



Arithmetic, enumerative induction and size bias

A. C. Paseau¹

Received: 30 December 2020 / Accepted: 5 May 2021 / Published online: 27 May 2021
© The Author(s) 2021

Abstract

Number theory abounds with conjectures asserting that every natural number has some arithmetic property. An example is Goldbach's Conjecture, which states that every even number greater than 2 is the sum of two primes. Enumerative inductive evidence for such conjectures usually consists of small cases. In the absence of supporting reasons, mathematicians mistrust such evidence for arithmetical generalisations, more so than most other forms of non-deductive evidence. Some philosophers have also expressed scepticism about the value of enumerative inductive evidence in arithmetic. But why? Perhaps the best argument is that known instances of an arithmetical conjecture are almost always small: they appear at the start of the natural number sequence. Evidence of this kind consequently suffers from size bias. My essay shows that this sort of scepticism comes in many different flavours, raises some challenges for them all, and explores their respective responses.

Keywords Mathematics · Arithmetic · Induction · Enumerative induction · Non-deductive methods in mathematics · Inductive scepticism · Mathematical justification · Goldbach's conjecture · Size bias · Frege · Mill

1 Introduction

By enumerative induction, I mean an inference from particular instances to a generalisation.¹ So from ' A_1 is F ', ' A_2 is F ', ..., to ' $\text{All } A\text{s are } F$ '. As an opening exercise, I invite you to think about a few enumerative inductions in mathematics, in which the generalisation is respectively about: \mathbb{N} (the natural numbers); the semi-closed,

¹ Enumerative induction is a species of non-deductive reasoning. Norton (2010) is a survey of induction. For the avoidance of doubt, with a minor exception in the penultimate section 'induction' in this paper is always used in its epistemological sense rather than in the sense of 'the principle of mathematical induction'.

This article belongs to the topical collection on Explanatory and Heuristic Power of Mathematics, edited by Sorin Bangu, Emiliano Ippoliti, and Marianna Antonutti.

✉ A. C. Paseau
alexander.paseau@philosophy.ox.ac.uk

¹ Philosophy Faculty, University of Oxford, Oxford, UK

semi-open real interval $[0, 2\pi)$; and \mathbb{C} (the complex numbers). In each case, assume no more about the relevant mathematics than is required to understand the statement in question—try not to import any background knowledge. Consider the evidential scenarios below and, before reading on, pause to think. Your task in each case is to gauge roughly how strong the evidence is for the statement in question.

1. The statement before you is the binary Goldbach Conjecture, *viz.* that every even number from 4 onwards is the sum of two primes. Imagine you know nothing about this conjecture other than it has been checked for every number up to 4×10^{18} .²
2. You are presented with the (to you) novel claim that $\sin^2(\theta) + \cos^2(\theta) = 1$ for all θ in $[0, 2\pi)$. (We may imagine that you know very little trigonometry beyond the definitions of sine and cosine, and that you can't deduce the identity from them because you don't know Pythagoras' Theorem.) Your evidence this time consists of 4×10^{18} values of θ drawn randomly from $[0, 2\pi)$. Each of these instances corroborates the identity.
3. Consider the Riemann Hypothesis (RH). It states that the nontrivial zeros of the zeta function $\zeta(s)$ all lie on the line with real part $\frac{1}{2}$.³ A *truncated* version of RH states that all nontrivial zeros of $\zeta(s)$ with modulus less than or equal to 1 all lie on the line with real part $\frac{1}{2}$. Let's suppose that, in this third case, your only evidence (or relevant knowledge) for truncated-RH consists of 4×10^{18} randomly chosen values of s drawn from the set of complex numbers of modulus less than or equal to 1. Once more, all these instances corroborate the hypothesis.

Remember not to assume any background knowledge in any of these cases. How strong do you think your evidence is in each case? Strong? Weak? No evidence at all?

If you're at all like me, your reaction was as follows. The evidence in scenario 1—for Goldbach's Conjecture—is weaker than the evidence in scenarios 2 or 3, for the trigonometric identity and truncated-RH respectively. In scenarios 2 and 3, the evidence is spread fairly uniformly across the relevant range of values.⁴ They are dotted all over the respective interval or region. Not so in scenario 1, in which the verified instances make up an initial segment of the natural numbers.

I think most mathematically minded people are like me (and perhaps you). They would regard the evidence in scenario 1 as weaker than that in scenario 2 or in scenario 3. And it's easy to put one's finger on why. In scenario 1, the evidence is potentially biased, as it consists only of the *first* 4×10^{18} natural numbers. Since the size of a natural number significantly affects its properties, the evidence in scenario 1 is biased with respect to size. Not so in scenarios 2 and 3, where the sample points—the evidence—don't seem biased in any obvious way, as they are drawn randomly from the entire set.

To put this in a broader context, let's compare this form of inductive scepticism to some others. The most general and strongest form takes inductive generalisations, in

² As has been done by Tomás Oliveira e Silva: see www.sweet.ua.pt/tos/goldbach.html. The induction here is over \mathbb{N} if we think of GC as the statement 'for all natural numbers n , if n is even and greater than 2 then it's the sum of two primes'.

³ The so-called critical line. The Riemann zeta function $\zeta(s)$ is the analytic continuation of $\sum_{n=1}^{\infty} \frac{1}{n^s}$ to $\mathbb{C} \setminus \{1\}$. Its trivial zeros are 2, 4, 6, \dots . See Borwein et al. (2008) and Franklin (1987) for some of the further non-deductive evidence behind RH.

⁴ Obviously, I'm using the term 'uniformly' in a non-technical sense here.

mathematics or elsewhere, to have *no* justificatory force. A narrower but still fairly broad form of scepticism takes the same line but restricts it to mathematics. According to it, verifying instances in mathematics may be heuristically useful but lacks justificatory force—belief in the conjecture is no more justified after the instances have been verified than before. My topic in this essay is a still narrower form of inductive scepticism, *prima facie* more plausible than either of these stronger ones. This is the scepticism illustrated by the difference in our reactions to scenario 1 on the one hand and scenarios 2 and 3 on the other. The scepticism at work here might be called ‘size-scepticism’. It is based on the worry that any initial segment of the natural numbers is a biased sample of the whole set, because it consists of ‘small’ numbers (in some appropriate sense).⁵

A closely related form of scepticism applies to enumerative induction over a domain that may naturally be given the structure of the natural numbers.⁶ An instance of great philosophical interest is induction over a domain of proofs, e.g. proofs in an axiomatic system such as first-order ZFC. No ZFC-proof in the system’s history has issued in a contradiction, a fact which some take as good grounds for ZFC’s consistency. As above, one might complain that all the ZFC proofs we have come across are small—they can be written, sketched or outputted by a human or computer. So, the complaint goes, they are a potentially biased sample of the domain of all ZFC proofs.

My aim in this essay is to explore size-scepticism. I shall point out that it applies only in limited contexts (§3), separate out its different versions (§4), and examine how these versions respectively deal with some challenges I raise (§§5–6). I consider size-scepticism about proofs more briskly in §7. We begin in §2 with a brief review of some of the relevant philosophical literature.

2 Sceptics

In the early sections of *The Foundations of Arithmetic*, Frege argues against Mill’s view that mathematics as a whole is founded on inductive evidence.⁷ Frege articulates his scepticism about enumerative induction in mathematics in the following passage, in which he unfavourably compares arithmetical inductions to scientific ones:

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. Our results must hold good for any other place and any other time, provided only that the conditions are the same. But in the case of the numbers this does not apply, since they are not in space or time. Position in the number series is not a matter of indifference like position in space (Frege 1884, §10).

⁵ As relatively subtle differences between the terms ‘evidence’ and ‘justification’ won’t be relevant, I use these terms interchangeably throughout.

⁶ Clearly, any countably infinite domain X may be given the well-ordering of the natural numbers: simply use the bijection from ω to X to define an ω -order on X . The adverb ‘naturally’ (intended intuitively, not in a mathematically precise sense) is inserted for wriggle room, to allow that perhaps not all countably infinite domains may naturally be ordered in this fashion.

⁷ A representative quotation from Mill: ‘With regard to axioms, we found that, considered as experimental truths, they rest on superabundant and obvious evidence.’ (Mill 1843/1973, Book II, Ch. VI, §1).

