Erkenn (2009) 70:211–235
DOI 10.1007/s10670-009-9154-1

ORIGINAL ARTICLE

Fifteen Arguments Against Hypothetical Frequentism

Alan Hájek

Received: 28 November 2008 / Accepted: 16 January 2009 / Published online: 10 February 2009

Abstract This is the sequel to my "Fifteen Arguments Against Finite Frequentism" (*Erkenntnis* 1997), the second half of a long paper that attacks the two main forms of frequentism about probability. Hypothetical frequentism asserts:

> The probability of an attribute *A* in a reference class *B* is *p*
> iff
> the limit of the relative frequency of *A*'s among the *B*'s would be *p* if there were an infinite sequence of *B*'s.

I offer fifteen arguments against this analysis. I consider various frequentist responses, which I argue ultimately fail. I end with a positive proposal of my own, 'hyper-hypothetical frequentism', which I argue avoids several of the problems with hypothetical frequentism. It identifies probability with relative frequency in a hyperfinite sequence of trials. However, I argue that this account also fails, and that the prospects for frequentism are dim.

Prologue

Over a decade ago, in a flurry of youthful zeal, I wrote a paper called "Thirty Arguments Against Frequentism" for the wonderful 3rd Luino Conference on Probability, Dynamics and Causality, organized by Domenico Costantini and Maria-Carla Galavotti. The conference was held in honor of Richard Jeffrey, and his paper "Mises Redux" (1992), a famous critique of the frequentist interpretation of probability, provided the inspiration for mine. The conference proceedings eventually appeared in *Erkenntnis* Vol. 45 (1997), and they were reprinted in

A. Hájek (✉)
Philosophy Program, Research School of Social Sciences, Australian National University, Canberra ACT 0200, Australia
e-mail: alan.hajek@anu.edu.au

Costantini and Galavotti (1997). In my original paper I distinguished two versions of frequentism—what I called *finite* frequentism, and *hypothetical* frequentism, and I marshaled fifteen arguments against each. Unfortunately, my paper was roughly twice the length allowed for the publications. Fortunately, this problem was easily solved—I simply cut the paper into two halves, one on each version of frequentism, and I submitted the first half! (Hájek 1997.) The second half remained on my computer's hard drive and in the hands of a few interested folk who solicited copies of it—until now.

It's a slightly curious feeling revisiting and revising that paper. It's like co-authoring a paper with someone with whom I agree mostly but not entirely, and with the co-author denied the right of reply. Perhaps I have mellowed in the intervening years, but I now see some frequentist responses that I didn't see then, as this version of the paper shows. I even offer the frequentist a positive proposal, *hyper-hypothetical frequentism*, which I argue avoids several of the problems with hypothetical frequentism. That said, I stand by all the main points of the original article, so I am happy to reprise them here. Anyway, I hereby appoint my current time-slice as the senior author; but all mistakes should be blamed on my younger self!

Hypothetical Frequentism

Probability is *long run* relative frequency—or so it is said. Here is a well-known engineering textbook saying so: "The probability of an event (or outcome) is the proportion of times the event would occur in a long run of repeated experiments" (Johnson 1994, p. 57). So the probability that the coin lands Heads is the relative frequency of Heads in a long run of tosses of the coin, the probability that the radium atom decays in 1,600 years is the relative frequency of such atoms that so decay in a long sequence of such atoms, and so on. What if the world is not generous enough actually to provide a long run of the relevant sequence of events? And how long is 'long', in any case? We can circumvent both concerns with a single stroke, by going *hypothetical*—by considering what things *would* be like if the run in question were of any length that we care to specify. (Notice the "would" in the quote above.) And since we are going hypothetical, we might as well make the most of it and consider the longest run possible: an infinite run. After all, whatever vagueness there may be in 'long run', an infinite run surely counts. So let us give as broad a characterization of hypothetical frequentism as we can, consistent with this commitment. It asserts:

(HF) The probability of an attribute A in a reference class B is p

iff

the limit of the relative frequency of occurrences of A within B would be p if B were infinite.

This characterization is meant to subsume various more specific accounts—for example, those of Reichenbach (1949) and von Mises (1957), the latter endorsed by Howson and Urbach (1993).

This account of probability will be my target—15 times over. Why so many arguments? Is this an exercise in overkill? Or worse, is it an exercise in *under*kill, my deployment of so many arguments betraying a lack of faith that any one of them actually does the job? On the contrary, as in *Murder On the Orient Express*, I think that many of the blows may well be fatal on their own (although in the book, the victim only received twelve of them). But there is good reason to accumulate the arguments: they successively cut off hypothetical frequentism's escape routes, making the addition of an epicycle here or a further clause there less and less promising. For example, some of the arguments work in tandem, setting up dilemmas for hypothetical frequentism: if it dodges one argument by retreating one way, another argument awaits it there.[1] Moreover, different frequentists may have different ambitions for their theory. Some might offer it as an analysis of the concept of 'objective probability'. Others might regard it as an explication, allowing that it might precisify or otherwise slightly revise a messy, pre-scientific concept so that it is fit for science. Still others might be content to identify a core usage of the words 'objective probability' in our theorizing, or some central strand of our thinking about it, and seek just to capture that. And so on. By piling on argument after argument, we thwart more and more of these ambitions. When we are done, I hope to have shown that probability is not recognizably a frequentist notion, however we squint at it.

Despite its eventual and overdetermined demise, hypothetical frequentism is a worthy target. It is certainly entrenched, arguably the most entrenched of all the major interpretations of probability. Just ask your typical physicist or engineer what *they* think probability is, if you need any convincing of this. It still has some currency among philosophers nowadays—I have already mentioned Howson and Urbach's endorsement of it. It seems to render precise our folk understanding of probability as having a close connection to long run frequencies, even when those long runs are not actualized. It resonates with the class of limit theorems known as the 'laws of large numbers'—more on that shortly. Moreover, hypothetical frequentism is recognizable in more recent frequency-based philosophical accounts of probability—notably van Fraassen's (1980) 'modal frequency' account. And there are ghosts of hypothetical frequentism in Lewis's (and his followers') 'best systems' accounts (Lewis 1994). They say, roughly, that objective probabilities ('chances') are given by indeterministic laws as stated by the theory of the universe that best combines simplicity, strength, and fit to the data. I believe that collectively the arguments here cast some doubt on the viability of these more recent accounts, too, so I hope the interest of the arguments extends beyond the narrower focus of this paper.

Since the target is worthy, it is worth pursuing at least some of the epicycles or further clauses, to see how they play out. Along the way I will suggest ways in which frequentism can be buttressed in the face of the problems that I point out. However, buttressing frequentism is one thing, saving it another.

I will present first some more broadly philosophical arguments, then more precise, mathematical arguments. Not all of them are original to me, although when

[1] Kenny Easwaran suggested that I might offer instead "Seven and a Half Dilemmas for Hypothetical Frequentism"!

they are not I think at least I have something original to say about them. And I hope it will be useful to have them gathered alongside more original arguments, so that the case against hypothetical frequentism can be assessed in its entirety.

Enough preliminaries; onwards to

The Arguments

1. *An Abandonment of Empiricism*

Frequentism has laudable empiricist motivations, and frequentists have typically had an eye on scientific applications of probability. Finite frequentism, the target of this paper's predecessor, admirably respected those motivations, its other failings notwithstanding. It says:

The probability of an attribute A in a finite reference class B is the relative frequency of actual occurrences of A within B.

But hypothetical frequentism makes two modifications of that account that ought to make an empiricist uneasy: its invocation of a *limit*, and of a *counterfactual*. Regarding the limit, any finite sequence—which is, after all, all we ever see—puts no constraint whatsoever on the limiting relative frequency of some attribute. Limiting relative frequencies are unobservable in a strong sense: improving our measuring instruments or our eyesight as much as we like would not help us ascertain them. Imagine what it would take to observe the limiting relative frequency of Heads in an infinite sequence of coin tosses—the observers would have to live forever, or be able to complete successive tosses in exponentially shorter intervals, Zeno-like, so that all the tosses are completed in a finite period of time. Hume would turn in his grave.

