Week 4: Indifference

Brian Weatherson

2023-09-25

Keynes discusses the Principle of Indifference, his new name for the best hope we have for getting numerical probabilities out of ignorance.

Logicality

First, a confession. Although I've talked about the *Treatise* a lot both here and in written work, before this class I'd never really sat down to read it from start to finish. I basically read it as a collection of somewhat connected papers, reading whatever chapter was relevant to my then interests. Relatedly, I always read it in light of how we now think about things, not ever really trying to get a sense of what it meant to Keynes, or in Keynes's time.

Now that's not entirely a bad way to do things. I think it's fine to have a very instrumental approach to history of philosophy, reading it to see what it can tell us about things that are independently interesting. Indeed, that's usually my view. But sometimes you can overdo it, and be too instrumentalist about how one reads. And that can lead to missing things that are, I think, somewhat interesting. On that note, a second attempt to get clear on what Keynes means by probabilities being **logical**.

There are four ways to understand an author when they say that their theory has some feature.

1. Look for their definition of the feature. That's not much help here since Keynes I don't think is ever particularly clear about what it is to be logical.

2. Look at common usage at the time of the term they use. That maybe would help, but I'm not going to have us sit through three volumes of *Principia Mathematica* to figure this out. I'll leave that to more serious historians.

3. Look at the argument they give for the feature. This turns out not to help because the argument Keynes gives is really bad. He just argues that probabilities are necessary and a priori. And while that's interesting, it really isn't enough to be logical in any interesting sense.

4. Look at how they apply that feature. And here we might be getting on to something. Because the discussion at the end of chapter 4, and into chapter 5, suggests something about a stronger thing Keynes might have meant. And it is, I think, interesting enough to digress into.

Chapter 4, and onto chapter 5, have an odd place in the history of philosophy. The first half of chapter 4 is the best place to find a clear early statement, in English, of the problems with Indifference. It's a devastating tour de force of objections. But the second half does not do what you expect if you only know the first half. It doesn't say "Indifference, begone!". Rather, it says that it's time to find a restricted version of Indifference that avoids the (many, many) problems, and that this restricted version will ultimately be the basis for all probability.

Why might that be relevant. Well, think about four properties that we expect a logical relation to have. To be concrete, let's focus on what it means for a particular sentence to be a logical truth, like a particular substitution instance of Peirce's Law: $((p \to q) \to p) \to p$.

1. **Necessity:** This sentence is true in all possible worlds.
2. **A Priori:** It is knowable a priori that this sentence is true.
3. **Analyticity:** This sentence is true in virtue of its meaning.[1]
4. **Formal:** This sentence is true (and perhaps necessarily, a priori true) in virtue of its form.

Now when Keynes introduces the idea that probability is logical, he only really talks about 1 and 2, and that's why I was kvetching last week that we don't get a lot here. And, I thought, we really need at least 3, and ideally 4. But we don't get it. And, as we now know, we can't get it. The argument for this has two steps. The first step is to identify 4 with the following constraint.

[1] There is a connection here to Keynes defining probability as a relation between sets. There is an argument, perhaps due to Michael Dummett, that Peirce's Law is not a logical truth because it isn't true in virtue of its meaning. And the argument turns on identifying the meaning of $\to$ with its role in a deduction system where arguments have many premises but just one conclusion. If you allow multiple conclusions, Peirce's Law is analytic. I can say much more about this if anyone's interested, but it's really a digression.

5. **Formal-New:** Any sentence with the same form as this one is also true (and ideally also has 1, 2, and 3).

The second step is to show that probability relations do not satisfy 5. That's easy enough given some concepts that we have and Keynes did not have. Start with the following definitions.

- An emerald is grue iff it is (a) examined by some human and green, or (b) not examined by any human and blue.
- Let h_1 be that lots of us have examined lots of emeralds in lots of places, and they've all been green.
- Let h_2 be that lots of us have examined lots of emeralds in lots of places, and they've all been grue.
- Let a_1 be that the next emerald a human examines will be green.
- Let a_2 be that the next emerald a human examines will be grue.