Frege more generally doubted that checking (finitely many) instances of an arithmetical generalisation could ever justify it. A contemporary author who, in his own words, takes his cue from Frege is Alan Baker. He writes:

The one distinctive feature of the mathematical case which ought to make a difference to the justification of enumerative induction (or so I shall argue) is the importance of size. By this I mean that the instances falling under a given mathematical hypothesis (at least in number theory) are intrinsically ordered, and furthermore that position in this order can make a crucial difference to the mathematical properties involved. (Baker 2007, p. 66)

Baker articulates this thought by means of a so-called *non-uniformity principle*:

...in the absence of proof, we should not expect numbers (in general) to share any interesting properties.... Hence establishing that a property holds for some particular number gives no reason to think that a second, arbitrarily chosen number will also have that property. (Baker 2007, p. 66)

He goes on to spell out his thought as follows:

Definition: a positive integer, n , is *minute* just in case n is within the range of numbers we can (given our actual physical and mental capabilities) write down using ordinary decimal notation, including (non-iterated) exponentiation. Verified instances of GC to date are not just small, they are minute. And minuteness, though admittedly rather vaguely defined, is known to make a difference. (Baker 2007, p. 67)

Baker's scepticism is based not on the size of the confirmatory sample being small (what Walsh (2014) calls *setwise-smallness*), but rather on each confirmatory instance being small (what Walsh (2014) calls *pointwise-smallness*). To co-opt Walsh's terminology, scientific inductions over physical domains are frequently setwise-small, as the number of instances is often small relative to the population. Usually, though, they are not pointwise-small in any clear sense: observed ravens, for instance, are not small compared to all the ravens at large. That's why Baker thinks there is room for a specifically arithmetical form of inductive scepticism.

Baker ends his discussion with both a normative and a descriptive conclusion: mathematicians ought not and in general do not 'give weight to enumerative induction *per se* in the justification of mathematical claims' (2007, p. 72).⁸ But following other writers, he allows that circumstantial reasons can come to the rescue of an enumerative induction, in arithmetic as well as elsewhere in mathematics. Other philosophers have expressed distrust of inductive evidence of mathematics, and in particular enumerative inductive evidence taken in isolation, in a way that suggests that they too are at least sympathetic to this type of scepticism.⁹

⁸ His discussion is explicitly focussed on arithmetic, and he cautions that his conclusions may not apply to other areas of mathematics (2007, p. 60). On p. 64, Baker also clarifies that he is perfectly happy to grant the rationality of inductive inference in empirical cases.

⁹ Mark Steiner (1975, pp. 102–8, esp. p. 107) for example discusses Euler's discovery that the value of the series $1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \dots$ agrees with $\frac{\pi^2}{6}$ to several decimal places. The reason he believes 'Euler had

Baker's scepticism has come under attack in Walsh (2014, section 2.2) and Waxman (2017, section 3.4). Broadly speaking, the complaint is that physical samples in a natural-scientific enumerative induction are biased in a similar way to the size bias of verified instances in an arithmetical enumerative induction. After all, physical samples' spatiotemporal distance to us is 'small'. Yet empirical inductions regarding 'nearby' events are often justified, which suggests that it's not enough simply to note that samples in arithmetical cases are small. Some further argument is required to suppose this involves problematic bias.

There is much to be said here, but since my focus will be different, let me state my conviction rather dogmatically. Although sympathetic to this response, I don't think it's dialectically very effective. Its effect is intended to throw the burden of proof back on to the sceptic, i.e. Baker. But there are disanalogies as well as analogies between the arithmetical and the empirical cases. It is open to Baker to reply that the analogies are outweighed by the disanalogies, as reflected by mathematicians' inherent wariness of merely enumerative inductive evidence in arithmetic. For it is, I believe, indisputable that they *are* wary of such evidence. They are certainly aware of conjectures whose first counterexample is a very large number, perhaps the most famous such being that $Li(x) > \pi(x)$, where $Li(x)$ is the logarithmic integral $\int_2^x \frac{dt}{\ln(t)}$ and $\pi(x)$ the number of primes up to x .¹⁰ The ratio between the two quantities has been proved to be asymptotically 1, and the first had long been observed to be greater than the second for all verified cases. Yet as Littlewood proved in the 1910s, this is true 'only' up to a number no greater than about 1.4×10^{316} .¹¹ Size-sceptics go further than this reasonable caution. They maintain that there are particular grounds to be sceptical in the case of an arithmetical generalisation because any initial segment of the natural numbers is biased with respect to size.

In short, size-scepticism has its source in mathematicians' instinctive distrust of enumeratively inductive evidence in arithmetic. Its philosophical defence by Baker has been questioned in the literature, yet not decisively so. Instead of pursuing this line of criticism, we shall examine others below. Most importantly, we shall show that the criticisms' effectiveness depends on which sort of size-scepticism is in question. A more subtle understanding—and assessment—of size-scepticism must distinguish these different sorts.

Footnote 9 continued

a right to be confident in his discovery beyond any doubt' (1975, p. 106) was the presence of background information. Steiner makes it clear that he thinks that without such background information, in particular the analogy between (finite) polynomials and infinite-degree power series, Euler's enumerative inductive evidence would have been worth little, if anything.

¹⁰ For this to be an arithmetical generalisation in the strict sense, restrict x to integer values. Other well-known examples include the Pólya Conjecture and the Mertens Conjecture. Obviously, we are considering 'natural' conjectures only; for any N , coming up with any old conjecture whose first counterexample is N is child's play.

¹¹ Upper bounds on this first counterexample are known as Skewes's numbers.

3 How much does it matter?

The toy scenarios encountered in §1 were rather artificial. Enumerative inductive evidence is often coupled with other non-deductive evidence for a conjecture. Examples are proofs of weaker versions, evidence from limiting behaviour, and the like. Let's look at a couple of examples in a schematic way. (Needless to say, we are not aiming for an exhaustive treatment or classification.)

1. Consider the following situations a mathematician may find herself in. (The quantifiers in each case range over the natural numbers.)¹²
 - (i) She wishes to know whether $\forall nFn$. She lacks any other relevant knowledge.
 - (ii) She wishes to know whether $\forall nFn$. This time, she also knows that for some N , $(\forall n \geq N)Fn$. However, she has no idea which might be the least such N .
 - (iii) She wishes to know whether $\forall nFn$. She also knows that for some N , $(\forall n \geq N)Fn$. On top of that, she knows an N for which $(\forall n \geq N)Fn$ is true, but doesn't know whether this particular N is the smallest such.

It's pretty clear, I take it, that the enumerative inductive evidence lends more support to the conjecture in cases (ii) and (iii) than in case (i), in which the enumerative inductive evidence is 'bare'—unaccompanied by any other information. Since cases of type (ii) and type (iii) regularly crop up, the relevance of size-scepticism to day-to-day mathematics should not be exaggerated.

The ternary Goldbach Conjecture, that every odd number ≥ 7 is the sum of three primes, illustrates the point. The Soviet mathematician Ivan Vinogradov first unconditionally proved the eventual truth of the ternary Goldbach Conjecture in the 1930s. One of the first upper bounds to emerge, much improved on over the decades, was $3^{3^{15}}$, thereby putting mathematicians in situation (iii). According to some accounts, Vinogradov was in situation (ii) when he first came up with his proof, since he knew of no upper bound. Finally, in 2013, the Peruvian mathematician Harald Helfgott proved the conjecture outright.

I won't venture to guess what percentage of cases are of type (ii) or (iii), if this is even a well-posed question. Fairly obviously, though, there are plenty of them.

2. In several cases, supplementary knowledge can turn the tables on the size-sceptic. Goldbach's Conjecture (GC) illustrates the point nicely. In the thought experiment that opened this paper, I asked you to imagine you had no knowledge relevant to GC other than all the cases up to 4×10^{18} in its favour. However, as Echeverría (1996) points out, the enumerative inductive evidence suggests that the number of ways in which a number can be written as the sum of two primes is broadly increasing, with oscillation—that is, not strictly increasing but with a steady increasing trend. This suggests that smaller numbers are more likely to be counterexamples than larger ones.¹³

¹² Not to complicate things, we may assume that in (ii) and (iii) the additional knowledge is deductive. Clearly, there are lots of other cases, e.g. intermediate cases between (ii) and (iii).

¹³ Of course, this evidence for thinking that early cases are hard cases is itself based on enumerative induction. Yet unless one is an out-and-out sceptic about enumerative induction, the circle here is virtuous. Enumerative inductive evidence is deployed to show that, so far as GC is concerned, the size bias works in favour of someone deploying enumerative inductive evidence to confirm it.