Related, finite frequentism made congenial the *epistemology* of probability. One can easily know the relative frequency of Heads in some finite sequence—it's as easy as watching, counting, and dividing. But how can one know what the limiting relative frequency of Heads would be in a hypothetical infinite sequence? To be sure, science appeals to quantities that are defined in terms of limits—think of velocity, or acceleration, or power—and we take ourselves to be able to know the values of such quantities, well enough. But the value of a limiting hypothetical relative frequency is unknowable in the strongest sense, for the reasons in the previous paragraph, and also since there is necessarily no fact of the matter of this value (as I will shortly argue), and knowledge is factive.

A commonly made criticism of one early version of Carnapian logical probability (1950), $c^{\dagger}$, is that it 'does not learn from experience': evidence regarding the properties of certain individuals does not change prior probabilities for other individuals having those properties. But to the extent that this is a problem, it affects frequentism—the interpretation whose very motivation is to take evidence seriously. Indeed, finite frequentism takes evidence *so* seriously that it conflates a certain kind of good evidence for a probability claim for the truth-maker of the claim itself. But ironically, hypothetical frequentism seems to suffer from $c^{\dagger}$'s problem more than $c^{\dagger}$

itself does. For while Carnap's $c^{\dagger}(h, e)$ is a function of e, and in that sense surely *does* take the evidence seriously, finite strings of data put absolutely no constraint on the hypothetical frequentist's probability. But again, finite strings of data are all that we ever see.

The frequentist could try to relieve this problem by restricting his account, only according probabilities to attributes whose relative frequencies converge *quickly* (in some sense to be spelled out more precisely). For such attributes, finite strings of data could be a good guide to the corresponding limiting behavior after all.[2] For example, there are many hypothetical sequences of coin tosses for which the relative frequency of Heads beyond the 100th toss never deviates by more than 0.1 from ½. In those sequences, the first 100 tosses reflect well the limiting behavior. More generally, the frequentist might choose an $\varepsilon > 0$ and an N such that the only attributes that have probabilities are those whose hypothetical relative frequencies beyond N remain within ε of their limiting values. This proposal arguably would help with the epistemological problem that I have raised: if a probability exists at all, then it can be ascertained fairly accurately with a comparatively small number of observations. But the proposal renders worse the problem, which we will see shortly, of probabilities being undefined when intuitively they should be defined—see argument 5. And it shifts the epistemological problem to one of knowing *whether* a probability exists at all, which finite frequentism at least made tractable.

Nor should the appeal to HF's counterfactual sit well with frequentism's forefathers: frequentism has gone modal. The frequentist can no longer obviously take the philosophical high ground when compared to a propensity theorist, who sees probabilities as certain dispositions. After all, dispositions are typically closely linked to counterfactuals about behavior under appropriate circumstances: to say that a piece of salt is soluble is (roughly[3]) to say that it would dissolve if it were placed in water, and so on. Hypothetical frequentism's counterfactual has a similar ring to it. In fact, it has a worse ring to it, by empiricist lights—a death knell.

2. *The Counterfactuals Appealed to are Utterly Bizarre*

For HF isn't just some innocent, innocuous counterfactual. It is infinitely more far-fetched than the solubility counterfactual. To focus our discussion, let us think of counterfactuals as being analysed in terms of a Stalnaker/Lewis-style possible worlds semantics (Stalnaker 1968, Lewis 1973). Taking hypothetical frequentism's statement literally—and I don't know how else to take it—we are supposed to imagine a world in which an infinite sequence of the relevant attribute occurs. But for almost any attribute you can think of, any world in which *that* is the case would have to be *very* different from the actual world. Consider the radium atom's decay. We are supposed to imagine infinitely many radium atoms: that is, a world in which there is an infinite amount of matter (and not just the 10^{80} or so atoms that populate the actual universe, according to a recent census). Consider the coin toss. We are supposed to imagine infinitely many results of tossing the coin: that is, a world in which

[2] Thanks here to Kenny Easwaran.

[3] I ignore various subtleties—e.g. so called 'finkish' dispositions.

which coins are 'immortal', lasting forever, coin-tossers are immortal and never tire of tossing (or something similar, anyway), or else in which coin tosses can be completed in ever shorter intervals of time… In short, we are supposed to imagine *utterly bizarre* worlds—perhaps worlds in which entropy does not increase over time, for instance, or in which special relativity is violated in spectacular ways. In any case, they sound like worlds in which the laws of physics (and the laws of biology and psychology?) are quite different to what they actually are. But if the chances are closely connected to the laws, as seems reasonable, and the laws are so different, then surely the chances could then be quite different, too.[4]

Note also a further consequence for the world that we are supposed to consider here. If there are infinitely many events of a given sort in a single world, then either time is continuous (as opposed to quantized), or infinite in at least one direction, or infinitely many events are simultaneous, or we have a Zeno-like compression of the events into smaller and smaller intervals. In any case, I find it odd that there should be such extravagant consequences for the truth-makers of probability statements in the actual world.

So what goes on in these worlds seems to be entirely irrelevant to facts about this world. Moreover, we are supposed to have intuitions about what happens in these worlds—for example, that the limiting relative frequency of Heads would be 1/2 in a nearest world in which the coin is tossed forever. But intuitions that we have developed in the actual world will be a poor guide to what goes on in these remote worlds. Who knows what would happen in such a world? And our confidence that the limiting relative frequency really would be 1/2 surely derives from actual-worldly facts—the actual symmetry of the coin, the behavior of other similar coins in actual tosses, or what have you—so it is really *those* facts that underpin our intuition.

Readers of Kripke (1980) will recognize a parallel here to his famous argument against a dispositional analysis of meaning. Kripke's skeptic challenges you to come up with a fact that determines that you should now compute the *plus* function rather than the *quus* function in order to accord with your past intentions. Response: your past *dispositions* constrain what you should now do. But as Kripke points out, you had only a finite set of dispositions, being a mortal being with a finite mind, so that underdetermines what you should do now. Response: then let's imagine away those limitations, considering instead what would be true if you had an infinite brain… But Kripke replies—and now the parallel should be clear—who knows what would be the case under such a bizarre supposition? It seems that any plausibility the infinite-case counterfactual has, it derives from the finite case.

[4] It only adds to the bizarreness if we add that the counterfactually repeated trials are to be *'identically prepared',* a phrase one sometimes hears the frequentist add. Actual repeated events differ in so many ways from each other—removing all of these respects of difference takes us still further from actuality. We can countenance probabilities involving certain supernova explosions, for instance; but can we even imagine the 'identical preparation' of a sequence of supernova explosions—let alone an infinite sequence of them? In fact, if the identity of indiscernibles is a necessary truth, then such identical preparation may be downright impossible.

Now, it may be objected to Kripke that the infinite brain is an idealization, much like an ideal gas or a frictionless plane, and that such idealization is a familiar part of science.[5] So it is; but I don't see how talk of idealization rescues hypothetical frequentism, regarded as an *analysis* of probability. HF asserts a biconditional (one which presumably is supposed to be necessarily true). According to it, the coin in my pocket lands Heads with probability 1/2 *if and only if* a certain bizarre counterfactual about a very different coin is true. I am puzzled by the claim that the two sides of this biconditional (necessarily) have the same truth value; indeed, I flatly deny it. Is my puzzlement supposed to vanish when I am told that HF involves an *idealization*? Does the biconditional suddenly become true if we think of it as an *idealization*? I don't know what that would even mean.

Perhaps I am not doing justice to the role that idealization might play in hypothetical frequentism; I did, after all, allow that different frequentists might have different ambitions for their theory. Very well then; perhaps they have an answer to this argument. But I also said that I have other arguments as backup, poised to scotch those ambitions. So let me continue.

