

“MISES REDUX” – REDUX: FIFTEEN ARGUMENTS AGAINST FINITE FREQUENTISM

ABSTRACT. According to finite frequentism, the probability of an attribute A in a finite reference class B is the relative frequency of actual occurrences of A within B . I present fifteen arguments against this position.

1. INTRODUCTION¹

The most widely accepted interpretation of probability is *frequentism*. Roughly, frequentism says: the probability that a coin lands heads when tossed is the number of times that the coin lands heads, divided by the total number of times that the coin is tossed; the probability that a radium atom decays in 1500 years is the number of radium atoms that so decay, divided by the total number of radium atoms; and so on. This should sound familiar – all too familiar – for somehow this notion still pervades much scientific thinking about probability. But it should be rejected, as I will argue here – fifteen times over.

To philosophers or philosophically inclined scientists, the demise of frequentism is familiar, I admit, even though it hasn't quite been universally accepted.² Familiar too are many of the arguments that I will present here – indeed, some of them were inspired by Richard Jeffrey's “Mises Redux” (1977) – though I hope it will be useful to have them gathered in one place. Other arguments in this paper are new, as far as I am aware. So even if the fact that there is bad news for frequentism is old news, I hope it is newsworthy just *how much* bad news there really is.

The stance that one takes on issues of philosophical methodology is important here. I will begin by saying some friendly things about the role of intuition in the philosophical analysis of objective probability, by way of preparation for the unfriendly things that I will say about frequentism as such an analysis. I will distinguish two frequentist analyses – *finite frequentism* and *hypothetical frequentism*. Although space limitations require me to confine my discussion to the former, many of the arguments that I will adduce will count equally against both.

2. OBJECTIVE PROBABILITY: INTUITIONS AND ANALYSIS

Probability is, I claim, a concept of both commonsense and science. The person on the street recognizes and understands it (at least to some extent) in locutions such as “the probability that this coin lands heads is 1/2”; the scientist recognizes and understands it (at least to some extent) in locutions such as “this electron is measured to be spin ‘up’ with probability 1/2”. Commonsense and science are joined by a two-way street: scientific theories are, after all, invented by people who share the folk’s conceptual apparatus, and who seek to refine it; and commonsense, in turn, partly incorporates some of these refinements, as scientific ideas become popularized. The concept of gravity, for example, was once a part of neither commonsense nor science, and now it is part of both. I believe the same is true of the concept of probability.

Many computer scientists, statisticians, physicists, economists . . . seem to speak as if probability simply *is* relative frequency – no ifs or buts, end of story.³ This is surely mistaken. We would do better to think of it as a putative *analysis* of our pretheoretical notion of probability, one which both informs and is informed by a more sophisticated scientific notion of probability. ‘Probability’, after all, is not just a technical term that one is free to define as one pleases. Rather, it is a concept whose analysis is answerable to our intuitions, a concept that has various associated platitudes (for example: “if X has probability greater than 0, then X can happen”). Thus, it is unlike terms like ‘complete metric space’ or ‘Granger causation’ or ‘material conditional’, for which there are stipulative definitions with which there is no sensible arguing, and no associated platitudes. What ‘probability’ is like, instead, are concepts like ‘space’ or ‘causation’ or ‘if . . . then’, concepts that can be the subject matter of analyses. It is fair game to dispute such analyses; and it is certainly fair game to dispute frequentism. I say this early on, to forestall any possible puzzlement about my project here (a puzzlement that I have already encountered from various computer scientists, statisticians . . .).

Furthermore, frequentism is at best an analysis of *objective* probability, sometimes called objective *chance* (to be distinguished, for example, from subjective probability, or degree of belief.) As we will see, however, it cannot be even that.

This is not to deny that probability and relative frequency have *some* sort of close connection. Subjectivists, propensity theorists, logical probabilists, and so on presumably all agree, for example, that an event with probability half should be expected to occur roughly half of the time, in some senses of the words ‘should’, ‘expected’, and ‘roughly’ (I would say this is another platitude). Moreover, I concede that finding out a relative frequency can

often be the best – and sometimes even the only – way of finding out the value of a probability. I do not deny the existence of some interesting relationship between the two; I am only disputing their identification.

So far I have taken as a starting point our commonsensical notion of probability, and I have regarded frequentism as an analysis of that (a bad one). But we could come to frequentism from another direction. Frequentism is, as I said at the outset, an *interpretation* of probability. More precisely, it is a putative interpretation of the axioms of probability theory – traditionally, those provided by Kolmogorov. So starting with a primitive, uninterpreted function P , defined over a certain set-theoretic substructure, which is non-negative, normalized, and additive, we might come to a frequentist understanding of P . Here again, intuitions have a role to play. For *many* quantities that have nothing to do with our intuitive notion of probability conform to Kolmogorov's axioms, and so in some sense provide an interpretation of them – think of mass, or length, or volume, which are clearly non-negative and additive, and which can be suitably normalized. They are not even in the running, however, because commonsense tells us that probability is simply something *else*. (Just try substituting any of them into the platitudes above!) Incidentally, this also shows that it is too glib to say that a satisfactory understanding of probability is provided as long as we find a concept of importance to science that conforms to the axioms – for that does not narrow down the field enough. In any case, the more strictly *philosophical* project of analysing our commonsensical concept would remain, much as the project of analysing our commonsensical concept of causation, say, would remain even if we had already done the job for the concept as it appears in science.