A well-known heuristic probabilistic argument suggests the same conclusion. The Prime Number Theorem states that the number of primes up to N tends to $\frac{N}{\ln N}$ asymptotically: $\pi(N) \sim \frac{N}{\ln N}$.¹⁴ Hence the number of distinct sums of primes no greater than $2N$ tends to $\frac{1}{2} \cdot (\frac{N}{\ln N})^2$ (dividing by two since each sum, with the possible exception of $N + N$, appears twice). Thus the typical even integer $\leq 2N$ can be written as the sum of two primes in about $\frac{1}{2} \cdot (\frac{N}{\ln N})^2 / N = \frac{N}{\ln^2 N}$ ways, a quantity that increases with N .¹⁵ This argument is admittedly very rough and ready, and, to stress, heuristic rather than demonstrative. But it can be improved upon to yield much better estimates that point to the same conclusion: the greater N is, the greater $G(N)$ is likely to be. Such arguments remain heuristic—GC has not been proved—but they make it very plausible that the first counterexample to GC, if one exists, will be a ‘small’ number.

In the case of GC, there are therefore conjecture-specific reasons for thinking that the earliest cases are the ‘hardest’. Anyone armed with this background knowledge will see the inductive evidence for GC as stronger than they would otherwise—but not for any *general* reasons linking size and inductive evidence.¹⁶

GC is by no means unique in this respect. Another example is Legendre’s Conjecture, also currently unproved, which states that for every positive integer N there is a prime between N^2 and $(N + 1)^2$. As in the case of GC, the enumerative inductive evidence suggests that not only is Legendre’s Conjecture true, but also that the number of primes between N^2 and $(N + 1)^2$ non-strictly increases as N does. And there are heuristic arguments for this conclusion too: the Prime Number Theorem implies that the number of primes between N^2 and $(N + 1)^2$ is asymptotic to $\frac{N}{\ln(N)}$, a quantity which increases with N . So for Legendre’s Conjecture, as for GC, it seems that smaller numbers are ‘hard’ cases.

In sum, given the details of a particular case and one’s background knowledge, ‘anti-size-scepticism’ may well be more justified than size-scepticism.

Many more cases could be added to the above. The moral is that size-scepticism may not affect all that many instances of inductive reasoning in arithmetic.¹⁷ Of course, there are conjectures for which we have specific reason to believe that early cases are easy cases.¹⁸ Such conjectures are diametrically opposed to GC and Legendre’s Conjecture (in this respect). My point is only that in many cases, no such size-scepticism is warranted.

¹⁴ Incidentally, coupled with the discussion in §2, this tells us that $\pi(N) \sim \frac{N}{\ln N} \sim Li(N)$.

¹⁵ The result seems to be mathematical folklore.

¹⁶ This example is also discussed by Baker (2007, pp. 69–70).

¹⁷ Whether enumerative inductive evidence, in combination with other types of non-deductive evidence, suffices to generate knowledge of the conjecture depends on the details of the case, as I discuss in Paseau (2015).

¹⁸ An example is the conjecture that all perfect numbers are even. (A natural number N is perfect iff its factors sum to $2N$.) Since Ochem and Rao (2012), we know that the smallest odd perfect number, if it exists, must be greater than $10^{1,500}$.

4 Varieties of size-scepticism

So far, we have taken scepticism about pointwise-small samples (in Walsh's terminology) to be a monolithic position. It's now time to distinguish several versions of it.

Suppose first that what drives your scepticism is the idea that small natural numbers are unlike larger ones with respect to the properties mathematicians are interested in. These numbers' small size affects how they behave mathematically. Thus any sample consisting of small numbers is potentially biased. This position itself subdivides. One variant has it that a sample set with smaller instances than another offers weaker inductive evidence than the latter. This is a *comparative* form of size-scepticism. Let's call it *c-scepticism* ('c' for 'comparative'). For the second variant, waive worries about the vagueness of 'small' for a moment. Suppose that the small numbers consist of a definite initial segment of the natural numbers, and that you believe small natural numbers are unlike larger ones with respect to the properties mathematicians are interested in. These numbers' small size affects how they behave mathematically. Thus any sample consisting of small numbers is potentially biased. Call this *s-scepticism* ('s' for 'small').

Suppose alternatively that your scepticism is driven by the idea that any convincing induction should consist of a representative sample. If a sample consists solely of small numbers, it will not be representative. Thus any sample consisting of small numbers is potentially biased with respect to the mathematical property of interest. Call this *u-scepticism* ('u' for 'unrepresentative').

Size-scepticism can therefore be considered a genus comprising (at least) three species:

The *c-sceptic* believes that an inference based on a sample is (in this respect) weaker than an inference based on another sample that contains larger instances than the first.

The *s-sceptic* believes that an inference based on a sample consisting only of small instances is (in this respect) weak precisely because the instances are small.

The *u-sceptic* believes that an inference based on a sample consisting only of small instances is (in this respect) weak because the instances are small and therefore unrepresentative.

These three types of size-sceptic often draw the same conclusion about the inductive strength of a sample set, even if they come to it in different ways. But not always: their conclusions can diverge. To see this, suppose—for the sake of concreteness—that a number is small iff it is smaller than 10^{100} . Suppose further that the enumerative inductive evidence consists of 10^{10} instances, each in the range 2×10^{100} to 3×10^{100} . Since no instance is small, the s-sceptic does not regard the sample as potentially biased.¹⁹ In contrast, the u-sceptic sees the evidence as potentially biased because it is all drawn from a particular range. Finally, the c-sceptic regards the sample as being evidentially stronger than an equal-sized sample drawn from the range 10^{100} to

¹⁹ At least not on the basis of her s-scepticism—she may of course regard it as potentially biased for other reasons. This sort of qualification is implicit throughout.

2×10^{100} , but weaker than an equal-sized sample drawn from the range 3×10^{100} to 4×10^{100} .

To further illustrate these three types of scepticism, consider a vaccinological analogy.²⁰ Imagine we are testing a new vaccine for its effectiveness in combatting a certain kind of virus. We find that the vaccine is highly effective in trials, but that the trial participants are all young children: they are aged between 0 and 9. You might have qualms about inferring that the vaccine is generally effective in tackling the virus on this basis. But you might do so for different reasons.

The c-sceptic is analogous to someone who would like see the vaccine trials extended to older people. They would prefer to test the vaccine on, say, adults aged 30 to 39 than young children. In fact, they believe testing the vaccine on (an equal number of) adults in the age range 40–49 would provide even better evidence than testing it on thirty-somethings. More generally, the older the vaccine-trial participants the better.

The s-sceptic is analogous to someone who believes that children under 10 have different physiologies than those of older children and adults, and in particular that they react differently to vaccines.²¹ So no vaccine tested on young children can be assumed to have the same effect on anyone older. It would be much better, on this view, to test the vaccine on (an equal number of) adults aged between 30 and 39, as their reactions to the vaccine are more likely to be representative of the population at large. But there is no reason to suppose that forty-somethings would be better than thirty-somethings; both those age ranges would in principle be equally good.

The u-sceptic is analogous to someone who believes that people of every age react differently to vaccines. This includes young children, but there is nothing special about them in this regard. No vaccine tested on young children can be assumed to have the same effect on anyone older, because their ages are from the same range. It would have been *no* better, on this view, to test the vaccine on (an equal number of) adults aged between 30 and 39.

The following table sums up the analogy between the three respective types of sceptic.

Numbers	Age of trial participants
c-sceptic	Someone who believes it's better to test an older person than a younger one
s-sceptic	Someone who believes it's better to test people older than young children
u-sceptic	Someone who believes it's better to test a diverse group of age ranges

You might well think that the analogue of c-scepticism is utterly implausible. Wouldn't it be daft, as the view clearly implies, to always prefer trialling vaccines on the world's oldest people—centenarians or (even better) supercentenarians? Prob-

²⁰ Inspired by an epidemiological analogy used by Alan Baker in a personal communication (cited with permission) to illustrate his size-scepticism. The point of my analogy is to illustrate the three varieties of size-scepticism I have distinguished.

²¹ A related fact is well-known: many medicines that work on adults are not recommended for children, and are even prohibited for paediatric use.

ably; but what this shows is no more than that the age-number analogy is, in this respect, strained. People's age is bounded, whereas the natural numbers are unbounded, which makes the vaccinological analogue of c-scepticism implausible. However, imagine for a second that people can live to any finite age and that time stretches infinitely back; that (as now) finitely many people are born every year; and that (very roughly as now) people's physiologies change monotonically with age, so that the closer in age two people are the more similar their physiologies, other things being equal. In such a scenario, c-scepticism would no longer be such a daft proposition. A 45-year-old would be more like anyone older than 40—an infinite amount of people—than a 35-year-old; and that same 45-year-old would be less like those younger than 40—only a finite amount of people—than the same 35-year-old. Generalising, it would make sense, in that imaginary scenario, to test as old a person as possible. C-scepticism's analogue would be vindicated.²²

The three variants of size-scepticism and their differences should now be tolerably clear. In assessing size-scepticism's plausibility, it will be crucial to have these differences firmly in mind, as we shall see. As an exegetical aside, Frege in *Foundations* seems to have been a wholesale sceptic about any form of enumerative induction in arithmetic, however (finitely) many and however varied the verified instances. In particular, Frege's inductive scepticism was more general and stronger than any of the size-scepticisms here described. It is less clear how to categorise Baker (2007), though a case can be made that its author leaned towards u-scepticism.²³ As my interests are not exegetical, I shall not dwell on this and turn instead to two different challenges for size-scepticism.