3. *There is no Fact of what the Hypothetical Sequences Look Like*

In fact the problem for the hypothetical frequentist's counterfactual is still worse than it was for the 'infinite brain' counterfactual. Here is a coin that is tossed exactly once, and it lands Heads. How would it have landed if tossed infinitely many times? Never mind that—let's answer a seemingly easier question: how would it have landed on the second toss? Suppose you say "Heads". Why *Heads*? The coin equally could have landed Tails, so I say that it would have. We can each pound the table if we like, but we can't both be right. More than that: neither of us can be right. For to give a definite answer as to how a chancy device would behave is to misunderstand chance.

Here I am echoing one of Jeffrey's (1992) main arguments against frequentism. In his words, "there is no telling whether the coin would have landed head up on a toss that never takes place. That's what probability is all about" (193). He says that it's like asking: what would be the mass of a tenth planet?[6] There's no fact of the matter. In fact, I would go further than him on this point: it's *worse* than asking that question. At least it's consistent with the concept of mass that we could answer the question about the tenth planet; and perhaps cosmologists could point to facts about the distribution of nearby matter, or what have you, that would dictate a sensible answer. But to say that there is a fact of the matter of how the second toss would land is to deny that the coin is a chancy system, whereas the point of the example is exactly that.

The frequentist will be quick to point out that what matters is not what the next toss would be, but rather what the limiting relative frequency would be. He might even add: while there may be no fact of the matter of how the coin would land on any given trial, there *is* a fact of the matter of what the limiting relative frequency

5 Thanks here to Darren Bradley.

6 Jeffrey wrote before Pluto arguably lost its title of being a planet; and before Eris arguably gained *its* title.

would be, namely one half. After all, he continues, all the nearest possible worlds in which the coin is tossed forever agree on that fact.[7]

I contend that this is not so. This is the upshot of the next few arguments.

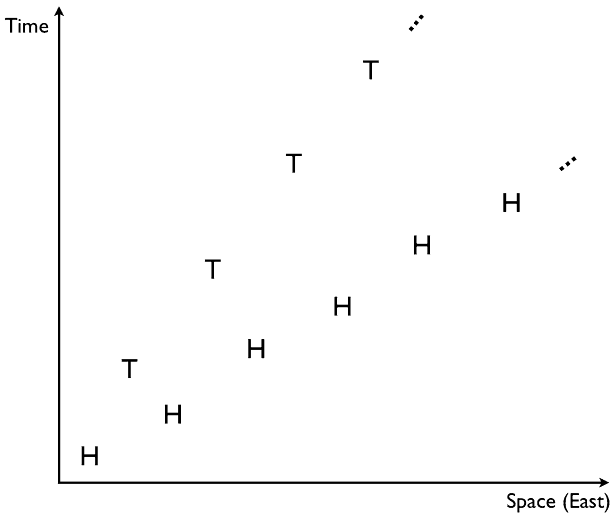
A good strategy for arguing against an analysis that involves a definite description of the form '*the F* such that…' is to argue against the presupposition that there is exactly one such *F*. There are two ways to do this:

1) argue that there could be more than one such *F*; and
2) argue that there could be less than one.

Sometimes one can do both, and one can here. HF invokes a definite description: *the* limit of the relative frequency of *A*'s among the *B*'s … I will now challenge it in both these ways.

4. *There Could be More than One Limiting Relative Frequency: The Problem of Ordering*

Suppose that I am tossing a coin on a train that is moving back and forth on tracks that point in a generally easterly direction. Suppose that the results happen to fall in a pattern as depicted in the 'space-time' diagram below. Think of the horizontal axis as the west-east spatial dimension, and the vertical axis as the temporal dimension.



Moving from left to right (west to east), we see the pattern: HTHTHTHTH… Moving upwards (earlier to later), we see the pattern: HHTHHTHHT… Imagine,

[7] The frequentist's idea here is somewhat reminiscent of the supervaluational treatment of vagueness. Even though there might be no determinate fact of the matter of which hair makes the difference between baldness and non-baldness, with different 'valuations' disagreeing on it, it is determinately true that everyone is bald-or-not-bald, since this is true on *all* valuations. But as we will shortly see, the analogy to supervaluating does not go through, much as the frequentist might like it to: it is *not* true that the limiting relative frequency is the same in *all* the nearest possible worlds.

as we can, that these patterns persist forever. What is the limiting relative frequency of Heads? Taking the results in their temporal order, the answer is 2/3, and I suppose this is the answer that the frequentist would be tempted to give. But taking them in their west-east spatial order, the answer is 1/2. Now, why should one answer have priority over the other? In other words, we have more than one limiting relative frequency, depending on which spatio-temporal dimension we privilege.

We can imagine making matters still worse for the frequentist. Now suppose that the train has an elevator inside, and that the results taken in the up-down spatial dimension happen to follow the pattern HTTHTTHTT… (I won't attempt to draw a picture of this!) Then the limiting relative frequency of Heads in that dimension is 1/3. Yet that is apparently an equally natural ordering of the results.

It is arbitrary to select, say, the temporal ordering of the results as the one for which the limiting relative frequency is to be ascertained. Indeed, it seems worse than arbitrary: for we often think of chances as themselves evolving over time—more on that later—but if we privilege the temporal ordering, there is a worry that the limiting relative frequencies will not so evolve. Moreover, suppose that all the trials of some experiment are simultaneous—there is a world, for example, in which infinitely many radium atoms are created at once, setting up infinitely opportunities for them to decay within 1,600 years, side by side. In that case they don't have a temporal ordering at all; how, then, should they be ordered? And as Einstein has taught us, one should really say 'simultaneous in a certain *reference frame*'. But that only confounds the frequentist further: by changing reference frame, one can change temporal orders of events—and possibly the limiting relative frequency, much as I did by changing the choice of dimension above.

The upshot is this. *Limiting relative frequency depends on the order of trials, whereas probability does not. They therefore cannot be the same thing.*

Frequentism has long been regarded as foundering on the reference class problem: its probability assignments must be relativized to a *reference class* of outcomes. Hypothetical frequentism founders on a further problem: even after we have settled upon a reference class, there is the further problem of ordering it. We might call this the *reference sequence problem*. Hypothetical frequentism turns the one-place property 'having probability p' of an event, into the *three*-place relation 'having limiting relative frequency p relative to reference class R, according to ordering O'.

The reference sequence problem also brings home how serious is the problem with HF's counterfactual: it makes no sense in general to say what the limiting relative frequency of some attribute *would* be, because different orderings will yield different answers.

In Sect. 2 I argued that HF requires the frequentist to countenance utterly bizarre worlds in which infinitely many trials take place. The frequentist could, I suppose, modify his account so that the infinitely many trials are not confined to a single world, but rather are taken *across* infinitely many worlds.[8] That might solve the problem raised in that section: none of the worlds need have an extravagant

[8] Thanks here to Kenny Easwaran.

ontology. However, it would do so by making the problem raised in *this* section worse: trivial cases aside, different *orderings* of the worlds will yield different limiting relative frequencies for a given attribute, and it is hard to see how any particular ordering could be privileged.

5. *There Could be Less than One Limiting Relative Frequency: the Limit may not Exist*

Suppose that we have a privileged ordering of the results of tossing a fair coin—as it might be, the temporal ordering in our frame of reference. Then, despite its fairness—or better, *because* of its fairness—there is no ruling out the sequence of results:

HT HHTT HHHHTTTT HHHHHHHHTTTTTTTT…

Here we have sequences of 2^n Heads followed by sequences of 2^n Tails. There is no limiting relative frequency for Heads here: the relative frequency sequence oscillates wildly, growing steadily through each run of Heads, then decreasing steadily through each run of Tails; and the runs are always long enough to counterbalance all that has come before. The hypothetical frequentist has to say that the probability of Heads does not exist. But I told you that it is a fair coin, so the probability is ½, and I do not appear to have contradicted myself.