3. VERSIONS OF FREQUENTISM

It is necessary to distinguish two variants of *actual* frequentism, and these from *hypothetical* frequentism. According to actual frequentism, the probability of an event or attribute is to be identified with its actual relative frequency: there is no need to 'leave the actual world', for all the requisite facts are right here. Now, if there happen to be infinitely many events or attributes of the requisite sort, then we cannot simply count the number of 'successes' and divide this by the total number of trials, since this will take the indeterminate form ∞/∞ . In that case, we take the limit of the relative frequency up to the n th trial, as n tends to infinity. Hypothetical frequentism keeps the intuition that probability is such limiting relative frequency, but applies when the actual world does not furnish the infinitely many trials required. It thus identifies probability with a *counterfactual*

limiting relative frequency: the limiting relative frequency *if* there were infinitely many trials.

I cannot discuss the problems with hypothetical frequentism here; and since the infinite variant of actual frequentism suffers from many of the same problems, my discussion of it is best left for another occasion also.⁴ So let me focus solely, then, on the version of actual frequentism in which there are only finitely many trials of the relevant sort – for short, *finite frequentism*:

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

Venn (1876), in his discussion of the proportion of births of males and females, concludes: “probability *is* nothing but that proportion” (p. 84, his emphasis). This I take to be finite frequentism at its purest. Reichenbach had such inclinations also, although his account in the end looked more like hypothetical frequentism; finite frequentist accounts were pursued in more detail by Cramér, Hempel and Putnam among others.

Of course, all of this was a long time ago. It might thus be thought that frequentism is at best of historical interest as far as the philosophical literature is concerned – its currency elsewhere I have already emphasized – unworthy of much scrutiny in these more enlightened times. Up to a point, this is true enough, I guess; but only up to a point. As I indicated earlier, frequentism still has its proponents among philosophically inclined statisticians, and even philosophers. Furthermore, some of the criticisms presented here have some force against more sophisticated accounts that have grown out of frequentist soil, for example Lewis’ (1994) ‘best system’ approach to objective chance, as I will argue at the appropriate point. Indeed, one wonders if an empiricist account of objective chance could be given that didn’t look a lot like frequentism, and I suspect that many of these arguments could be adapted accordingly against any such account.

Any aspiring frequentist with serious empiricist scruples should not give up on finite frequentism lightly. The move to hypothetical frequentism, say, comes at a considerable metaphysical price, one that an empiricist should be unwilling to pay. Finite frequentism is really the only version that upholds the anti-metaphysical, scientific inclinations that might make frequentists of us in the first place. In any case, at first blush, it is an attractive theory. It is a reductive analysis, whose primitives are well understood; it apparently makes the epistemology of probability straightforward; unlike the classical

and logical theories of probability, it appears to be about the world; and it seems to be inspired by actual scientific practice.

However, at second blush, it does not look nearly so good: it runs afoul of many important intuitions that we have about probability. Or so I will argue.

4. THE ARGUMENTS

We are almost ready for the arguments. Why so many of them? It might make you suspicious that I am uneasy about them, substituting quantity for quality. It recalls Flew's 'leaky buckets' metaphor for philosophical arguments: to paraphrase him, one watertight one is better than fifteen leaky ones. Suffice to say that I think the arguments here are pretty watertight, and some of them are pretty decisive on their own.

One reason for giving so many arguments is this. You might agree with me that frequentism cannot be an analysis of our concept of (objective) probability; however, you might think that it is, so to speak, a successful *partial* analysis of the concept, one that captures an important and central strand in our thinking about probability (even if there are other such strands). In my ecumenical moments, even I feel some temptation to concede this: perhaps we should let a thousand flowers bloom, with frequentism being one of them. Or you might think that frequentism is a good *explication* of the concept of probability – a cleaned-up surrogate for a messy, ambiguous, vague, and even confused concept, one suitable for use in science and clear-headed discourse. But I do think that the many arguments here successively chip away at frequentism's possible domain, steadily reducing its interest.

Another reason for giving so many arguments is that it shows just how dim are the prospects for retrenching frequentism in favor of some close relative of it. A single class of counterexamples to it might prompt one to add a single epicycle in order to save it. What I hope to make clear, by piling on ever more arguments against frequentism, is that the problems are not an artifact of some particular presentation of it, ones that would go away with a little clever cosmetic surgery. No – the problems with frequentism run deep.

