

G. Tenenbaum. *Introduction to Analytic and Probabilistic Number Theory*, volume 163 of *Graduate Studies in Mathematics.* American Mathematical Society, 2015.

K. Trlifajová. Sizes of countable sets. *Philosophia Mathematica*, 32:82–114, 2024.

R. von Mises. *Probability, Statistics and Truth.* George Allen & Unwin, London, UK, 1928. Reprinted by Dover, Mineola, NY, 1981.

S. Wenmackers. Infinitesimal probabilities. In J. Weisberg and R. Pettigrew, editors, *Open Handbook of Formal Epistemology*, pages 199–265. PhilPapers Foundation, 2019. URL `https://philpapers.org/archive/WENIP.pdf`.

S. Wenmackers and L. Horsten. Fair infinite lotteries. *Synthese*, 190:37–61, 2013.

T. Williamson. *Vagueness.* Routledge, London, UK, 1994.

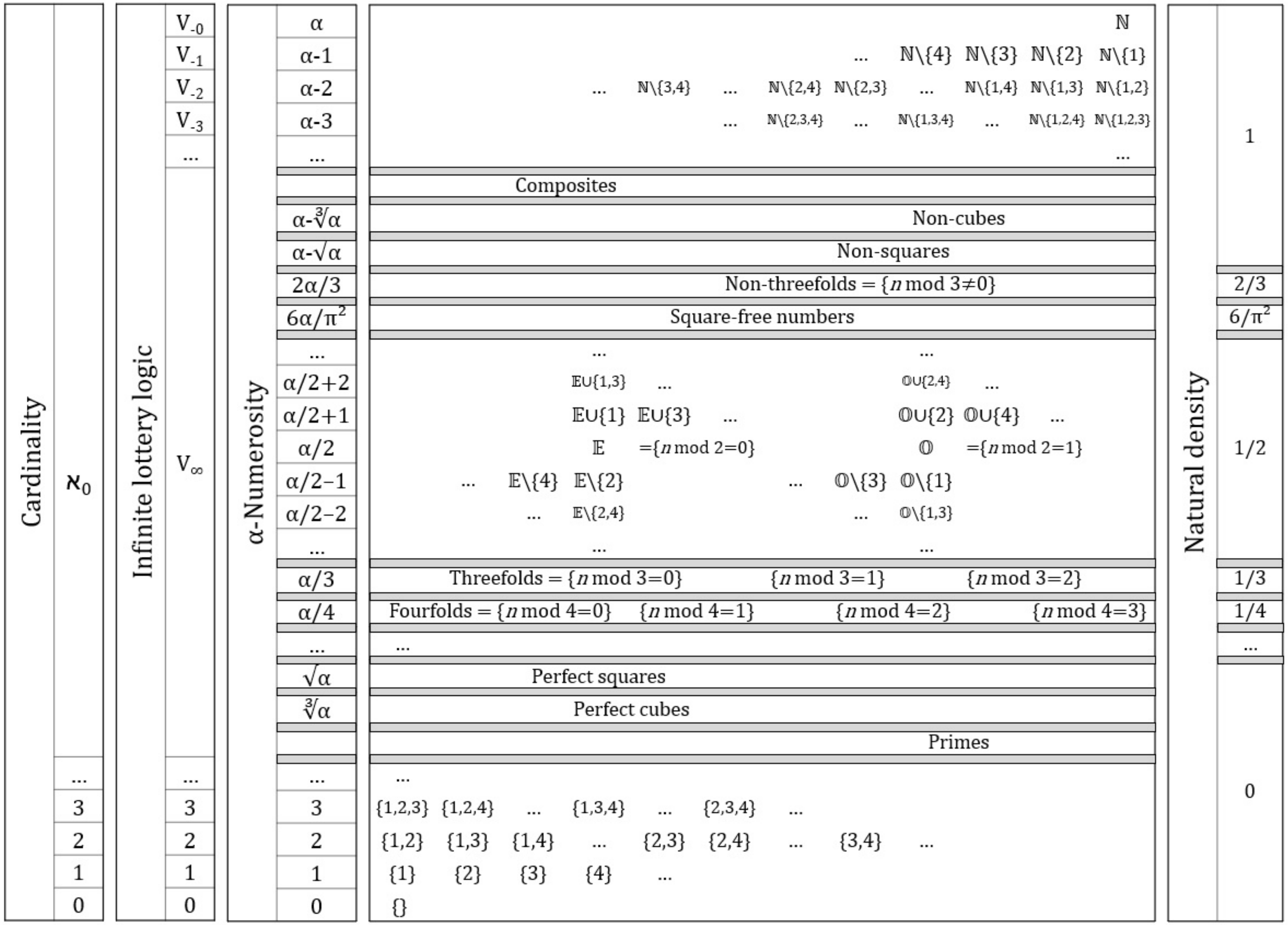


Figure 2: The central part of the diagram depicts part of $\mathcal{P}(\mathbb{N})$ stratified by the subset relation. Various vertical axes represent different measures of size and large cells indicate coarse-graining. Grey bands in the central diagram as well as in two scales indicate omitted parts.

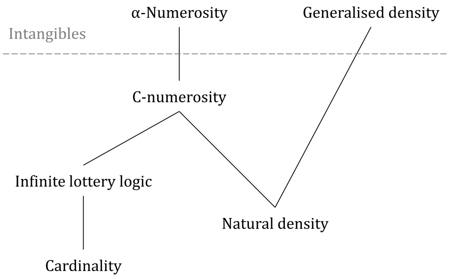


Figure 3: Partial order of the distinctions in set sizes made by various formalisms: more fine-grained towards the top. The dashed line indicates that intangibles occur above it.