It is no solution to rule this sequence out as impossible. For if this sequence is impossible, then surely all sequences are impossible: all sequences are equal co-habitants of the infinite product space, and they all have exactly the same probability. So this sequence is no more impossible than your favorite 'well-mixed' sequence. In short, there is no guarantee that the limit of the relative frequency of Heads is defined.

6. *The Limit may Equal the Wrong Value*

There are still further sequences—indeed, uncountably many of them, exactly as many as the 'well-behaved' sequences—in which the limit exists, but does not equal what it 'should' (namely 1/2). The fair coin could land Heads forever: HHHHH… Now, ironically, no reordering of the results (temporal, east-west, up-down, or anything else) will save it: on any ordering the limiting relative frequency of Heads is 1. Again, this cannot be ruled out any more than your favorite nicely mixed sequence of Heads and Tails.

The frequentist might bite the bullet and say "if that is really the sequence of results, then the probability of Heads really is 1. The coin sure looks like it deterministically yields only Heads". But then let the sequence be a single Tail, followed by Heads forever: THHHHH…. Or a finite sprinkling of Tails, followed eventually by Heads forever. Or the sequence TH THH THHH THHHH THHHHH, in which a single tail is always followed by ever-lengthening sequences of Heads. Perhaps this still looks too patterned to be indeterministic; so randomize where the T's appear. We can still see to it that the limiting relative frequency of Heads is 1.

The upshot is that the hypothetical limit of the relative frequency of Heads may not be what the frequentist wants it to be. In fact, the limit may exist, but equal any value between 0 and 1: this is one way of demonstrating that the cardinality of the set of 'badly-behaved' sequences is exactly the same as the cardinality of the set of 'well-behaved' sequences. The data could be misleading in every possible way.[9] Not only is this compatible with our hypothesis that the coin is fair—*it is implied by it!*

Interlude: Three Replies on Behalf of the Frequentist

At this stage, the frequentist that dwells deep inside you may want to fight back. In the years that have passed since writing this paper's predecessor, even I have found my inner frequentist (I told you that I may have mellowed!)—up to a point. I will now let him speak, offering three responses to the last four objections. They all imagined various ways in which the hypothetical limiting relative frequency for some attribute could be badly behaved: by there not being a fact regarding its value, by having different values on natural reorderings, by being undefined, or by being defined but having the wrong value. The three responses will each challenge the claim that these things could happen, in the relevant sense of 'could', insisting that these cases pose no real threat to frequentism.

But fear not; I have not lost my nerve. In each case, I think the response can be safely rebutted, as I will argue.

Frequentist response #1: In the nearest *possible worlds the limiting relative frequency is what it 'should' be*

> The frequentist responds:
> Granted, the errant sequences that have been imagined are possible—there are possible worlds in which they occur. But as Stalnaker and Lewis have taught us, what matters to the truth conditions of counterfactuals are the *nearest* possible worlds in which the antecedents are realized. And in each imagined case, the nearest possible worlds would display the limiting relative frequency that they 'should'. For example, in all the nearest worlds in which we toss a fair coin infinitely often, its limiting relative frequency of Heads is ½. To be sure, there are other worlds in which the limiting relative frequency is something else, or is undefined; but those worlds are *less similar* to the actual world. After all, something anomalous, probabilistically speaking, is going on in those worlds, and they are more remote for that very reason.

I reply: I began with a coin that was tossed exactly once, and that landed Heads, and I asked how it would behave on its hypothetical subsequent trials. For all that has been said so far, it may be reasonable to think that the nearest possible world in which it is tossed forever is one in which it lands Heads forever. After all, what could be more similar to the coin's actual behavior than a continuation of that behavior? Perhaps this is not so compelling after just a single trial; so consider instead the result of actually tossing the coin 1,000 times. It is extremely unlikely to

9 This recalls the problems with 'operationalism', mentioned in my previous paper's discussion of finite frequentism.

land Heads exactly 500 times; for definiteness, suppose that it does so 471 times. The hypothetical frequentist may want to claim that the limiting relative frequency of Heads would not be 0.471, but rather 1/2, if the sequence were infinitely extended: in all possible worlds in which the sequence is so extended, the limiting relative frequency is 1/2. But why think this? There is a sense in which 'correcting' what was observed in this way involves a gratuitous modification of the facts, an uncalled-for extra departure from actuality.[10] If the hypothetical frequentist insists that the coin is fair, I ask him what makes him so sure. Is it something about the coin's physical make-up, for example its near-perfect symmetry? Then he starts sounding like a propensity theorist. Is it the fact that the relative frequency in 1,000 trials was roughly ½? Then he starts sounding like a finite frequentist—except he has no business rounding off the decimal of 0.471. Perhaps he is more of a Lewisian 'best systems' theorist, insisting that rounding off the decimal provides a gain in simplicity that more than offsets the loss in fit—but again that is to favor a rival theory. He surely does not directly intuit that the nearest infinitely-tossed worlds are all limiting-relative-frequency-equals-1/2 worlds. Rather, that intuition (if he has it at all) must be based on belief that the probability of Heads is ½, grounded in some *other* fact about the coin. But then it seems that the truth-maker for his probability claim is this other fact, rather than the limiting relative frequency.

In any case, in insisting that if the coin were tossed infinitely often, the limiting relative frequency *would* be ½, the hypothetical frequentist is apparently denying that it *might not* be ½—for I take it that the 'would' and 'might not' counterfactuals are contraries. But now it is my turn to insist on something: the chanciness of the coin implies that the limiting relative frequency *might not* be ½—it might be something other than ½, or it might not exist at all. That's what probability is all about.

Frequentist response #2: The strong law of large numbers underwrites the 'would' counterfactuals

> The frequentist responds:
> The strong law of large numbers tells us that a fair coin *would* land Heads with limiting relative frequency ½. In general, it says:
>
> > For a sequence of independent trials, with probability p of a 'success' on any given trial,
> >
> > $Pr(\text{limiting relative frequency of successes} = p) = 1.$
>
> Said another way: those 'badly behaved' sequences that do not have the correct limiting relative frequency have collectively probability 0, and they are in this sense pathological. That's why we may safely ignore them.

I reply: This move won't work, for a number of reasons:

Firstly, whether or not a sequence is pathological is not immediately read off its probability. For as I have said, all sequences, including the patently well-behaved ones (whatever that exactly means) have the same probability, namely 0.

[10] I thank Daniel Nolan for suggesting a version of this point to me.

Secondly, the law of large numbers itself has several probabilistic references in it, both tacit and explicit. The notion of 'independent' trials is tacitly probabilistic: it means that the probability of conjunctions of trials equals the corresponding product of probabilities. p is a probability, the constant probability of a success from trial to trial. So is Pr, the 'meta'-probability whose value is 1. How are these probabilities to be understood? As limiting relative frequencies? Take for instance the meta-probability statement, that essentially says that a set of sequences of a particular sort has probability 1. Should we give a frequentist translation of this: the limiting relative frequency, within an infinite meta-sequence of sequences, of sequences of that sort, is 1? But that is not quite right, for even the frequentist should admit that this limit is not *certain* to be 1—that statement in turn is true only 'with probability 1'. So we have a meta-meta-probability, whose value is 1. Should we give a frequentist translation of this? Not unless we are fond of infinite regresses.

Thirdly, an appeal to the law of large numbers from a frequentist should strike one as quite odd. Jeffrey notes that according to the frequentist, the law of large numbers is a tautology. To be sure, tautologies are true, and we often appeal to the 'tautologies' of mathematics and logic in our theorizing. So I would put the point a little differently: according to the frequentist, the law of large numbers admits of a one-line proof! Little wonder that the limiting relative frequency equals the true probability, with probability one. According to the frequentist, it is an analytic truth that the limiting relative frequency and the true probability are one and the same!