5 Frontloading of evidential value I

This section and the next present two 'frontloading' arguments. They appear to show that enumerative inductive strength is concentrated 'early on' (in a sense that will be made precise) in the natural-number series. The question is whether this affects size-scepticism, and if so which variety.

Let E , a finite subset of \mathbb{N} , consist of our enumerative inductive evidence for a particular arithmetical conjecture. In other words, E is the set of known instances of a generalisation over the natural numbers. Let the function v be our evidential function, with domain all finite subsets of the natural numbers. In this section, we assume only that v 's codomain is the closed unit interval $[0, 1]$ with the usual order; the higher v 's value in $[0, 1]$, the stronger the evidence. Evidential values may be thought of as measuring the subject's rational degree of confidence in the generalisation in question, though without commitment to the whole panoply of probabilistic ideas.

Consider next the following evidential principle:

²² Needless to say, whether c-scepticism itself is vindicated depends on the strength of the analogy between the imaginary-scenario-vaccine-testing case and the numerical case.

²³ A relevant remark appears on p. 68 of Baker (2007), in which Baker likens the enumerative inductive evidence to an 'unvaried' range of instances.

More is Better

If $n \notin E$ then $v(E \cup \{n\}) > v(E)$.

This principle, presumably uncontroversial, captures the idea that more evidence is better than less.

Next, define $l = \lim_{n \rightarrow \infty} I_n$, where $I_n = v(\{0, 1, \dots, n\})$. By *More is Better*, if $m < n$ then $I_m < I_n$; and since 1 is an upper bound for the I_n , the limit l exists. The real number l itself, of course, may be 1 or smaller than 1, but it has to be greater than 0 (by *More is Better*). So we deduce that $0 < l \leq 1$.

Now by the definition of a limit, for any $\epsilon > 0$, however small, there is an N_ϵ such that for any $N^* \geq N_\epsilon$,

$$l - \epsilon < I_{N_\epsilon} \leq I_{N^*} < l$$

It follows in particular that $|I_{N^*} - I_{N_\epsilon}| < \epsilon$. Here's another way of putting it: if ϵ is chosen to be much smaller than $l - \epsilon$, almost all the evidential value stems from the first N_ϵ instances of the enumerative induction. The next $N^* - N_\epsilon$ instances add very little evidential value, however large the difference between N_ϵ and N^* may be.

The evidential value of any finite amount of numerical instances is therefore concentrated almost entirely in a fixed initial segment. Whatever conjecture you wish to test, the value of further instances beyond some finite bound will be vanishingly small. This initial segment therefore provides the lion's share of the confirmation. At first sight, this appears in tension with the broad size-sceptical idea. But to avoid jumping to conclusions, let's consider how the result respectively affects our three varieties of size-scepticism (as defined in §4).

The c-sceptic believes that the smaller the numbers in one's (finite) evidence set, the lesser the confirmation. The mathematical point just noted is incompatible with her view. To take a concrete example, for some N , the evidential value of knowing the first N instances of the Goldbach Conjecture is $10^{10^{10}}$ -fold greater than the additional value of knowing the next 10^{10^N} cases.²⁴ Yet N is minuscule compared to 10^{10^N} ! So the result seems to flatly contradict c-scepticism.

As for the s-sceptic, she believes that small numbers are those below some bound. (Here and throughout, we set aside vagueness, for simplicity.) In that case, she can accept the result but insist that, in the previous few paragraphs' terminology, the relevant N_ϵ for any given arithmetical conjecture and small ϵ is (much) greater than any small number. In other words, suppose quite a few—how many will depend on the specific conjecture—non-small numbers make up part of the enumerative inductive evidence. The s-sceptic can maintain that the evidence should then be almost as convincing as any amount of (finite) enumerative inductive evidence could ever be.

The u-sceptic can respond in a very similar way. When it comes to inductive inference over the natural numbers, the u-sceptic does not have an *eo ipso* preference for evidence sets with larger numbers. Her preference, rather, is for diverse evidence sets, which are more representative. So the u-sceptic can similarly accept the result; but she will insist that the relevant N_ϵ for any given arithmetical conjecture and small ϵ

²⁴ This remark easily follows from the above result.

is (much) greater than any small number. For example, N_ϵ must be greater than any minute number on Baker's view. As long as quite a few—how many will depend on the specific conjecture—non-small numbers are included as part of the enumerative inductive evidence, then this evidence should be almost as convincing as any amount of (finite) enumerative inductive evidence could ever be.

In summary, a simple argument suggests that if evidential value is measured by a real number in the interval $[0, 1]$ then for any $\epsilon > 0$ there is an N_ϵ such that knowing arbitrarily many finite instances beyond N_ϵ adds no more than ϵ in evidential value. In contrast, all the evidential value finite evidence can yield, bar ϵ , is owed to the first N_ϵ cases—and for sufficiently small ϵ , this is the lion's share. I suggested this mathematical point is inconsistent with one version of size-scepticism: c-scepticism, which takes larger instances as providing more inductive confirmation than smaller ones. In contrast, both s-scepticism and u-scepticism are consistent with it. Naturally, one might query the application of a probabilistic framework in this context, something I'll return to in the next section.

6 Frontloading of evidential value II

In §5 we assumed no more than that the valuation function v is the closed real interval $[0, 1]$ with its standard order. In this section, we turn to a treatment closer to orthodox Bayesianism, though still distinct from it, so merely 'Bayesian-like'. According to orthodox Bayesianism, logical truths must be believed to degree 1. As a consequence, a subject who is certain of the axioms of arithmetic must be equally certain of their consequences. Orthodox Bayesianism, however, cannot make sense of the evidential situation mathematicians find themselves in. For mathematicians are disposed to believe the Peano Axioms with credence 1 or close to 1, but have no opinion about whether some decidable instances of GC, such as $10^{10^{100}}$ say, is true or not. One could argue that mathematicians are simply not living up to the ideals of rationality through their inability to follow their beliefs' logical consequences. Another reaction is to suppose that there is a less idealised sense of rationality in which mathematicians are *not* guilty of irrationality simply by believing axioms without believing all their consequences. For the rest of this section, let's assume this latter, more realistic, sense of rationality. Since orthodox Bayesianism is clearly incapable of modelling it, we shall have to go against some of its tenets, while holding on to other parts of the framework.

6.1 C-scepticism

We start with c-scepticism. Let p_i be the probability that number i has property F ,²⁵ probabilities being understood as credences. Suppose further that the subject's credences in the instances of the arithmetical generalisation $\forall nFn$ are independent. This is a fairly plausible assumption for anyone with no relevant knowledge of the conjecture; and, as we shall explain below, it is in any case superfluous. In a probabilistic setting, in which we assume no credential dependence between instances, it follows

²⁵ We blur the difference between predicates and the properties they denote.

that:

$$Pr(\forall x Fx) = \prod_{i=0}^{\infty} p_i$$

The obvious way to define the infinite product $\prod_{i=0}^{\infty} p_i$ is as $\lim_{N \rightarrow \infty} \prod_{i=0}^N p_i$. Since the sequence $(\prod_{i=0}^N p_i)_{N \in \omega}$ —that is, $(\prod_{i=0}^0 p_i)$, $(\prod_{i=0}^1 p_i)$, $(\prod_{i=0}^2 p_i)$, \dots —consists of non-increasing real numbers bounded below by 0, it must have a limit, and this last is what we define $\prod_{i=0}^{\infty} p_i$ as. (As a preview of §7, we mention that the product is unchanged however we permute the p_i .)

One way of modelling c-scepticism is to suppose that $p_i > p_j$ iff $i < j$ —the smaller i is, the larger p_i is. This embodies the idea that the smaller the number, the greater the subject’s degree of belief that it is a true instance of the generalisation; so to test the conjecture, the larger numerical instance the better. But in that case, the subject’s credence in the universal statement must be 0, since

$$Pr(\forall x Fx) = \prod_{i=0}^{\infty} p_i < p_1^k$$

for any non-negative (integer) k , so that the left-hand product is 0. No amount of finite evidence, wherever in the number line it may fall, could then lift this credence from 0 to a positive amount. This way of cashing out c-scepticism therefore has the unacceptable implication that no universal generalisation can be supported to any positive degree by any enumerative inductive evidence.