Now the problem for Keynes. First, a_1/h_1 is very high. Second, a_2/h_2 is very low. Third, a_2/h_2 has the same form as a_1/h_1. Then we can reason.

1. Probability does not satisfy **Formal-New**.
2. Therefore, probability does not satisfy **Formal**.
3. Therefore, probability is not logical, contra Keynes.

Now question, where would Keynes reject this reasoning? Maybe he wouldn't, maybe he'd say that oops, hadn't thought about that example, back to the drawing board. And look, that's not an unreasonable guess; he really didn't know about 'grue'. But let's see if we can do a little better.

When I was a student, and probably when any of you were students, there was no sharp distinction made between **Formal** and **Formal-New**. I think in the textbooks I used, **Formal-New** was just taken to be the translation of **Formal** into a mathematically precise language. But there are two important things to note about this. First, it really is a theory of how to understand 4. Second, it wasn't (to the best of my knowledge) a theory in Keynes's time.[2] What happens if we reject the analysis of **Formal** as **Formal-New**, and hence the step from 1 to 2 in the last argument?

There are some twentieth century precedents for this? Not, to be sure, one Keynes could possibly have known about, but there are

[2] Again confession time: I haven't read nearly enough of *Principia Mathematica* to be sure about this, but I'm pretty confident that it is from circa the 1930s.

precedents. In so-called positive free logic, the inference from Fa to $\exists x : Fx$ is valid iff F is 'positive'. So you can infer from *Brian is Australian* to *There is an Australian*, but not from *Santa Claus is unreal* to *Something is unreal*. This is usually spelled out as a constraint on the possible interpretations of the predicates: F is only allowed to designate some positive property, like *Australian* and unlike *unreal*. In some non-normal modal logics, if p is a primitive sentence-letter, then $\diamond p$ is an axiom. Again, it's important here to put constraints on what sentence letters can designate. You can't have p mean $a \neq a$, for example. These are obscure logics (to put it mildly), but they are definitely logics.[3] Maybe some similarly restricted notion of 'same form' can rescue Keynes here. That actually seems plausible to me, but I want to try a more twenty-first century defense instead.

The understanding of **Formal** as **Formal-New** follows a very common twentieth century paradigm. We understand claims of the form *X is true in virtue of Y* as invariance claims: No variation in *X* without variation in *Y*. That's what's going on here. **Formal** is an in virtue of claim: Peirce's Law is true (/necessary/a prior) in virtue of its form. **Formal-New** is an invariance claim: you can't change the truth (/modal status/epistemic status) of an instance of Peirce's Law without changing its form. Merely changing its non-formal elements (i.e., the meanings of p and q) won't be enough. In the twentieth century, these understandings of 'in virtue of' claims as invariance claims were everywhere in metaphysics, theory of mind, logic, epistemology, and elsewhere.[4]

In the twenty-first century, there has been a huge, discipline wide, rejection of the equation of these two types of claims. This largely traces to work Kit Fine did in the early 2000s, though it didn't really take hold until around 2009. It's often thought that that the connection fails in both directions.

You can have invariance without the in virtue of claim being true. I'm Australian in virtue of being born in Australia. But it's not true that: I'm Australian in virtue of God knowing that I was born in Australia. And that's the case even if it's necessary that God knows everything, and so there's no variation in my nationality without variation in God's knowledge of where I'm both.

And (I think, though not everyone agrees), you can have in virtue of claims being true without the invariance claim being true. Assume,

[3] You can read about both of them in Graham Priest's quite incredible *Introduction to Non-Classical Logics*. He thinks they are pretty weird, which is something coming from *Graham Priest*, but he does carefully set them out.

[4] The invariance claims were usually taken to be necessary truths, but they need not be. Jerry Fodor's asymmetric counterfactual dependence theory of mental content is an invariance theory where the invariance is contingent.

as I think is right, that lying is wrong other things equal, but sometimes there are good reasons to lie. And assume that x lies in a situation where there was no good reason. Then I think this is simply true: x did something wrong, and it was wrong in virtue of being a lie. But I think this isn't true: x did something wrong, and the moral status of x's act couldn't change without changing whether it was a lie. The last thing is false because if there were special circumstances, it would have been ok to lie.