In fact, the frequentist should regard the law of large numbers as strangely coy. The law is stated cautiously: the convergence to the true probability happens *only* with probability one, while the frequentist thinks that it happens with certainty. And, as probability textbooks are fond of reminding us, 'probability one' is weaker than 'certainty'.

The frequentist might put up some resistance at this point: "Granted, the required convergence must be qualified: the limiting relative frequency equals the true probability *with probability one*. And that means in turn that, in an infinite meta-sequence of sequences of such trials, the limiting relative frequency of 'good' sequences is 1—again, with the qualification *with probability one*. And so on. But each step in this regress adds further information. The regress is not vicious, but informative." In response to this, it might be helpful to reflect on why the hypothetical frequentist believes in probability in the first place. Presumably it is because he can't predict with certainty the outcomes of certain experiments (and maybe he thinks that these outcomes are unpredictable in principle). He wishes that at least he could predict with certainty their relative frequency in a finite sequence of observations; but he still can't do that. So he tries to comfort himself with the thought that at least he could make the prediction with certainty in a hypothetical infinite sequence; but even that is not quite right. So he retreats to the true statement that the relative frequency does what it is supposed to in this hypothetical sequence 'with probability one'. (He just can't get rid of those wretched modalities, can he?!) The strong things that he wants to say are not true, and the true things that he ends up saying are not strong enough for his purposes.

Frequentist response #3: Probabilities are only defined in collectives

The frequentist responds:
HF is on the right track, but following von Mises (1957) and Church (1940), an extra detail is required. Probabilities are hypothetical relative frequencies in *collectives.* A collective is an infinite sequence in which limiting relative frequencies are well-behaved in a particular way. More formally, a collective is an infinite sequence $\varpi = (\varpi 1, \varpi 2, \ldots,)$ of attributes (thought of as all the outcomes of a repeatable experiment), which obeys two axioms:

Axiom of Convergence: the limiting relative frequency of any attribute exists.
Axiom of Randomness: the limiting relative frequency of each attribute is the same in any recursively specifiable infinite subsequence of ϖ.

The latter axiom is meant to capture the idea that the individual results in the collective are in a strong sense unpredictable: there is no gambling system for betting on or against the various possibilities that would improve the gambler's odds. For example, the sequence:

HTHTHTHTHTHT …

is not a collective, since it violates the axiom of randomness. The limiting relative frequency of 'H' in the infinite subsequence of odd-numbered trials is 1, while in the infinite subsequence of even-numbered trials it is 0. This corresponds to a predictability in the sequence that a gambler could exploit: she could guarantee wins by betting on 'H' on exactly the odd-numbered trials.

Now we can answer some of the arguments to hypothetical frequentism of the previous sections. In the train example (argument 4), the spatial ordering of outcomes was exactly this sequence, hence it was not collective; nor was the temporal ordering of outcomes, for that matter. And a sequence in which the limiting relative frequency of an attribute does not exist (argument 5) is not a collective, since it violates the first axiom. So hypothetical frequentist probabilities should not be defined in these cases.

I reply: The axiom of randomness both excludes and includes too much. It excludes too much: one may speak of a *process* as being random, irrespective of its outcomes. Repeatedly tossing a fair coin is a paradigmatically random process, but its outcomes may not form a collective. (Make the example quantum mechanical if you prefer.) And it includes too much: it judges as random the sequence HHHHHH… After all, the limiting relative frequency of H is the same in *every* subsequence (namely, 1), so *a fortiori* it is the same in every infinite recursively specifiable subsequence. But if you ask me, that is as *non*-random as a sequence can be.[11] And more to the point, I appealed to exactly this sequence in argument 6, so it was certainly fair game to do so. Moreover, I could change the train example so that

[11] Now perhaps you may reply that it *is* random: it is the degenerate case of a random sequence of outcomes of a coin that is guaranteed to land Heads (e.g. it is two-headed). Then let the sequence be a single tail, followed by Heads forever: THHHHH…. Or the sequence TH THH THHH THHHH THHHHH, in which a single tail is always followed by ever lengthening sequences of Heads. These sequences are also *non*-random, yet they are still collectives.

the spatial and temporal orderings of the outcomes are *both collectives*, but with different limiting relative frequencies for Heads. Admittedly, I would not be able to convey to you exactly what the orderings look like—to do so would require my giving a recursive specification of the Heads outcomes, which the axiom of randomness rules out. But nor was von Mises able to give us an example of a collective (for the same reason)—apart from examples that are surely *counter*-examples, degenerate cases such as HHHH…

According to a von Mises-style hypothetical frequentist, we know that for any coin that has a well-defined probability of landing 'Heads', the following counterfactual is true:

> if the coin were tossed infinitely often, it would either yield a sequence in which the placement of Heads outcomes is not recursively specifiable, or else a degenerate sequence in which all infinite subsequences have the same limiting relative frequency (0 or 1).

I claim that we don't know that. If the coin were tossed infinitely often, it *might not* yield either kind of sequence—the 'might' here being epistemic, signifying what is consistent with all that we know. We are in no position to rule out other things that such a coin could do—see my previous arguments for some examples!

This ends the interlude; back to main arguments.

7. *Hypothetical Frequentism's Order of Explanation is Back-to-Front*

Von Mises introduced collectives because he believed that the regularities in the behavior of certain actual sequences of outcomes are best explained by the hypothesis that those sequences are initial segments of collectives. This seems curious to me: we *know* for any actual sequence of outcomes that they are *not* initial segments of collectives, since we know that they are not initial segments of infinite sequences—period. In fact, often we know rather more than that—e.g., I can know that this coin will only ever be tossed once (I can destroy it to make sure of that). But even if, completely implausibly, we believed that an actual sequence *was* an initial segment of a collective, how would that explain, let alone *best* explain, the regularity in this initial segment? It is not as if facts about the collective impose some constraint on the behavior of the actual sequence. Something that *would* impose such a constraint—probabilistic, to be sure—is a single case probability that is fixed from trial to trial. For example, we explain the fact that our coin landed Heads roughly half the time in 1,000 tosses with the hypothesis that its single case probability of Heads is roughly ½, constant across trials, and that the trials are independent of each other. We then appeal to the usual Binomial model of the experiment, and show that such a result is highly probable, given that model. But this is not the hypothetical frequentist's explanation. Von Mises famously regarded single case probabilities as "nonsense" (e.g. 1957, p. 17). I leave it to you to judge whether that epithet is better reserved for the explanation of the coin's behavior that adverts to the wild fiction of its results belonging to a collective.

Generalizing to any hypothetical frequentist account: I maintain that its order of explanation is back to front. The fact that we have IID trials of a random experiment explains their long run frequency behavior, rather than the other way round. Compare the case of repeated sampling from a fixed population with mean μ, and the sample mean $\bar{X}$ for large samples. We do not account for μ being what it is on the basis of $\bar{X}$ being what it is; rather, the order or explanation is the other way around.

8. *The Limit Might Exist When it should Not*

In Sect. 5 I considered cases in which a given probability exists, but the limiting relative frequency does not. Now consider reverse cases: a given probability does *not* exist, but the limiting relative frequency does (as it so happens).

One way of making this point involves an appeal to a result from measure theory. It turns out that, given certain plausible assumptions, it is impossible to assign a uniform probability distribution across the [0, 1] interval, such that every subset receives a probability value. Certain subsets remain as probability gaps: so called 'non-measurable' sets. Let *N* be such a set. Imagine throwing an infinitely thin dart at random at the [0,1] interval, and consider the probability that the dart lands inside *N*. Let us agree that *N* has no probability (for to say otherwise would be to attribute a measure to a non-measurable set). Now imagine performing this experiment infinitely often, and consider whether the limiting relative frequency of landings inside *N* exists or not. It had better not, according to the frequentist, since we have agreed that the probability does not.