The same moral applies to c-scepticism modelled in the following way: $p_i \geq p_j$ iff $i \leq j$, and $p_k > p_{k+1}$ for some k .²⁶ If the prior credence in $\forall x Fx$ is to be non-zero, our discussion in fact points to something like the opposite of c-scepticism. For example, under the assumptions just stated, we may prove:²⁷

Observation

Suppose $Pr(\forall x Fx) = \prod_{i=0}^{\infty} p_i \neq 0$. Then for any $0 < p_i < 1$, eventually all $p_j > p_i$. (‘Eventually all’ means ‘all but finitely many’.)

Clearly, *Observation* tell against c-scepticism. Prior to investigation, no p_i is equal to 1.²⁸ It follows that any instance of the conjecture is more dubious than all but finitely many of the instances that follow it. *Observation* thus supports an opposite moral to the c-sceptic’s.

To illustrate the point numerically, we model a form of inverse c-scepticism in which p_i strictly increases as i increases. Suppose that, for a given arithmetical generalisation, $p_i = (\frac{1}{2})^{3-(i+1)}$, so that $p_0 = (\frac{1}{2})^{\frac{1}{3}}$, $p_1 = (\frac{1}{2})^{\frac{1}{2}}$, etc. Clearly, p_i is an increasing function of i , meaning that $p_i < p_j$ iff $i < j$. This means that, before the evidence

²⁶ Since $\prod_{i=k+1}^{\infty} p_i < (p_k)^n$ for all n , and the right-hand side tends to 0 as n tends to ∞ .

²⁷ By an easy argument left to the reader.

²⁸ Or perhaps p_i is 1 for a few small i , when the result is obvious, e.g. for very small cases of GC.

comes in, smaller numbers are perceived as more likely to be counterexamples to the conjecture than smaller ones. A little algebra shows that

$$Pr(\forall x Fx) = \prod_{i=0}^{\infty} p_i = \prod_{i=0}^{\infty} \left(\frac{1}{2}\right)^{3^{-(i+1)}} = \left(\frac{1}{2}\right)^{\frac{1}{2}} \approx 0.7071$$

Suppose a subject has verified the generalisation's first 10^{10} cases. Her updated credence in the generalisation should then be:

$$Pr(\forall x Fx | \forall x \leq 10^{10} Fx) = \prod_{i=10^{10}+1}^{\infty} \left(\frac{1}{2}\right)^{3^{-(i+1)}},$$

a number extremely close to 1. Before considering how these results affect the s-sceptic and the u-sceptic, let's consider a potential response. For the c-sceptic faced with these results is liable to complain about the probabilistic framework the results assume.

6.2 Questioning the framework

We know that the standard Bayesian framework cannot be applied wholesale to mathematics, since for example it forces logical truths to have probability 1.²⁹ A certain sort of philosopher might complain that because we lack a good mathematical model for them (in particular, Bayesians have so far found mathematics a hard nut to crack), we shouldn't trust judgments of this type. However, this would be to put the cart before the horse. It would be to suppose that however systematic, reliable, and informally understood some phenomena might be, in the absence of an undergirding theory they are not to be trusted. Compare the situation before the emergence of probability in early modern mathematics and long before its twentieth-century axiomatisation. Would it have been unreasonable to think some claims more credible than others, to believe evidence can confirm a hypothesis, etc., in the absence of a mathematical model? Is it unreasonable now for anyone who has not taken a course in formal epistemology?

Clearly, the answer is no. But to this point, one might respond: fine, we have no satisfactory formal model for non-deductive reasoning in mathematics. Yet our discussion in §6.1 *did* assume a model: not a Bayesian model *per se*, but a Bayesian-like one. And what reason is there to think that *it* correctly models the phenomena in question? C-scepticism's critics are likely to make two points in response.

To begin with, it seems hard to deny that degrees of confidence in these types of mathematical cases may be measured by real numbers in the interval $[0, 1]$. Degrees of confidence are strongly, even if not constitutively, linked to betting behaviour. And clearly prior to investigation one can bet on whether the number 10^{10} is the sum of two primes. Or whether all the first M numbers are. If the subject regards each case as independent, then according to the more descriptive notion of rationality here targeted, the subject's overall credence in the generalisation should be calculated as above. And if the cases are *not* independent, we may take the probability p_i to be that of the i^{th} number's conforming to the conjecture given that all previous ones do, in which case

²⁹ Though we call it 'Bayesian', the version of Bayes's Theorem used in §6.1, in which the events are independent, is extremely simple. We use the label 'Bayesianism' because the probabilities here are all construed as credences, and because we make more sophisticated use of Bayes's Theorem when the events are not independent in a couple of places in §6.

the overall probability is once more the product of the relevant probabilities. This is a natural picture, and the burden is on the *c*-sceptic to fault it. She cannot rest content with pointing out that it conflicts with the for-us unattainable requirement of logical omniscience.

The second point is that there is no shortage of models trying to improve upon orthodox Bayesianism. Classic attempts include Hacking (1967) and Garber (1983). In the latter, for example, sentences such ‘*A* entails *B*’, where *A* and *B* may be arithmetical sentences, are treated as atoms, to avoid the assumption of logical omniscience. Without a high credence in ‘*A* entails *B*’, a subject may then rationally have high credence in *A* but low credence in *B*, even if *A* entails *B*. More recently, formal epistemologists have started to develop more sophisticated Bayesian-like models accounting for agents with limited cognitive resources. One such is that of Gaifman (2004) and another that of Skipper and Bjerring (2020). The latter for instance avoids the assumption that if *A* entails *B* then one’s credence in *B* should be no lower than one’s credence in *A*, yet in other ways preserves as much of the Bayesian framework as possible. Very roughly, Skipper and Bjerring’s idea is to use a step-based model of bounded logical reasoning, in which the number of inference steps an agent is able to perform is modelled by a natural number *n* so that any consequences of principles more than *n* steps away are not transparent to the agent.

As this is not the place for an in-depth review of any of these models, we note simply that the name of the game in this particular subfield of formal epistemology is to reach a theoretical understanding of reasoning by bounded creatures whilst holding on to as much of the traditional framework as possible. Though all such models are controversial,³⁰ the best candidates vindicate something like the ‘Bayesian-like’ model assumed. In particular, a subject may have credence 1 in the Peano Axioms but low credence in some of their consequences, and may well be agnostic about an instance of the Goldbach Conjecture even if it follows from axioms they have full confidence in. Indeed, any proposed model of partial belief in mathematics that aims to be more realistic than orthodox Bayesianism will be judged by how well it accords with these and similar aspects of mathematical reasoning.

The *c*-sceptic might be moved by the two responses so far, and accept that Bayesian-like models that reject logical omniscience can correctly capture *some* forms of non-deductive reasoning in mathematics. But she might deny that Bayesian-like models can be applied to *infinite* cases. For she might urge that any form of Bayesianism, orthodox or heterodox, is ill-equipped to deal with these. It’s widely appreciated that assigning probabilities to all subsets of an infinite event space can be tricky or impossible. For instance, what should one’s credence be that ticket *i* will win in a fair infinite lottery with ticket numbers 0, 1, 2, ...? All tickets must be assigned the same probability, by assumption. This probability cannot be positive, since by countable additivity the probability of some ticket being the winning one will then be infinite rather than 1. Nor can it be 0, however, since the probability of some ticket or other being the winning one will then be 0.

But there is, in turn, a response to the *c*-sceptic. Notice that there are all sorts of consistent probability assignments under which the overall probability of an arithmetical

³⁰ As noted in for example in Easwaran (2011, p. 324) or Skipper and Bjerring (2020).

generalisation is non-zero. As we saw, *none* of these vindicates c-scepticism. In fact, quite the opposite: *all* such assignments support a sort of inverse c-scepticism.

For an analogy, suppose someone makes a claim about what probability assignments in (countably) infinite lottery cases look like. For example, she might contend that all such assignments must have non-decreasing probabilities, i.e. that any rational agent should believe that ticket i will win to degree no greater than ticket j will whenever $i < j$. By means of a simple mathematical argument,³¹ we point out to her that in a Bayesian framework there are *no* such consistent assignments. However, there are all sorts of probability assignments consistent with Bayesianism in which the probabilities are decreasing.³² The response to this argument cannot be that because a Bayesian approach cannot model infinite fair lotteries, the original claim about non-decreasing probabilities stands. Clearly, the original claim is impugned by the fact that it conflicts with all cases in which probabilities *can* be assigned consistently.