I suspect the best way to interpret Keynes, the way that makes him say the most interesting and plausible things, is to similarly say that he rejects the understanding of **Formal** as an invariance claim. True instances of Indifference are true in virtue of their form. That does not mean that anything with the same form is also true, any more than saying that x's action is wrong in virtue of lying means that all lies are wrong. What it does mean is that when you have a correct instance of Indifference, it's the formal features of the claim that *explain* why it is true, and which *make* it true. (Compare: In normal circumstances, it's that it's a lie that explains why it is wrong, and which makes it wrong.)

I don't mean to say this strategy will work. It certainly doesn't do enough to save the details of Keynes's theory, which we've known at least since Carnap's *Logical Foundations of Probability* (1950) don't work out. But maybe it saves the idea that if (a) there is a restricted formulation of Indifference that works and (b) all true probability claims are ultimately derivable from true instances of this restricted formulation, then (c) it would be fair to call the theory logical.

"Indifference"

Every time I read this chapter I'm surprised again at learning Keynes is responsible for the name "The Principle of Indifference". It has stuck so thoroughly. And it is so much better than "The Principle of Non-Sufficient Reason", which would make everyone think we were about to get into PSR and Leibniz and the cosmological argument. Anyway, it's a good name, for what may or may not be a good principle.

Remember that Keynes thinks it is a wide-open question whether an argument may even have a probability. So let a be some proposition

wholly disconnected from *h*. Then even if we'd convinced ourselves that *a / h* couldn't be any number other than 0.5, it doesn't follow for Keynes that it is 0.5. It might just be non-numerical.

Counterexamples to Indifference

Keynes kind of goes to town on the PoI in paragraphs 4-9. There are two big things to note here, neither of which I think is always properly reflected in the secondary literature on this problem.

1. This whole discussion is an in-house argument among people who have broadly epistemic theories of probability. If your response to these examples is "We need to know more about the physical setup, or underlying facts, or ...", then that means you really don't believe the epistemic theory of probability. That's fine - it's a controversial theory. But Keynes isn't arguing with you here; he's talking to his fellow epistemic theorists.
2. There are a *lot* of examples here. Don't think that by finding some solution to one of them, you've found some general solution to the problems with indifference.

One that isn't there - the cube factory

The cube factory example that van Fraassen uses in *Laws and Symmetry* isn't quite there in Keynes. A similar example is. Van Fraassen's example is that we know a factory produces perfect cubes of side length at most 2cm. Given that knowledge, and nothing else, what is the probability that the next cube has a side length under 1cm? Related question. We know that a factory produces perfect cubes of volume at most 8cm^3. Given that knowledge, and nothing else, what is the probability that the next cube has a volume under 1cm^3? Uh oh, these are the same question, but PoI (Principle of Indifference) says they get different answers.

Keynes will have an example like this, but it's really striking how much variety there is to his counterexamples. One rather often sees people try to resolve the cube factory problem and thereby conclude that they have somehow rescued some version of the PoI.

Three Book Colours

We know nothing about a book. What's the probability that it (or I guess its cover) is red? What about blue? What about black? If PoI says that they are each 0.5, we get a contradiction given the background knowledge that nothing is red, blue, and black.

Small question: Are these contraries? Surely some covers are blue and red. Does getting clear on the nature of the claim make it more plausible that the PoI wouldn't apply.

Adding a Detail

Assume we know nothing whatsoever about an object A. Consider the proposition *A is a book*. Since we know nothing about whether it is or isn't a book, PoI says that its probability is 0.5. Now consider the proposition *A is a red book*. Since we know nothing about whether it is or isn't a red book, PoI says that its probability is 0.5. Putting these together, we get that the probability that A is red, conditional on being a book, is 1. That's absurd.