What does it mean for a limiting relative frequency not to exist? The only way that this can happen is if the relative frequency sequence oscillates forever, the oscillations never damping to arbitrarily small levels. This in turn only happens when successively longer and longer runs of one result predominating are followed by successively longer and longer runs of another result predominating. (Recall the HT HHTT HHHHTTTT HHHHHHHHTTTTTTTT… sequence.) But what reason do we have to think that our sequence of dart-landings inside and outside *N* will display such persistent instability? I challenge the frequentist to make sense of probability distributions such as the usual Lebesgue measure over [0,1], with its corresponding non-measurable sets.

My strategy here is to find objective probability gaps that hypothetical frequentism wrongly fills. Perhaps other, less esoteric examples will do the job. For example, *free acts* may be the sorts of things to which objective probabilities simply don't attach. Nevertheless, in virtue of their very freedom, they may have stable relative frequencies. Suppose that I repeatedly may choose to raise my left hand or right hand. I can drive the relative frequency of each to whatever value I like—and if you are happy to entertain staggering counterfactuals about my making these choices infinitely often, then you should be happy to allow the innocuous further assumption that I can freely steer the limiting relative frequency anywhere I want. And yet my free choice may lack an objective probability. The frequentist sees objective probability values, when there may be none. If you don't like this example, but you like some other example of a probability gap, then I am sure that I can run my argument using it instead.

Another way of making my point is to consider gerrymandered sequences of trials of highly heterogeneous events. Let the first trial determine whether the next person to cross the New South Wales border has an even number of hairs. Let the second trial determine whether Betelgeuse goes supernova or not in the next million years. Let the third trial determine whether my dog Tilly catches the next ball I throw to her or not … Consider the highly disjunctive 'event type' *D* that occurs when any of these events occur, and imagine keeping a tally of *D*'s relative frequency as each trial is completed. For all we know, *D* could have a well-defined limiting relative frequency; the hypothetical frequentist then regards that as *D's* objective probability. But it is not clear that *D* is the sort of event type that could have an objective probability at all. Unlike, say, repeated tosses of a coin, there is nothing *projectible* about it. If the relative frequencies happen to stabilize, that is by fluke, and not because of some stable feature of the event type itself. I take objective probabilities to be such stable features.

A brief aside. I said earlier that I believe that collectively my fifteen arguments also threaten Lewisian 'best systems' accounts of chance. I can now be more specific about that claim: I believe that the last five arguments (Sects. 4–8) can easily be rewritten to target such accounts.

9. *Subsequences can be Found Converging to Whatever Value You Like*

For each infinite sequence that gives rise to a non-trivial limiting relative frequency, there is an infinite *subsequence* converging in relative frequency to any value you like (indeed, infinitely many such subsequences). And for each subsequence that gives rise to a non-trivial limiting relative frequency, there is a sub-subsequence converging in relative frequency to any value you like (indeed, infinitely many sub-subsequences). And so on. There are, moreover, subsequences converging to no value at all (again, infinitely many). This is reminiscent of the problem of ordering discussed earlier: infinite sequences provide an embarrassment of riches. It is another way in which hypothetical frequentism faces a reference sequence problem. It is also another way of making the point that there is no such thing as 'the' infinite counterfactual extension of a given finite sequence. Far from there being a single such extension, there are infinitely many; and far from them agreeing on the limiting relative frequency, they collectively display every disagreement possible.

10. *Necessarily Single-Case Events*

A fatal problem for finite frequentism is the notorious problem of the single case, and I discussed it in my previous paper. For example, a coin that is tossed exactly once necessarily has a relative frequency of either 1 or 0 of Heads, yet the probability of Heads can surely be intermediate. Hypothetical frequentism appears to solve the problem by going hypothetical—by sending us to other possible worlds in which the coin is tossed repeatedly. However, consider an event that is *essentially* single case: it *cannot* be repeated. For instance, some cosmologists regard it as a genuinely chancy matter whether our universe is open or closed—apparently certain quantum fluctuations could, in principle, tip it one way or the other—yet whatever it is, it is 'single-case' in the strongest possible sense. Either we can make no sense of the limiting relative frequency in HF for this case, or it trivially collapses to 1 or 0.

11. *Uncountably Many Events*

The previous problem involved cases where there are *too few* events of the requisite kind; now I want to consider cases where there are *too many*. HF assumes a *denumerable* sequence of results. It runs into trouble if we have *non*-denumerably many events of the relevant sort; how are we to form a merely denumerable sequence of them? For instance, each space-time point may have a certain property or not—say, the property of having a field strength of a certain magnitude located there. What is the probability that a given point has this property? The trouble is that there are uncountably many such points.

I suppose the frequentist might imagine a denumerable sequence of 'random' selections of points (whatever that might mean) and the limiting relative frequency with which the points have the property. But my question was not about such a sequence, which forms only a tiny subset of the set of all points. It's as if the frequentist wants to pay heed to the evidence—the pattern of instantiation of the property—but only a *little* heed. After all, he identifies the probability of the property's instantiation at a point with its relative frequency of instantiation in a very small set (comparatively speaking) of representative points. And which denumerable subset is to form this set? As in the problem of reordering (argument 4) and of subsequences (argument 9), the frequentist faces an embarrassment of riches. Any limiting relative frequency will be instantiated by *some* denumerable sequence, trivial cases aside.

The problem of necessarily single-case events, and of uncountably many events, are two ends of a spectrum. HF speaks only to the 'middle' cases in which denumerable sequences of trials of the relevant kind are both possible and exhaustive. But probabilities should not be held hostage to these seemingly extraneous facts about event cardinality.

12. *Exchangeability, and Independent, Identically Distributed Trials*

Consider a man repeatedly throwing darts at a dartboard, who can either hit or miss the bull's eye. As he practices, he gets better; his probability of a hit increases:

P(hit on $(n + 1)$th trial) > P(hit on nth trial).

Hence, the trials are not *identically distributed*. Still less are they *exchangeable* (meaning that the probability of any sequence of outcomes is preserved under permutation of finitely many of the trials). And he remembers his successes and is fairly good at repeating them immediately afterwards:

P(hit on $(n + 1)$th trial | hit on nth trial) > P(hit on $(n + 1)$th trial)

Hence, the trials are not *independent*.

For all these reasons, the joint probability distribution over the outcomes of his throws is poorly modeled by relative frequencies—and the model doesn't get any better if we imagine his sequence of throws continuing infinitely. More generally, in attributing the probabilities to the outcomes that are blind to trial number, the frequentist regards the trials as being identically distributed and indeed exchangeable. And the probability he assigns to the outcome of a given trial is insensitive to the outcomes of any finite number of *other* trials. To be sure, this probability is sensitive to the outcomes of *all* the other trials. All too sensitive—the individual trial completely loses its autonomy, its probability entirely determined by what *other* trials do.

The hypothetical frequentist might reply that we should 'freeze' the dart-thrower's skill level before a given throw, and imagine an infinite sequence of hypothetical tosses performed at exactly that skill level.[12] For example, the probability that he hits the bull's eye on the 17th throw is putatively the limiting relative frequency of hits in an infinite sequence, on every trial of which he has exactly the ability that he actually has on the 17th throw. But this really is to give up on the idea that relative frequencies in the actual world have anything to do with probabilities. Indeed, this seems like a convoluted way of being a propensity theorist: all the work is being done by the thrower's 'ability', a dispositional property of his, and the hypothetical limiting relative frequency appears to be a metaphysically profligate add-on, an idle wheel.

Our dart-thrower is hardly far-fetched (unlike the counterfactuals that our frequentist will deploy in describing him). He can't be easily dismissed as a 'don't care' for an account of probability. On the contrary, I would have thought that systems that have memories, that self-regulate, and that respond to their environments in ways that thwart IID probabilistic models, are the norm rather than the exception.

13. *Limiting Relative Frequency Violates Countable Additivity*

Kolmogorov's axiom of countable additivity says: given a countable sequence of disjoint propositions, the probability of their union equals the sum of their individual probabilities. Hypothetical frequentism violates this axiom: there are cases in which the limiting relative frequency of the union does not equal the sum of the limiting relative frequencies. For anyone who holds sacred Kolmogorov's full set of axioms of probability, this is a serious blow against frequentism: it means that frequentism is not an interpretation of the entire Kolmogorov probability calculus.