Now the parallel here is certainly not strict, because the c-sceptic's original claim was not about probabilities. But it does suggest that c-scepticism is impugned by the fact that it conflicts with all cases in which probabilities *can* be consistently assigned.

In sum, the epistemological model employed in the present section is undeniably simple. But if the c-sceptic is to challenge its conclusions, she must explain what exactly is wrong with it—other than that it goes against orthodox Bayesianism.

6.3 S-scepticism and u-scepticism

Let's now examine how the discussion in §6.1 affects s-scepticism and u-scepticism. Although the notion of 'small' these scepticisms invoke may be vague, we can for the sake of argument pretend it's precise. So we may take small numbers to be all and only the first S numbers, from 0 to $S - 1$ inclusive. Cast in our probabilistic framework, both s-scepticism and u-scepticism can then be expressed as:

$$Pr(\forall x Fx) \lesssim Pr(\forall x Fx | (\forall x \leq S - 1) Fx)$$

In words: the probability that $\forall x Fx$ is true is approximately the same as, and only a little less than, the probability that $\forall x Fx$ is true given that its first S instances hold. Consequently, learning that the first S instances conform to the hypothesis renders the generalisation more probable, but only by a very small amount.

Consider now the natural numbers broken up into consecutive blocks. Let's say that the first block consists of the first S numbers, and that all other blocks consist of S numbers as well.³³ Thus the first block is made up of the numbers 0 to $S - 1$, the second block of the numbers S to $2S - 1$, and so on. Therefore instead of the blocks being individual numbers, like so:

³¹ To wit: one of the probabilities must be non-zero if their sum is to be 1. But if the probabilities are non-decreasing, all subsequent probabilities must be at least as large as this positive probability, so that the sum of all probabilities is infinite. Contradiction.

³² For example, let the probability that ticket i will win be $2^{-(i+1)}$.

³³ The argument still goes through even if subsequent blocks after the first are of different (non-empty) size.

$$\boxed{0} \quad \boxed{1} \quad \boxed{2} \quad \cdots \quad \boxed{n} \quad \cdots$$

each block comprises S -many numbers, like so:

$$\boxed{0, 1, 2, \dots, S - 1} \quad \boxed{S, S + 1, \dots, 2S - 1} \quad \cdots \quad \boxed{nS, nS + 1, \dots, (n + 1)S - 1} \quad \cdots$$

Now take the probability p_i to represent a subject’s prior credence that all the instances of an arithmetical conjecture in the i^{th} block of S numbers are confirmatory. *Observation*, in §6.1, still holds, this time with the p_i interpreted as the subject’s credence that all the elements in the i^{th} S -block conform to the conjecture. Now the probability p_0 in almost all relevant cases is less than 1, since we have not checked all small instances of the conjecture. So we can deduce from *Observation* that verifying the first S instances—all the small ones—provides more evidence for the hypothesis than verifying any later block of S numbers does, bar finitely many such blocks.

As a numerical illustration, suppose

$$0.7 = Pr(\forall x Fx) \lesssim Pr(\forall x Fx | (\forall x \leq S - 1) Fx) = 0.701$$

Thus p_0 , the prior probability that the first S -block conforms to the generalisation, equals $\frac{0.7}{0.701}$ (about 0.9986). Only finitely many p_i can be smaller than p_0 , the maximum number being the greatest k such that $(\frac{0.7}{0.701})^{(k+1)} > 0.7$. Since the maximum such k equals 248, at most 248 other S -blocks can have lower prior probability than the first. To put it another way, verifying the first S cases provides *more* evidence for the generalisation than verifying any other S -block in the partition, with the possible exception of no more than 248 of them.

How does this affect s-scepticism? According to the s-sceptic, inferences based on pointwise-small samples are potentially biased. In probabilistic terms, this means that p_0 —not equal to 1, since verifying the first S cases provides *some* evidence—should be fairly high. Yet, as we have seen, only finitely many p_i can be smaller than p_0 . If there are k of these, then all but k of the p_i are in fact higher than p_0 . This means that if we had to rank S -blocks in terms of how likely they are to provide counterexamples to the conjecture, the first S -block would be among the $k + 1$ most likely ones. It is hard to square this with s-scepticism, which maintains that small instances (i.e. those drawn from the first S -block) are potentially biased. For if the first block is biased then all but $(k + 1)$ of the infinitely many ones are, and in fact any block can play the role of the first one in this argument. So the right conclusion to draw seems to be that the size of a numerical instance is not in itself an indicator of potential bias.

Let’s turn finally to u-scepticism. According to the u-sceptic, inferences based on pointwise-small samples are potentially biased because they are unrepresentative. In probabilistic terms, this means that p_0 (< 1 , as explained) should be fairly high. The fact that only finitely many p_i can be smaller than p_0 is compatible with u-scepticism. For the u-sceptic can maintain that any S -block is unrepresentative, so all the p_i should be very close to 1. It is only when the members of several of these blocks have been verified that the conjecture as a whole is confirmed. And a virtue of u-scepticism, compared to c-scepticism and s-scepticism, is that it does not distinguish one S -block

from any other. This accurately reflects the inherent symmetry, since for instance a result such as *Observation* still holds however the S -blocks are permuted.

(There remains the question, of course, of the potential arbitrariness of the number S . Both the s-sceptic and the u-sceptic must justify their particular choice. But that is not the challenge we are considering here.)

Observe in passing that we have assumed the probabilistic independence of instances. But this is inessential: identical results can be established if we don't assume independence, and instead (in §6.1 terms) let p_N represent the probability of N being F given that the numbers 0 to $N - 1$ are all F .³⁴ If instances are verified in increasing order, an analogue of *Observation* follows, in its original §6.1 version or in its block version in this section.

In sum, the simple probabilistic model is incompatible with c-scepticism and sits uncomfortably with s-scepticism. But, so far as I can tell, it is consistent with u-scepticism. If this is along the right lines, the size-sceptic would be well advised to be a u-sceptic.

7 Other orders

In this section, we consider what happens when we vary the order on the natural numbers. The principal motivation for doing so was mentioned in §1. In philosophy of mathematics circles, one often hears the following argument casually expressed, approvingly or disapprovingly: 'we haven't found an inconsistency in ZFC in the past 100-plus years, so there isn't one'. You might play down this enumerative inductive evidence on account of the fact that all the ZFC-proofs we have come up with are in some sense small.

Despite its often being aired, I have found no extended discussions of the argument in the literature, merely scattered remarks. Hartry Field believes we can have inductive knowledge of the consistency of axiomatic theories (Field 1989) and adds, specifically about ZF, that 'if it weren't consistent someone would probably have discovered an inconsistency in it by now' (Field 1989, p. 232). In a discussion of Field's argument, Dummett voiced scepticism about the value of such evidence though, interestingly, not on size-sceptical grounds.³⁵ Crispin Wright (1994, sec. 4) similarly expressed dissent. More recently, and in a slightly more extended discussion than most, Dan Waxman (2017, p. 95) takes the enumerative inductive case for consistency to provide 'fairly weak' justification. John Burgess sees set theory's consistency as 'a reasonable assumption ... partly because by now we have long experience of working with the axiom system in daring ways without falling into contradiction' (2015, p. 116).³⁶ Something like Burgess's claim might also be read into a remark of Gödel's (1947, p. 519). John Mayberry takes the fact that set theorists have tried to find an inconsistency in systems much stronger than standard set theory to 'carry considerable weight as evidence' (1977, p. 164). Yet he takes it to be, in a somewhat opaque turn of phrase,

³⁴ Here we appeal to Bayes's Theorem.

³⁵ See Dummett (1991, p. 313), and Waxman (2017, p. 92) for decisive criticism of Dummett.

³⁶ 'Daring ways' of course points to something more than bare enumerative inductive inference.

‘entirely of a *practical* nature’ (1977, p. 164).³⁷ I too have expressed some sympathy with this sort of evidence for consistency (Paseau 2011). But no one to date has really examined the argument in any detail. Although our discussion cannot but remain preliminary, I shall bring one aspect of the question into more critical focus, namely, the value of the evidence that no hitherto-instantiated proof in set theory has yet resulted in a contradiction.

In earlier sections, we assumed that the natural numbers are ordered in the standard way. In this section, we’ll consider other orders. These arise from different orderings on countably infinite sets other than the natural numbers. Our leading example is the set of all past or present instantiated proofs of a given formal theory, including proof sketches.³⁸ These may be ordered in many different ways, of which we’ll consider three, concentrating on the third.³⁹ We extract some philosophical morals at the end of the section. We will not need to distinguish between c-, s- and u-scepticism, so speak of size-scepticism in an undifferentiated sense.