Despite the title and fanfare, I don't want to be too fussy about how the arguments are counted. Not all of the arguments are completely independent of each other; in fact, several of them might be regarded as stemming from a single intuition (that probability statements can obey a certain sort of 'counterfactual independence'). On the other hand, elsewhere I might

combine under the one heading two or more arguments that could be separated. I will distinguish the arguments in a way that I hope is natural.

I will begin with some general arguments that I think are telling against any form of frequentism; then, arguments specifically against finite frequentism.

GENERAL PROBLEMS CONCERNING ANY VERSION OF FREQUENTISM

1. The Reference Class Problem

We think that various events straightforwardly have unconditional probabilities, and indeed we even have theories that tell us what some of these probabilities are. But it seems that frequentism delivers only conditional probabilities — or in any case, relativized probabilities.⁵ Von Mises (1957) writes: “It is only the notion of *probability in a given collective* which is unambiguous” (p. 20). Suppose I am interested in my probability of dying by age 60. What I want is an *unconditional* probability. I can be placed in various reference classes: the set of all living things; the set of all humans; the set of all males; the set of all non-smoking males who exercise occasionally; the set of all philosophers; the set of all Woody Allen fans ... Each of these reference classes will have its own associated relative frequency for death by age 60. But I’m not interested in my probability of death *qua* philosopher, say. To repeat, I want an unconditional probability.

Here we confront the notorious ‘reference class problem’: a given event or attribute has more than one relative frequency; and according to the frequentist, this means that the event or attribute has more than one probability.

There is some irony here. Frequentists have been quick to mock Carnapian logical probability on the grounds that it must always be relativized to a choice of language, and no single language seems to be the canonical one. But a parallel problem is practically alluded to in the very name ‘relative frequency’ — frequencies must always be relativized to a choice of reference class, and no single reference class seems to be the canonical one.

I see only one possible way out for the frequentist. He should insist that *all* probability is really conditional; and that a putative unconditional probability statement is really elliptical for a conditional probability statement in which the condition is tacit. He could maintain that probability theory could still do a lot of work for us. For example, knowledge of inequalities between conditional probabilities might be all that we need in order to control our environment in desirable ways, modifying our behavior beneficially. (When you see that the conditional probability of death by age 60,

given smoking, is substantially greater than it is given non-smoking, you see a good reason to quit smoking – at least when you have ruled out other explanations for this correlation, such as the existence of a common cause, also on the sole basis of conditional probability information.) In short, rather than seeing the reference class problem as a problem, the frequentist could embrace it.

Perhaps this gives the frequentist a way out of the reference class problem. But he should admit that this ‘eliminativism’ regarding unconditional probability is somewhat radical, if only because science seems to abound with statements of unconditional probability. And of course he can no longer pretend to be giving an interpretation of Kolmogorov’s axioms.

2. Typing Events may Change the Probability

Every event, in all its myriad detail, is unique. This is just Leibniz’s principle of the identity of indiscernibles, applied to events. So if you are going to group an event with others, you will have to allow differences between them. But there needs to be a guarantee that these differences make no difference to the probability. The thought must be that whatever the differences are, they are not relevant. The bulge in the carpet then moves over to the notion of ‘relevance’. It had better not be probabilistic relevance, on pain of circularity. But what is it, then?

Plausibility drains from finite frequentism especially when the putative reference class is too heterogeneous, or too small. Unfortunately for the frequentist, these problems work in tandem, so that solving one tends to exacerbate the other. Homogeneity can be enhanced by raising the admission standards into a reference class, demanding greater similarity between the individuals or events; but that reduces the number of individuals or events that can be admitted.⁶

3. Probabilities of Local Events can be Counterfactually Independent of Distant Events

Let me first continue my plea on behalf of commonsense, with a little homily on philosophical argumentation. Sometimes arguments against a philosophical position attempt to show that the position has *internal* difficulties (and some of my arguments against frequentism are of this form). On the other hand, sometimes arguments begin with commonsensical intuitions that are supposed to be dear to us, and then deploy these intuitions against a philosophical position with which they clash. For example, we might begin with the following commonsensical intuition: this fire’s burning my hand is a matter solely involving a small region of space-time containing the fire, my hand, and little else; and then deploy this intuition against

Hume's regularity account of causation. 'Intuition-based' arguments are often not as damaging as 'internal-difficulty' arguments; and when faced with an intuition-based argument, a proponent of the philosophical position in question might simply retort that the argument is question-begging, and that we should revise our intuitions. Nevertheless, depending on the strength of those intuitions, and the weight that we attach to common-sense, such arguments can still have some pull on us. Much philosophy – the greater part of it, I would say – proceeds in just this way. (Think especially of philosophy's famous thought experiments: the Chinese Room, Twin Earth, and so on.) The argument that I want to turn to now is of this form.

Here is a radium atom; its probability of decaying in 1500 years is $1/2$. If radium atoms distant from it in space and time had behaved differently, this probability would still have been $1/2$. I submit that probability statements about things can be (and perhaps typically are) counterfactually independent of what other things of that kind happen to do; and probability statements about local events can be (and perhaps typically are) counterfactually independent of distant events of the same kind. But according to relative frequentism, this is not so.