To see this, start with a countably infinite event space—for definiteness, consider an infinite lottery, with tickets 1, 2, 3, … Let A_i = 'ticket i is drawn'. Suppose that we have a denumerable sequence of draws (with replacement), and as it happens, each ticket is drawn exactly once. Then the limiting relative frequency of each ticket being drawn is 0; and so according to the hypothetical frequentist, $P(A_i) = 0$ for all i, and so

$$\sum_{n=1}^{\infty} P(A_n) = 0.$$

But $A_1 \cup A_2 \cup A_3 \cup \ldots$ is an event that happens every time, so its limiting relative frequency is 1. According to the hypothetical frequentist, this means that

[12] Thanks to Kenny Easwaran and Aidan Lyon for suggesting versions of this reply.

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = 1,$$

a violation of countable additivity.

The hypothetical frequentist cannot help himself to various important limit theorems of probability theory that require countable additivity for their proof. The failure of countable additivity also gives rises to failures of conglomerability (Seidenfeld et al. 1998), which is perhaps more troubling for credences than for objective probabilities, but still troubling enough, especially since credences are supposed to coordinate with objective probabilities à la Lewis' Principal Principle (1980). And at least *some* objective probabilities seem to be countably additive, since they are parasitic on lengths, areas, or volumes of regions of space, which are themselves countable additive. As van Fraassen (1980) points out, for events of the form 'the outcome is in *Q*', where *Q* is a measurable region of n-dimensional Euclidean space, we may define a probability that is proportional to the 'volume', or Lebesgue measure of *Q*; and Lebesgue measure is countably additive (hence, so is anything proportional to it). These probabilities, then, cannot be limiting relative frequencies.

14. *The Domain of Limiting Relative Frequencies is not a Field*

Probabilities are usually defined on a *field* on some set Ω—a collection of subsets of Ω closed under complementation and intersection. (Indeed, Kolmogorov went further, defining probabilities on a *sigma*-field, which is closed under complementation and *countable* intersection.) However, the domain of limiting relative frequencies is not a field. There are cases in which the limiting relative frequency of *A* is defined, the limiting relative frequency of *B* is defined, but the limiting relative frequency of the intersection *AB* is undefined.

De Finetti (1972) writes:

> Two people, *A* and *B*, are gambling on the same sequence of coin tossings; *A* bets always on Heads, *B* changes between heads and tails at intervals of increasing length. Things may be arranged (proof omitted in this translation) so that the [relative] frequency of successes for *A* and for *B* tends to 1/2, but the [relative] frequency of joint success *AB* oscillates between the lower limit 0 and the upper limit 1. (75).

Let's clean up de Finetti's infelicitous notation—'*A*' and '*B*' start out as people, but end up as events that can be conjoined. Let *A* be the event 'the first person succeeds' and *B* be the event 'the second person succeeds'. It's unfortunate that the proof is omitted. In fact, there is surely a mistake here: the relative frequency of joint success *AB* cannot oscillate as high as 1 in the long run, since the relative frequency of *AB* is bounded above by the relative frequency of *A*, and its limit is 1/2. But the relative frequency of *AB can* oscillate between roughly 0 and ½ in the long run, and that suffices to show that the domain of limiting relative frequencies is not a field.

For example, suppose that the first person bets on H on each toss. The second person bets on oscillating runs of length 10^n, as follows: H on the first trial; T for the next 10 trials; H for the next 10^2 trials, T for the next 10^3 trials, etc. In fact, the coin lands HTHTHTHT …. Then we have:

relative frequency(A) $\rightarrow$ 1/2.
relative frequency(B) $\rightarrow$ 1/2.

But
relative frequency(AB) has no limit,
since it oscillates between roughly 0 and 1/2 forever. For joint successes happen *none* of the time throughout a particular run of the second person betting on T, and eventually these amount to nearly all of the trials; but joint successes happen *half* the time during a particular run of her betting on H, and eventually *these* amount to nearly all of the trials.

I will conclude with one further argument based on the mathematics of limiting relative frequencies. In the prequel to this paper, I noted that according to finite frequentism, there are no irrational probabilities. To be sure, the move to hypothetical frequentism solves this problem. But an analogue of this problem holds even for hypothetical frequentism.

15. *According to Hypothetical Frequentism, there are no Infinitesimal Probabilities*

A positive infinitesimal is a number greater than zero, but smaller than every positive real number. The hypothetical frequentist's probabilities are always either rational numbers, or the limits of sequences of rational numbers. In either case, they are real-valued. This means that they cannot be (positive) infinitesimal-valued. Yet infinitesimal probabilities have appealed to, and have been appealed to by, various philosophers—e.g. Lewis (1980), Skyrms (1980), McGee (1994), and Elga (2004). If a probability function over an uncountable space is to be 'regular'—it assigns 0 only to the empty set—it must make infinitesimal probability assignments. (See Hájek 2003 for a proof.) Moreover, infinitesimal probability assignments seem to be well motivated by various random experiments with uncountably many possible outcomes—e.g., throwing a dart at random at the [0, 1] interval. And infinitesimal probabilities can apparently help to solve a variety of problems—for example, the 'shooting room problem' (Bartha and Hitchcock 1999). To be sure, in some of these applications the probabilities in question may be regarded as subjective. But it is somewhat troubling that they *cannot* be objective, if objective probabilities are understood along frequentist lines.

A Parting Offering: Hyper-Hypothetical Frequentism

The last objection prompts me to end with a tentative positive proposal as to how hypothetical frequentism might be refined, so as to evade this objection and several others besides. The objections I have in mind turned on certain mathematical facts about the limits of infinite sequences of rational numbers (arguments 4, 5, 13, 14, and 15). The proposal, then, is to appeal to something that preserves the idea of having such infinite sequences, but for which the mathematics 'looks' finite: namely the theory of *hyperreal numbers*. I can only sketch the idea here—but I hope that someone familiar with the theory will see how the details might be filled in further.

Non-standard models give us a way handling infinite and infinitesimal quantities in a way that makes them behave just like finite quantities: these are so-called *hyperfinite* models. One construction that I find particularly user-friendly treats hyperreal numbers as equivalence classes of sequences of reals. *What is the equivalence relation?* Two sequences are regarded as equivalent iff they agree term by term almost everywhere. *What does 'almost everywhere' mean?* We must impose an additive measure on the positive integers **N**, which takes just the values 0 and 1, and which assigns **N** measure 1, and all finite sets measure 0. But there are infinitely many such measures—*which one are we talking about?* That just means that there are as many different ways of defining the hyperreals as there are ways of making this arbitrary choice of the measure. So imagine this choice being made. Among the hyperreals so chosen are infinite numbers, such as this one: the set of all sequences equivalent to $\{1, 2, 3,\ldots\}$. Call that number K. Among them are also infinitesimals, such as this one: the set of all sequences equivalent to $\{1, 1/2, 1/3, \ldots\}$. This will turn out to be $1/K$. For further details, see Lindstrom (1989).

We want to have most of our cake, and to eat most of it too: we want a genuinely infinite sequence of trials, yet we want to avoid some of the problems, encountered previously, that were artifacts of our use of infinite sequences. The trick, I suggest, is to suppose that we have a hyperfinite number of appropriate trials; and now define probability as simply relative frequency (note: *not* limiting relative frequency) among these. More precisely, let us introduce *hyper-hypothetical frequentism* as follows[13]:

(HHF) The probability of an attribute A in a reference class B is p

iff

the relative frequency of occurrences of A within B would be p if B were hyperfinite.