The first way of assigning numbers to proofs is by giving them Gödel numbers. For how to do so, open your favourite textbook on Gödel’s Incompleteness Theorems and adapt the numbering given there to the theory of interest. The resulting numbers may then be ordered in the usual way. The size of a proof will then vary according to the choice of Gödel numbering, so that proof P_1 is assigned a number smaller than P_2 according to one Gödel numbering but a larger number according to another. Not all Gödel numbers of proofs we have come across need be small; for example, they need not be minute in Baker’s technical sense (§2), since there is nothing stopping us from, say, assigning a symbol in a proof the non-minute Gödel number $10^{10^{10}}$. But Gödel numbers of instantiated proofs will all be small in a sense that can be made precise for the particular Gödel numbering used.

A second way of ordering proofs is by the number of steps they contain. Now this measure will depend to some extent on the presentation of the theory, in particular on the proof system. A proof in a natural-deduction system, for example, will typically contain a different number of steps from a similar one for the same conclusion and from the same axioms in a Hilbert-style system. Moreover, each type of system has many variants, depending for example on exactly which propositional connectives are used. Yet once a proof system has been chosen, the length of a proof can stand in for its numerical size, and the previous discussion carries over to proofs thus numbered.⁴⁰ Given a standard choice of system, it is safe to say that all proofs we will ever come across or produce are small in a sense that can be made precise, and certainly all minute in Baker’s technical sense.

The third way is to order proofs according to which types of axioms they use. Suppose for example you find ZFC’s Axiom Scheme of Replacement dubious for some reason but have no beef against the system’s other axioms. We may divide ZFC-

³⁷ Italics in the original. Mayberry contrasts practical reasons with mathematical ones.

³⁸ The class of concretely instantiated full-dress formal proofs, say on a blackboard, in a printed text or on a screen, is small compared to the much ampler class of proof sketches.

³⁹ This is not an exhaustive list. Another way, for example, might be to order proofs by their proof-theoretic complexity.

⁴⁰ The ordering on proofs will not be linear, since antisymmetry fails, as different proofs will be of the same length.

proofs into two categories, depending on whether they use instances of Replacement or not. Next, stipulate that Replacement-using proofs come later in the ordering than those that do not make essential use of Replacement, and employ some standard Gödel numbering within each category to order the proofs. If no proofs of the same category are given the same Gödel number, the resulting order would then be of type $\omega + \omega$. As we said, the philosophical justification for an order of this kind would be that the Replacement scheme is suspect in some way, so that proofs that avoid it—the first ω ones—are simpler or more credible. They therefore appear earlier in the ordering than any proofs that make essential use of Replacement. For another example, consider a system of second-order arithmetic. Here one could stipulate infinitely many categories, a proof of category n being one that uses a Σ_n -induction axiom but no Σ_m -induction axiom for any $m > n$. We may further stipulate that any proof of category i comes before any proof in category j in the ordering iff $i < j$, and number the proofs within each category using some standard Gödel numbering. The resulting order on proofs would be of type $\omega + \omega + \dots + \omega + \dots$, i.e. ω^2 .

We might wonder how in this type of case we should go about working out the probability of the universal generalisation $\forall n F(n)$ in terms of the probability of each of its instances. The instances may be indexed by elements of set I , with an order \triangleleft on it, p_i representing the probability that the i^{th} instance conforms to the hypothesis. So the question is: what is $\prod_{i \in I} p_i$? (Assuming that the index set I is countably infinite and each p_i is in $[0, 1]$.) There is in fact a natural way to define this product, which at first sight seems to depend on an arbitrary choice but upon inspection turns out not to do so.

The natural definition of $\prod_{i \in I} p_i$ is as follows. Since I is countably infinite, there exists a bijection b from I to the set of natural numbers. We may define $\prod_{i \in I} p_i$ as $\prod_{j=0}^{\infty} p_{(b^{-1}(j))}$ with the infinite product defined as the limit of its finite approximations, as usual, i.e. $\prod_{j=0}^{\infty} p_{(b^{-1}(j))} = \lim_{N \rightarrow \infty} \prod_{j=0}^N p_{(b^{-1}(j))}$. This is well-defined because each p_i is in $[0, 1]$. The definition is natural because the choice of bijection does not matter: the product is identical whatever bijection b we choose.⁴¹ To put it informally, ordering the probabilities in any way we like does not affect their product.

The method just described of defining $\prod_{i \in I} p_i$ for I an arbitrary countably infinite set (and each p_i in $[0, 1]$) is entirely general and does not depend on I 's ordering \triangleleft . When

⁴¹ Proof: let b_1 and b_2 be two bijections from I to \mathbb{N} and let $l_i = \lim_{N \rightarrow \infty} \prod_{j=0}^N p_{(b_i^{-1}(j))}$ for $i = 1, 2$. Clearly, l_1, l_2 are in $[0, 1]$ and for any $N, l_1 \leq \prod_{j=0}^N p_{(b_1^{-1}(j))}$ ($i = 1, 2$) as the probabilities p_i are all in $[0, 1]$. Suppose w.l.o.g. that $l_1 < l_2$, and that $\epsilon > 0$ and N_ϵ are such that $\epsilon < |l_1 - l_2|$ and $\prod_{j=0}^{N_\epsilon} p_{(b_1^{-1}(j))} < l_1 + \epsilon$. Let $M_\epsilon = \max\{b_2(b_1^{-1}(j)) : 1 \leq j \leq N_\epsilon\}$. The following inequalities hold: $\prod_{j=0}^{M_\epsilon} p_{(b_2^{-1}(j))} < l_1 + \epsilon < l_2$. But by the definition of l_2 , we also have $l_2 \leq \prod_{j=0}^{M_\epsilon} p_{(b_2^{-1}(j))}$. Contradiction. Hence $l_1 = l_2$ and the choice of bijection does not matter.

\triangleleft happens to be a well-order, another natural definition of $\prod_{i \in I} p_i$ is available. The idea is to exploit the well-ordering \triangleleft by multiplying probabilities together at successor ordinals and by taking limits at limit ordinals.⁴² A proof similar to the proof that the choice of bijection in the general case doesn't matter establishes that this definition yields the same limit as in the general case.⁴³

It should be clear that if \triangleleft is not an ω -order, there is in general no reason to suppose that the evidence must be frontloaded in the sense of §5 or §6. Suppose for instance that \triangleleft has order-type $\omega + \omega$. In that case, the probability of a counterexample to the generalisation in question appearing in the first ω -block may be close to 0, yet the probability of its appearing in the second ω -block may be much higher. In this scenario, a sort of size-scepticism would be justified, since any enumerative inductive evidence from the first ω -block in favour of the generalisation would count for little. In the ZFC-example mentioned earlier, such proofs would correspond to those which do not use instances of Replacement, which in this example are regarded as more dubious. For a very similar example, replace instances of Replacement with the Axiom of Infinity, which a certain type of finitist might suspect of leading to contradiction. Proofs that do not use Infinity would all be considered 'small', as they are all in the first ω -block of the $\omega + \omega$ ordering. The finitist in question would thus regard them as constituting

⁴² Suppose the well-ordering is of order-type γ for some ordinal $\omega \leq \gamma < \omega_1$, and that indices have been relabelled to correspond to ordinals in the well-order. We define the product as follows (where λ is a limit ordinal):

$$\begin{aligned} \prod_{\xi < 0} p_\xi &= 1 \\ \prod_{\xi < \beta^+} p_\xi &= p_\beta \cdot \prod_{\xi < \beta} p_\xi \\ \prod_{\xi < \lambda} p_\xi &= \lim_{\beta \rightarrow \lambda} \prod_{\xi < \beta} p_\xi \end{aligned}$$

In the third equation, which represents the limit case, the limit taken in the equation's right-hand side is in \mathbb{R} , of course. Now $\prod_{\xi < \beta} p_\xi$ for ordinals $\beta < \gamma$ is a weakly decreasing sequence of real numbers bounded below by 0, since $0 \leq \prod_{\xi < \beta} p_\xi \leq \prod_{\xi < \alpha} p_\xi \leq 1$ if $\alpha < \beta$. So all limits in the transfinite recursion are well-defined. We may then simply define $\prod_{i \in I} p_i$ as $\prod_{\xi < \gamma} p_\xi$.