And I submit that probability statements about a thing at a time can be counterfactually independent of the behavior of that thing at other times. The chance that this coin lands heads *now* does not depend on how the coin will land in the future – as it were, the coin doesn't have to 'wait and see' what it happens to do in the future in order to have a certain chance of landing heads now. To put the point crudely, though vividly: it's almost as if the frequentist believes in something like backward causation from future results to current chances. Put more carefully: the frequentist believes that the *future* behavior of the coin places constraints on the chance that the coin lands heads *now*, and in that sense, that chance is counterfactually dependent on the future behavior. Indeed, if the coin has yet to be tossed, the future behavior fully determines that current chance, according to the finite frequentist.⁷ I think the frequentist has things backwards: surely it is the coin's probability of landing heads that gives rise to its statistics, rather than the other way round.

And so it is in general. Frequentism suffers much the same fate as Hume's theory of causation. The fact that *this* flame burned my hand does not depend on whether *other* flames happen to be contiguous with *other* hands getting burned. The intuition behind this argument is that probability, like causation, is a far more private matter than that.

Digressing briefly: I said earlier that some of my arguments have force against certain more sophisticated analyses of objective probability that

could be thought of as refinements of frequentism, notably Lewis' (1994). It runs roughly as follows. The laws of nature are those regularities that are theorems of the *best theory*: the true theory of the universe that best balances simplicity, strength, and likelihood (that is, the probability of the actual course of history, given the theory). If any of the laws are probabilistic, then the chances are whatever these laws say they are. It seems that according to Lewis, the probability that this radium atom decays in 1500 years does depend on what other, perhaps distant atoms (and perhaps not just radium atoms) happen to do – at least this will be so under the assumption that there are sufficiently many such atoms. For if many such atoms had decayed, say, much earlier than they actually did (something the best theory will admit has positive chance), then plausibly the best theory would have had different radioactive decay laws – it would have needed to in order to have reasonable likelihood. In particular, plausibly the decay law for radium would have been different, and hence so too the decay probability for this particular radium atom.

I think that the next two arguments also carry some weight against the Lewis analysis, and perhaps even the two after that (with some smallish modifications); but let us return to our discussion of (finite) frequentism.

4. An Argument from Concern

Let me pursue a variation on the ‘counterfactual independence’ theme. I am inspired here by Kripke’s (1980) famous ‘argument from concern’ against Lewis’ counterpart theory, according to which entities cannot be genuinely identified across possible worlds:

[According to Lewis] if we say ‘Humphrey might have won the election (if only he had done such-and-such)’, we are not talking about something that might have happened to *Humphrey* but to someone else, a “counterpart”. Probably, however, Humphrey could not care less whether someone *else*, no matter how much resembling him, would have been victorious in another possible world. (p. 45)

Arguments from concern have the form: “If *A* were the correct analysis of *B*, then our concerns would be such-and-such; they are not; hence *A* cannot be the correct analysis of *B*.¹⁰” To be sure, such arguments are defeasible, and what our concerns happen to be is, I suppose, a highly contingent matter. Still, much as thought experiments can be a source of philosophical insight—even though our responses to them are surely highly contingent—I believe such arguments can be also. And they may serve a

distinctive function, trading as they do not only on our beliefs, but also on our desires (fears, regrets, . . .).

It is natural to think that my probability of dying by a certain age is a property of *me* (or perhaps me plus my immediate environment). Natural, though von Mises goes out of his way to deny it: “We can say nothing about the probability of death of an individual even if we know his condition of life and health in detail. The phrase ‘probability of death’, when it refers to a single person, has no meaning at all for us” (p. 11). Also: “It is utter nonsense to say, for instance, that Mr. X, now aged forty, has the probability 0.011 of dying in the course of the next year” (pp. 17–18). But surely it is just such ‘nonsense’ that Mr. X really cares about, when he is concerned about his probability of death. Of course, he may well unselfishly care about his fellow citizens too, and he may be concerned to find out how high the death rate is among people of his type. But to the extent that his concerns are directed to himself, the other people can drop out of the picture (much as Kripke would say that Humphrey’s counterparts can drop out of the picture when it comes to Humphrey’s concern about losing the election). The statistics about others like him may give him good *evidence* as to his own chance of dying, but the *fact* that he ultimately cares about is a fact about himself – one expressed by a meaningful ‘probability of death’ statement that refers to a single person.

GENERAL PROBLEMS CONCERNING ACTUAL FREQUENTISM (BOTH FINITE AND INFINITE)

5. Actual Frequentism Commits one to a Surprisingly Rich Ontology

Here is another variant of the ‘counterfactual independence’ argument, though sufficiently different to merit separate treatment, I think. Nicolas of Autrecourt once said words to the effect that, from the existence of one object, one cannot deduce the existence of others. But according to finite frequentism, the existence of a non-trivial probability for an event *does* imply the existence of other, similar events in the actual world. The fact that I have a non-trivial probability for dying by age 60 proves that I am not alone in the world, according to the frequentist. Or consider some probability statement about my own mind: for example, that it will deteriorate by age 60 with probability 0.1. According to the finite frequentist, this means that a tenth of the people out there with minds like mine experience such deterioration by age 60 – which of course implies that there *are* other such minds. Now there’s a quick argument against solipsism for you!