We not only recover irrational probabilities this way—we even recover infinitesimal probabilities. For example, suppose that our B consists of K trials, and consider an event that happens exactly once. Its relative frequency is $1/K$, which is infinitesimal. And we can distinguish it in probability from another event that happens twice, hence whose relative frequency is $2/K$. And so on for any event that happens a finite number of times. But standard hypothetical frequentism conflates the probabilities of events that happen finitely many times: they all are regarded as having probability 0.

The mathematics of hyper-hypothetical frequentism in other respects looks finite, and this gives it several further advantages over hypothetical frequentism. Now, one cannot change the relative frequency by reordering the trials (just as one cannot in finite sequences). Probabilities are now *guaranteed* to exist (and not just 'almost

[13] I thank Aidan Lyon for suggesting this name.

certain' to), just as they were for finite actual frequentism: after all, we are simply counting and taking ratios, not taking limits as well. The taking of subsequences is no more allowed than it was for finite actual frequentism: I said that we are to take the relative frequency among the K trials, and not some smaller number of trials. So probabilities will always be uniquely defined. They are hyperfinitely additive—that is, given a sequence of j disjoint propositions, where j is a non-standard integer which may even be infinite, their probabilities add in a finite-looking way. And their domain of definition is a field, as it should be; indeed, probability assignments are closed under hyperfinite unions and intersections.

That's the good news; now I should be forthright about the bad news. It begins with what I breezily called the "arbitrary choice" of measure when constructing our hyperreal model: demons lurk behind the words "arbitrary" and "choice". It is often a sign of a weakness in a philosophical position when it requires an arbitrary choice, particularly if that choice seems not to correspond to any natural property or distinction that we might care about. I imagined *choosing a measure* over $\mathbf{N}$; but what could favor one choice over another one? (Arbitrariness entered again when I imagined a hyperfinite set of K trials; why K rather than some other hyperfinite number?) Moreover, I *imagined* choosing a measure over $\mathbf{N}$; but in fact we cannot fully specify this measure, even in principle. Cognoscenti will recognize this problem as the indefinability of the ultrafilter that corresponds to the 'measure 1' sets in the construction that I sketched above—see my 2003 for further discussion. Hyperreal numbers are in this sense *ineffable*. To be sure, this seems to be a general cause for concern about the use of hyperreal probabilities, the enthusiasm for infinitesimal probabilities that I reported notwithstanding. My final argument above, then, is weakened to that extent. That said, at least the problem of arbitrariness for my proposal has a counterpart for hypothetical frequentism that is worth highlighting now. After all, that account privileged *denumerable* reference classes among all the infinite cardinalities—why that order type rather than another? Exposing the arbitrariness inherent in this would give us a close variant of argument 11.

And what about all my other objections to hypothetical frequentism? Well may you ask. They remain, as far as I am concerned. My proposal is still a complete abandonment of empiricism; it has us imagine utterly bizarre counterfactuals (even more so than HF does); there is no fact of what the counterfactual hyper-hypothetical sequences will look like; and so on. I offer my proposal as an improvement on hypothetical frequentism, not as a rescue of it. I believe it is beyond rescuing.

Conclusion

Finite frequentism confuses good methodology concerning probability for probability itself: while it respects the practice of experimental science, it disrespects too many of our central intuitions about probability. Hypothetical frequentism strives to have the best of both worlds: the scientific spirit of finite frequentism, yet with an injection of the modality that our commonsensical notion of probability seems to require. However, in the final analysis, it runs afoul of both science and commonsense. Its abandonment of finite frequentism's empiricism is costly, and it fails to pay its way with corresponding benefits; moreover, it suffers from other technical difficulties of its own. I have suggested a way to refine hypothetical frequentism so as to avoid some of these difficulties, with hyper-hypothetical frequentism—but in the end I think it amounts only to a little cosmetic surgery.[14]

References

Bartha, P., & Hitchcock, C. (1999). The shooting-room paradox and conditionalizing on 'measurably challenged' sets. *Synthese, 118*, 403–437. doi:10.1023/A:1005100407551.

Carnap, R. (1950). *Logical foundations of probability*. Chicago: University of Chicago Press.

Church, A. (1940). On the concept of a random sequence. *Bulletin of the American Mathematical Society, 46*, 130–135. doi:10.1090/S0002-9904-1940-07154-X.

Costantini, D., & Galavotti, M. (Eds.). (1997). *Probability, dynamics and causality–essays in honor of R. C. Jeffrey*, Dordrecht: Kluwer.

de Finetti, B. (1972). *Probability, induction and statistics*. London: John Wiley & Sons.

Elga, A. (2004). Infinitesimal chances and the laws of nature. *Australasian Journal of Philosophy*, *82*, (March), 67–76. (Reprinted from *Lewisian themes: The philosophy of David K. Lewis,* by F. Jackson, G. Priest (Eds.), 2004, Oxford: Oxford University Press.)

Hájek, A. (1997). 'Mises Redux'—Redux: Fifteen arguments against finite frequentism, *Erkenntnis 45*, 209–227. (Reprinted from Costantini and Galavotti (1997). Reprint forthcoming in *Philosophy of probability: Contemporary readings,* by A. Eagle (Ed.), 2009, Routledge.)

Hájek, A. (2003). What conditional probability could not be. *Synthese, 137*(3), 273–323. doi:10.1023/B:SYNT.0000004904.91112.16.

Howson, C., & Urbach, P. (1993). *Scientific reasoning: The Bayesian approach* (2nd ed.). La Salle: Open Court.

Jeffrey, R. (1992). Mises Redux. In *Probability and the art of judgment*. Cambridge: Cambridge University Press.

Johnson, R. A. (1994). *Miller & Freund's probability & statistics for engineers* (5th ed.). New Jersey: Prentice Hall.

Kripke, S. A. (1980). *Wittgenstein on rules and private language*. Cambridge, MA: Harvard University Press.

Lewis, D. (1973). *Counterfactuals*. Oxford/Cambridge, MA: Blackwell/Harvard University Press.

Lewis, D. (1980). A subjectivist's guide to objective chance. In Richard C. Jeffrey (Ed.), *Studies in inductive logic and probability* (Vol. 2, pp. 263–293). Berkeley and Los Angeles: University of California Press. (Reprinted in Lewis (1986)).

Lewis, D. (1986). *Philosophical papers* (Vol. 2). Oxford: Oxford University Press.

Lewis, D. (1994). Humean supervenience debugged. *Mind, 103*, 473–490. doi:10.1093/mind/103.412.473.

Lindstrom, T. (1989). An invitation to nonstandard analysis. In N. J. Cutland (Ed.), *Nonstandard analysis and its applications* (pp. 1–105). Cambridge: Cambridge University Press.

McGee, V. (1994). Learning the impossible. In Ellery Eells & Brian Skyrms (Eds.), *Probability and conditionals*. Cambridge: Cambridge University Press.

Reichenbach, H. (1949). *The theory of probability*. Berkeley: University of California Press.

Seidenfeld, T., Schervish, M. J., & Kadane, J. B. (1998). Non-conglomerability for finite-valued, finitely additive probability. *Sankhya Series A, 60*(3), 476–491.

Skyrms, B. (1980). *Causal necessity*. New Haven: Yale University Press.

[14] There are many people to thank for discussions of this material over its many years of gestation. My younger self thanks especially Fiona Cowie, Peter Gerdes, Ned Hall, David Hilbert, Carl Hoefer, Richard Jeffrey, Marc Lange, David Lewis, Daniel Nolan, Brian Skyrms, Bas van Fraassen, Jim Woodward, and Lyle Zynda. My current self thanks especially Paul Bartha, Darren Bradley, Kenny Easwaran, Branden Fitelson, Aidan Lyon, Ralph Miles, and Daniel Nolan.

Stalnaker, R. (1968). *A Theory of Conditionals*. Studies in logical theory, *American philosophical quarterly monograph series* (Vol. 2). Oxford: Blackwell.
van Fraassen, B. (1980). *The scientific image*. Oxford: Clarendon Press.
von Mises, R. (1957). *Probability, statistics and truth*, revised English edition. New York: Macmillan.