⁴³ It is crucial for these definitions that all the probabilities p_i lie in $[0, 1]$. Things would be otherwise if they were not. Consider for example the set $\{\frac{n+2}{n+1} : n \in \mathbb{N}\} \cup \{\frac{n+1}{n+2} : n \in \mathbb{N}\}$. Arranged as the ω -sequence

$$\frac{2}{1} \cdot \frac{1}{2} \cdot \frac{3}{2} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{3}{4} \cdot \frac{5}{4} \cdot \frac{4}{5} \cdot \frac{6}{5} \cdot \frac{5}{6} \dots,$$

these numbers' infinitary product is 1 according to the definition of a product for a well-order. (Argument: the product of the first $2n$ is 1, that of the first $2n + 1$ is $\frac{n+2}{n+1}$, so the limit of the finite products is 1.) Whereas arranged as the $\omega + \omega$ -sequence

$$\frac{2}{1} \cdot \frac{3}{2} \cdot \frac{4}{3} \cdot \frac{5}{4} \cdot \frac{6}{5} \dots \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{4}{5} \cdot \frac{5}{6} \dots,$$

they do not define a product. (Argument: the product of the first n is $n + 1$, so the limit of the initial ω -product does not exist.) Finally, arranged as the ω -sequence

$$\frac{1}{2} \cdot \frac{2}{3} \cdot \frac{3}{4} \cdot \frac{4}{5} \cdot \frac{5}{6} \dots \frac{2}{1} \cdot \frac{3}{2} \cdot \frac{4}{3} \cdot \frac{5}{4} \cdot \frac{6}{5} \dots,$$

their infinitary product is 0. (Argument: the product of the first n is $\frac{1}{n+1}$, so the product of the first ω is 0. It then remains 0 thereafter.) Not so when all the numbers are in $[0, 1]$: in this case, the product is well-defined whatever the order.

little, if any, evidence for ZFC's consistency. For a third example—perhaps not a very realistic one but one that illustrates the point neatly—suppose \triangleleft is of order-type $\omega + 1$, with the last instance being the most controversial, more so than the first ω ones. In this last case, any 'small' enumerative inductive evidence—drawn from the first ω cases—would count for less than the single instance that is the final case.

None of this should be surprising, however. If initial credibility affects an instance's place in the order, then we are 'baking in' a certain kind of order bias. Putting more dubious proofs later in the order forces early cases to be easy cases.

What's the upshot? Once we move beyond the natural numbers and consider the class of proofs in a formal system, the picture becomes more complicated. In summary:

1. Suppose we order proofs in the first way, by assigning them Gödel numbers in the customary fashion. You might think that size-scepticism about the enumerative evidence for consistency would stand or fall with size-scepticism about arithmetical evidence. But this is moot. For suppose you agree with Frege that position in the number series matters, unlike position in space (see §2). If you buy that thought, you may well discern a potentially relevant difference between numbers and proofs. The position of a proof in the sequence ordered by a Gödel number *is* a matter of indifference, since the numbering is arbitrary along several dimensions and extrinsic to the proof itself. But this is quite unlike the position of a number in the number series. Size-scepticism about the evidential value of non-contradictory proofs is correspondingly less attractive.
2. Suppose proofs are ordered by the number of steps they contain. The situation is then very similar to the arithmetical case. And observe that a proof's length is a fairly natural number to associate with it, unlike any Gödel number we might assign it. But it's much less clear, to me at any rate, that 'the length of a proof is not a matter of indifference, like position in space', to adapt Frege's words. Perhaps the length of a proof is more analogous to proximity in space than it is to numerical size. In any event, the case for size-scepticism here remains to be made.
3. Suppose finally that we order proofs in the third way, according to the types of axioms they use. Whether size-scepticism gets a foothold depends entirely on *how* the proofs are ordered. We could 'bake in' size-scepticism, or avoid it, depending on the ordering chosen.

A fuller discussion building on these preliminary remarks would have to consider the value of the enumerative inductive evidence for different set theories, and how it complements other non-deductive evidence. The non-deductive evidence for ZFC includes for example the fact that no contradiction has ever been derived from its axioms, as well as the fact that people have tested it in 'daring ways', as John Burgess put it, together with the fact that ZFC has an intuitive model—the iterative universe of sets. Non-iterative set theories or stronger iterative set theories such as ZFC + 'there exists a supercompact cardinal' have rather different non-deductive support. We leave these investigations for a further occasion.

8 Conclusion

Many mathematicians and philosophers are sceptical about the value of enumerative inductive evidence in arithmetic, especially in the absence of further supporting evidence. ‘Size-scepticism’ of this sort is justified by the following thought: enumerative inductive evidence for an arithmetical conjecture consists exclusively of ‘small’ instances—those that appear very early on, in some sense, in the natural number sequence. This essay set out two sorts of cases in which size-scepticism does not come into play (§3), distinguished three varieties of size-scepticism (§4), and raised and assessed some challenges for all three (§§5–6). We saw that the strength of these challenges depended sensitively on the variety of size-scepticism in question. Finally, §7 contained some preliminary remarks about enumerative inductive evidence for the consistency of set theory.⁴⁴

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Baker, A. (2007). Is There a Problem of Induction for Mathematics? In M. Leng, A. Paseau, & M. Potter (Eds.), *Mathematical Knowledge* (pp. 59–73). : Oxford University Press.
- Borwein, P., Choi, S., Rooney, B., & Weirathmueller, A. (2008). *The Riemann Hypothesis: A Resource for the Afficionado and Virtuoso Alike*. : Springer.
- Burgess, J. P. (2015). *Rigor and Structure*. : Oxford University Press.
- Dummett, M. (1991). *Frege: Philosophy of Mathematics*. : Duckworth.
- Easwaran, K. (2011). Bayesianism II: Applications and Criticisms. *Philosophy Compass*, 6, 321–332.
- Echeverría, J. (1996). Empirical Methods in Mathematics. A Case-Study: Goldbach’s Conjecture. In G. Munévar (Ed.), *Spanish Studies in the Philosophy of Science* (pp. 19–55). : Kluwer.
- Field, H. (1989). *Realism. Mathematics and Modality*. : Oxford University Press.
- Franklin, J. (1987). Non-deductive Logic in Mathematics. *The British Journal for the Philosophy of Science*, 38, 1–18.
- Gaifman, H. (2004). Reasoning with Limited Resources and Assigning Probabilities to Arithmetical Statements. *Synthese*, 140, 97–119.
- Garber, D. (1983). Old Evidence and Logical Omniscience in Bayesian Confirmation Theory. In J. Earman (Ed.), *Testing Scientific Theories* (pp. 99–132). : University of Minnesota Press.
- Gödel, K. (1947). What is Cantor’s Continuum Problem? *American Mathematical Monthly*, 54, 515–525.
- Hacking, I. (1967). Slightly More Realistic Personal Probability. *Philosophy of Science*, 34, 311–325.
- Mayberry, J. (1977). The Consistency Problem for Set Theory: An Essay on the Cantorian Foundations of Mathematics (II). *British Journal for the Philosophy of Science*, 28, 137–170.
- Mill, J. S. (1843/1973). A System of Logic. In J. M. Robson (Ed.), *Collected Works*. : University of Toronto Press.

⁴⁴ I would like to thank Alan Baker, Dan Waxman, Jeffrey Schatz, Sorin Bangu and two anonymous *Synthese* referees for comments on an earlier draft. I completed this paper while holding a Research Fellowship awarded by the Leverhulme Trust, whose support I gratefully acknowledge.

- Norton, J. (2010). 'A Survey of Inductive Generalization', pp. 144, unpublished ms. available from <http://www.pitt.edu/~jdnorton>, accessed between October 2014 and January 2020.
- Ochem, P., & Rao, M. (2012). Odd Perfect Numbers are Greater than 10^{1500} . *Mathematics of Computation*, 81, 1869–1877.
- Paseau, A. C. (2011). 'Mathematical Instrumentalism, Gödel's Theorem and Inductive Evidence', *studies in the history and philosophy of science*, 42, 140–9.
- Paseau, A. C. (2015). Knowledge of Mathematics without Proof. *The British Journal for the Philosophy of Science*, 66, 775–99.
- Skipper, M. & Bjerring, J.C. (2020), 'Bayesianism for Non-ideal Agents', *Erkenntnis*, vol. & pp. tbc.
- Steiner, M. (1975). *Mathematical Knowledge*. : Princeton University Press.
- Walsh, S. (2014). Empiricism, Probability, and Knowledge of Arithmetic: A Preliminary Defense. *Journal of Applied Logic*, 12, 319–348.
- Waxman, D. (2017). *Freedom, Truth, and Consistency*, PhD dissertation, New York University.
- Wright, C. (1994). About "The Philosophical Significance of Gödel's Theorem": Some Issues. In B. F. McGuinness & G. Oliveri (Eds.), *The Philosophy of Michael Dummett, Synthese Library* (Vol. 239, pp. 167–202). : Kluwer.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.