I’m being a little facetious here. It’s not that implying the falsehood of solipsism is a bad thing – on the contrary. And of course the would-be

solipsist-frequentist will simply deny that there are any non-trivial probabilities about my mind. What's troubling, though, is that statements of probability about a mind, an object, or an event, seem to be simply irrelevant to the existence of *other* minds, *other* objects, *other* events of the same sort, right here in the actual world. Moreover, we can even put lower bounds on *how many* such entities there are – for example, at least 9 other minds in the case just considered, and of course some multiple of 10 minds in total. It is often true that the required things do indeed exist, and in the numbers required (at least 9 other minds, for example). But sometimes they do not – see the problem of the single case below; and in any case, these things simply don't seem to be *implied* by the corresponding probability statements.

6. Actual Frequentism = Operationalism about Probability

The finite frequentist definition of probability sounds a lot like an *operational* definition. Like the operational definitions of temperature in terms of actual thermometer measurements, or mental states in terms of actual behavior, we have probability being defined in terms of the results of some actual ‘measurement’ (put in scare quotes, since the results might not always be observed): in this case, the results of trials of the relevant sort. ‘Measurement’ of the probability is mistaken for the probability itself.

Operationalism has hit hard times, of course. And rightly so – the arguments are well known. To rehearse just one of these arguments: we want to be able to say that measurements can be misleading (“the thermometers were poorly calibrated”), but an operational definition doesn’t let us say that. Likewise, if the frequentist has his way, we can’t say that the chance of the coin landing heads really was $1/2$, but that there was an unusually high proportion of tails in the actual sequence of tosses. And yet that could be a very natural thing to say.

7. Chance is Supposed to Explain Stable Relative Frequencies

Why do we believe in chances? Because we observe that various relative frequencies of events are stable; and that is exactly what we would expect if there are underlying chances with similar values. We posit chances in order to explain the stability of these relative frequencies. But there is no explaining to be done if chance just *is* actual relative frequency: you can’t explain something by reference to itself. Here I am echoing a well-known argument due to Armstrong (1983) against the ‘naïve regularity theory’ of lawhood (that laws are simply true universal generalizations). Compare: we posit laws of nature in order to explain regularities, so they had better not simply *be* those regularities, as a naïve regularity theory of lawhood

would have it. (Indeed, the demise of frequentism is parallel to that of the naive regularity theory in many respects.)

I have presented firstly some general arguments that work equally well against any of the versions of frequentism that I have mentioned – and indeed, in some cases, even against more sophisticated refinements thereof; and then some further arguments against actual frequentism, irrespective of the size of the reference classes. But finite frequentism also has its own characteristic problems, all really stemming from simple mathematical facts about ratios of (finite) natural numbers.

PROBLEMS SPECIFIC TO FINITE FREQUENTISM

8. Attributes with No Occurrences have Undefined Relative Frequencies. Chance Gaps

Relative frequencies are undefined for attributes that have no occurrences: 0/0 has no determinate value. But I contend that such attributes can have probabilities nonetheless.

Imagine two different worlds, each with a single die. In the first world, the die is tossed a number of times, but in the second it is never tossed. There's a sense in which both dice can be said to have well-defined probabilities for landing 6, say, but according to finite frequentism, only the first does. By analogy, in the first world, the die is weighed, but in the second it is never weighed; nonetheless, both dice have masses.

Ironically, von Mises adduces considerations similar to mine about dice in order to argue for his *opposite* conclusion that probability is relative frequency: “The probability of a 6 is a physical property of a given die and is a property analogous to its mass, specific heat, or electrical resistance” (p. 14). Exactly! But taking the analogy at face value, the conclusion ought to be that probability is an intrinsic property of chance devices (such as dice) – something a propensity theorist might say. The analogy to those other properties presumably appealed to von Mises because of his thoroughgoing positivism, bordering on operationalism. He regarded the mass of a die, for example, as the limit of a sequence of ever improving measurements of the mass. But who among us now would want to say *that*?

All this may devolve into a clash of intuitions, as so many philosophical debates do (which is not to say that the debates are worthless); but I think that the intuitions on my side are perhaps even more compelling in the examples in the next section.

9. If B Occurs Once, A has Probability 0 or 1: Local Determinism

Now suppose that we toss a certain coin exactly once. It lands heads. Then the relative frequency of heads is 1. But we don't want to be committed to saying that the probability of heads is 1, since we want to allow that it could be an indeterministic device. (Change the example to Stern-Gerlach measurements of electron spin, if you think that coin tosses are deterministic; and imagine a world in which there is only the one coin, if you think that the results of tossing other coins, when there are any, are relevant.) And in general, an event that only happens once (according to any sensible standard for typing it) does not automatically do so with probability 1. Such an *a priori* argument for local pockets of determinism is surely too good to be true!⁸

Of course it isn't true. Consider now a radioactive atom that obeys an indeterministic decay law, but as it so happens, there is exactly one such atom in the entire history of the universe (cf. Lewis' (1994) "unobtainium"). Are we to say that its probability of decay is 0 or 1, over any time interval, simply because for each such interval the relative frequency of decays is either 0 or 1? So with probability 1 it decays exactly when it does? This contradicts our supposition that it obeys an indeterministic decay law. An innocuous supposition, surely.

Many experiments are most naturally regarded as being unrepeatable – a football game, a horse race, a presidential election, a war, a death, certain chancy events in the very early history of the universe. Nonetheless, it seems natural to think of non-extreme probabilities attaching to some of them. This, then, is another notorious problem for frequentism: the so-called *problem of the single case*.

10. Universal Generalizations and Existential Statements

Certain statements are 'single case' in virtue of their very logical form: for example, universal generalizations and existential claims. Some people think that non-trivial (objective) probabilities attach to such statements – as it might be, 'the probability that all ravens are black is 0.9', or 'the probability that there exist tachyons is 0.1'. If there is sense to be made of such probabilities, then it is not the frequentist who can make it, for such statements only get one opportunity to be true or false. How do you count cases in which a universal generalization, or an existential statement, is true? What is the reference class, 0.9 of whose instances are 'all ravens are black' instances? I suppose one could imagine counting possible worlds: (in the limit?) 10% of all possible worlds are 'there exist tachyons'-worlds. But this is hardly an attractive proposal, and in any case, it is certainly not finite frequentism.

An ecumenical frequentist might acknowledge some further, non-frequentist sense of probability that covers such cases (“let a thousand flowers bloom . . .”), insisting that frequentism still holds sway elsewhere. The point of this argument is to identify certain sorts of probability statements that people have found quite intelligible, even though they are (virtually) unintelligible on a frequentist analysis. And various real-life frequentists are not so ecumenical.⁹

So far, the problems have involved very low numbers of instances of the attributes in question, namely 0 or 1. So the reaction might be: “frequentism was never meant to handle cases in which there are no statistics, or only a single data point; but in decent-sized samples it works just fine.” This is the intuition encapsulated in the catchy but all-too-vague slogan “Probability is *long run* relative frequency.” The reaction is wrong-headed: problems remain even if we let our finite number of trials be as large as we like.

11. Intermediate ‘Probabilities’ in an Deterministic World

The problem of the single case was that certain relative frequencies are guaranteed to be extreme (0 or 1), even when they are the results of indeterministic processes. This is an embarrassment for frequentism, because such indeterminism is thought to be incompatible with extreme (objective) probabilities — hence those relative frequencies cannot be probabilities. Now let’s turn this thought on its head: *determinism*, it would seem, is incompatible with *intermediate* (objective) probabilities: in a deterministic world, nothing is chancy, and so all objective chances are 0 or 1. But determinism is no obstacle to there being relative frequencies that lie between these values. Remember Venn’s example: the probability of a male birth is simply the proportion of male births among all births. This proportion is presumably roughly 1/2; but the process that determines a baby’s sex could well be deterministic nonetheless.

12. Finite Frequentism Generates Spurious Biases

Consider a coin that is perfectly fair, meaning by this that it lands heads with probability equal to 1/2, and likewise for tails. Yet it might not come up heads exactly 1/2 of the time in actual tossing. In fact, it would be highly unlikely to do so in a huge number of tosses, say 1,000,000.

If the number of tosses is 1,000,001, it would be more than unlikely to do so — it would be downright impossible. So the finite frequentist thinks that we would then be wrong in saying that the coin is perfectly fair. Put simply: according to finite frequentism, it is an *analytic* truth that any coin that is tossed an odd number of times is biased. Now there’s a startling bit of *a priori* reasoning for you.

Likewise, we do not need to leave our finite frequentist arm-chairs to ‘discover’ the biasedness of all n -sided dice that are tossed a number of times that is not divisible by n (a coin can be regarded, after all, as just the special case in which $n = 2$). And so on for other chance processes. If only all empirical matters could be settled so easily!

Furthermore, there is a ‘graininess’ to the possible biases of the coin, or the dice. Toss them n times; the relative frequencies must all be multiples of $1/n$. So not only can the finite frequentist assure us that various coins and dice are biased – he can even put severe constraints on the possible extents of the biases!

He should resist the temptation to reply to all of this: “When we say that the coin is fair, we really mean that it lands heads with probability *approximately* equal to $1/2$, and likewise for tails.”¹⁰ Firstly, there is no guarantee that the fair coin will land heads even approximately half the time. Secondly, we can at least *imagine* a genuinely fair coin, one that moreover is tossed an odd number of times; but the finite frequentist thinks that this is on a par with imagining an uncolored red object – namely, imagining gibberish. Finally, we should not let too much hinge on the choice of the example. Consider if you prefer certain Stern-Gerlach spin measurements, which are perhaps ‘fairer’ than the coin is; or consider the half-life of radium; or whatever your favorite example might be.

Ironically, the longer the finite run of coin tosses (or whatever), the more *unlikely* it is that the relative frequency exactly equals the value that it ‘should’. To be sure, the probability that the relative frequency is *near* the value that it ‘should’ equal increases. But if frequentism is supposed to be an analysis of probability, near enough is not good enough.

13. Finite Frequentism Generates Spurious Correlations

Let us say that A is *spuriously correlated* with B if $P(A|B) \neq P(A)$, and yet A and B are not causally related. The finite frequentist will see spurious correlations all over the place. We can be pretty sure that, say, the relative frequency of people who die by the age of 60 is not exactly the same in general as it is among people who wear green shirts. In fact, we can be absolutely sure that this is so if the smaller sample size happens not to divide the larger one. To see the point, pretend that there are 10 people in our sample reference class, and that 7 of them wear green shirts. (Note that 7 does not divide 10.) Then all relative frequencies within the whole sample must be a multiple of $1/10$, while within the green shirt sample they must be a multiple of $1/7$. Now, there is no way for a multiple of $1/10$ to equal a multiple of $1/7$ (apart from the trivial cases of 0 and 1, which are uninteresting). The finite frequentist translates this as: there is no way

for the probabilities to agree. In other words, a correlation – presumably, a spurious one – between death by the age of 60, and the wearing of green shirts, is guaranteed in this case. Again, it is startling that such results can be derived *a priori*!

14. All Irrational Probabilities, and Infinitely Many Rational Probabilities, 'Go Missing'

There's a good sense in which most of the numbers between 0 and 1 are irrational (uncountably many are, only countably many aren't). Yet a finite relative frequency can never take an irrational value. Thus, any theory which gives such values to probabilities is necessarily false, according to finite frequentism, irrespective of its subject matter. That's certainly a quick refutation of quantum mechanics! For example, according to finite frequentism, the radioactive law for radium is false for all time periods that have irrational probabilities for decay – which is to say that it is false almost everywhere.

Reply number 1 (à la Reichenbach, and very similar to one that we saw above): we can approximate an irrational value as closely as we like, provided we have a sufficiently large (finite) number of trials.

Counter-reply: again, this misses the point. The thesis before us is not that probability is approximately relative frequency, but that it *is* relative frequency. We have an *identification* of probability with relative frequency. Of course, it implies that we can approximate probability values as closely as we like with relative frequency values – anything approximates itself as closely as we like! – but it is a much stronger claim. The point about approximation might be appropriate in justifying relative frequentism as good methodology for discovering probabilities; but our topic is the analysis of probability, not its methodology.

Reply number 2: Bite the bullet, and deny that there are such things as irrational probabilities. No experiment could ever reveal their existence.

Counter-replies: Firstly, this would mean that the truth about various probabilistic laws is more complicated than we think. For instance, the radioactive decay laws would involve step functions, rather than smooth exponential curves. Secondly, the reply smells of positivism. Thirdly, we can imagine possible worlds that instantiate irrational probabilities, even if the actual world turns out not to be one of them. We surely do not want to say that quantum mechanics is not only false, but *logically* false.¹¹

(By the way, *a fortiori* infinitesimal probabilities are ruled out by finite frequentism – and indeed, by any version of frequentism – yet such probabilities may nonetheless have an important role to play. For sympathetic

discussion of infinitesimal probabilities, see for example Skyrms (1980), pp. 177–187.)

Moreover, according to finite frequentism, infinitely many rational probabilities ‘go missing’ also. This is related to the point I made earlier about the ‘graininess’ of finite relative frequencies, for a given sample size. All rational values that fall between the endpoints of the grains will be ineligible as probability values, according to the finite frequentist.

Note that the last few arguments did not require any assumptions about what we take to be the relevant reference classes. As I indicated before, I have misgivings about including in the reference class of a certain coin, the results of tossing *other*, very different or distant coins. But even waiving those misgivings, the last arguments still go through. Include if you like the results of various other coins when determining the probability for *this* coin; indeed, include if you like the results of all coins that ever were tossed, are tossed, and ever will be tossed. Since there will still be only finitely many trials in the reference class, still the frequentist will have to say: all probabilities will be guaranteed to be rational, and in fact, all multiples of a certain finite fraction; spurious correlations with other appropriately chosen factors can be guaranteed; and it is discoverable from the arm-chair that if the total number of tosses is odd, the coins are biased.

15. Non-Frequentist Considerations Enter our Probabilistic Judgments: Symmetry, Derivation from Theory ...

We should regard the various cases above as fatal for finite frequentism, because they provide bullets that cannot easily be bitten. We know that coins and dice cannot so easily be ‘shown’ to be biased, because we sometimes have independent grip on what their various chances are. We know that probabilities of radioactive decay cannot so easily be ‘shown’ to be rational, because quantum mechanics says otherwise. There are other sources of our probability judgments besides relative frequencies – for example, symmetry considerations, and derivation from scientific theories that we already subscribe to. When there’s a conflict between relative frequency and one of these other sources, the latter often wins.

5. CONCLUSION

In this space, I could not give voice to various responses to these arguments on behalf of finite frequentists (although I did give voice to quite a few). They would doubtless reject the starting points of some of the arguments, particularly those that were ‘intuition-based’; other arguments they would

perhaps grant me, remaining untroubled by their conclusions. That should hardly be surprising: most philosophical debates seem to go the same way.

I do, however, think that finite frequentism is about as close to being refuted as a serious philosophical position ever gets. This becomes clear once we have separated the question of how probabilities are discovered from the question of what probabilities *are*. (A good way to find out if a man is a bachelor is to ask him; but we wouldn't want to analyse 'bachelor' as one who answers 'yes' to the question.) To put my position in the form of a slogan: 'Finite frequentism: reasonable methodology, bad analysis'.¹²

NOTES

¹ This paper is an edited version of the first half of my talk "Thirty Arguments Against Frequentism", presented at the Luino Conference. I wanted this paper to reflect that talk, while meeting the reasonable length constraints that this volume required. So I have omitted my lengthy discussion of hypothetical frequentism, hoping to present that on another occasion. I thank the editors for their forbearance.

² I know this from various conversations I have had, though catching such frequentists out of the closet and in print is not so easy. Shafer (1976) comes close in his definition of chance: "... the proportion of the time that a particular one of the possible outcomes [of a random experiment] tends to occur is called the chance of that outcome" (p. 9), and closer still when he drops the qualification "tends to" four pages later.

³ Witness Frieden (1991): "The word 'probability' is but a mathematical abstraction for the intuitively more meaningful term 'frequency of occurrence'" (p. 10).

⁴ The discussion takes place in my manuscript "Fifteen Arguments Against Hypothetical Frequentism", the second half of my talk at Luino.

⁵ The distinction, I take it, is between a conditional probability of the form $P(B|A)$ and a relativized probability of the form $P_A(B)$. The former presupposes that $P(A)$ is defined, the latter does not.

⁶ This resembles somewhat the tension between simplicity and strength in the competition for the 'best' theory of the universe, central to Lewis' (1994) account of chance: raising the standards for admission is like an increase in strength, with a corresponding loss in simplicity. More on that shortly.

⁷ It is hardly better to propose instead that the chances evolve over time exactly as the corresponding relative frequencies evolve over time: for that would mean that the yet-to-be-tossed coin has an undefined chance of landing heads – see the 'chance gap' objection below.

⁸ I cannot pause for further discussion on the connection between determinism and objective chance. I admit that the connections I assume here are not uncontroversial. See Lewis (1986) pp. 117–121 for a fuller treatment.

⁹ Frieden, for example – see footnote 3.

¹⁰ I suppose we already knew that no actual coin lands heads with probability $1/2$ *exactly*, and tails with probability $1/2$ *exactly*, if only because some tiny amount of probability goes to the coin landing on its edge; and perhaps even some (tinier still) amounts of probability go to the coin landing on each of the two edges of its edge.

¹¹ Here I construe logic broadly to include analytic truths.

¹² I am grateful to many people for discussions of this material, especially Jim Bogen, Alex Byrne, Fiona Cowie, Ned Hall, David Hilbert, Marc Lange, Brian Skyrms, Nigel Thomas, Jim Woodward, and Lyle Zynda.

REFERENCES

- Armstrong, D. M.: 1983, *What is a Law of Nature?* Cambridge University Press.
- Frieden, B. R.: 1991, *Probability, Statistical Optics, and Data Testing*, Springer-Verlag.
- Jeffrey, Richard: 1977, 'Mises *Redux*', in R. E. Butts and J. Hintikka (eds.), *Basic Problems in Methodology and Linguistics*.
- Kripke, Saul: 1980, *Naming and Necessity*, Oxford University Press.
- Lewis, David: 1986, *Philosophical Papers*, vol. II, Oxford University Press.
- Lewis, David: 1994, 'Humean Supervenience Debugged', *Mind* 103, 473–490.
- Reichenbach, Hans: 1949, *The Theory of Probability*, University of California Press.
- Shafer, Glenn: 1976, *A Mathematical Theory of Evidence*, Princeton University Press.
- Skyrms, Brian: 1980, *Causal Necessity*, Yale 1980.
- Venn, John: 1876, *The Logic of Chance*, 2nd ed., Macmillan and Co.
- von Mises, Richard: 1957, *Probability, Truth and Statistics*, Macmillan.

California Inst. of Technology
Division of Humanities and Social Sciences
228-77
Pasadena CA 91125
U.S.A.