Having Names for Parts

We know nothing about the world other than that it contains people, each of whom inhabit a point, and that it has regions which have points in them and that Great Britain, France, the British Isles, and Ireland, are among those regions. B is one of the people in the world. By PoI, each of the following are equiprobable: *B is in France, B is in Great Britain, B is in Ireland, B is in the British Isles*. We then learn that the British Isles just are Great Britain plus Ireland.[5] Either the PoI leads to a contradiction here, or each of these has probability 0, or learning that a region has named parts tells us *a lot* about the probability of people being in it. None of these is particularly plausible.

[5] Did Keynes have to do last minute changes to this example after Irish independence?!

Volume and Density

This is the closest we get to the cube factory (and the one that Keynes most clearly attributes to someone else). Assume we know nothing about a substance except that it's specific volume[6] is between 1 and 3.

[6] i.e., the number of cubic meters occupied by 1kg of the stuff

PoI suggests that the probability that the specific volume is between 1 and 2 is 0.5. Compare, we get a substance and no nothing except that its specific density[7] is between 1/3 and 1. By PoI the probability that the specific density is between 1/3 and 2/3 is 0.5. But this is inconsistent with what we say about volumes.

[7] i.e., the inverse of the specific volume; this is not a term in widespread usage these days

Keynes uses a similar example when worrying about averaging prices in his 1909 essay on Index Numbers. Should we average the dollars per unit, or the units per dollars, when finding out the average price of various things? He argues that there is no particular reason to prefer one to the other.

Bertrand's Chord Example

I'm not going to try explaining this one here, but we possibly should talk through it in class. Again, it feels somewhat like the cube factory, but I'm not sure it is subject to the same objections. (I do think the sphere example is basically the same as the chord example, but I might be misunderstanding the geometry here, since I was never very good at solid geometry.)

The real point Keynes makes is at the end of paragraph 7 about f. Given the intuitions behind the PoI, (a) any two equally sized ranges of a function should have equal probability, and (b) probability should be preserved under bijections. (E.g., the probability of the side length being between x and y just is the probability of the volume being between x^3 and y^3. But given there are monotonic increasing functions that do not preserve ratios of ranges, this can't work.

Magnitudes and Learning

This is in a sense another one of the counterexamples to PoI, but I've separated it out because it has such dramatic consequences. Start with a simple case. There is an urn with two balls in it, and all we know is that each ball is black or white, but we do not know the distribution. What does the PoI tell us should be the probability distribution. It turns out there are two different answers here that seem equally plausible at first blush.

First answer. There are three possibilities:

1. Two black;
2. One black, one white;
3. Two white.

Second answer. There are four possibilities:

1. The first is black, and the second is black.
2. The first is black, and the second is white.
3. The first is white, and the second is black.
4. The first is white, and the second is white.

On the first way of thinking, the PoI says that the probability of an all white urn is 1/3; on the second way of thinking, the PoI says that the probability of an all white urn is 1/4. Keynes goes on to give a reason for thinking the second way is right, and I guess as far as it goes that seems fair.

But think about an example that isn't obviously different in form to this one. There are n swans in the world, and all we know *a priori* is that swans are no colour other than black or white. What are the possibilities that PoI says we should be indifferent between. Again, two possibilities jump out.

The first says that we should be indifferent between the $n+1$ possibilities of the form, *There are no white swans*, *There is one white swan*, *There are two white swans*, etc. The second says that we should be indifferent over the 2^n distribution of black-white over the different swans.

The second does seem in some ways more natural. But it has a shocking consequence. It makes inductive learning impossible. Consider all questions of the following form.

- Assume you start with the probability distribution recommended by the PoI, and learn that the first k swans are white. What, as a function of k, is the probability that the next swan is white?

Answer, on the second model, is that it is 1/2 no matter what. That seems absurd, so either (a) the PoI is wrong, or (b) it has been misapplied here, or (c) learning does not go by conditionalisation.

I think this is a problem even for the weaker, heavily qualified, version of the PoI that Keynes ends up endorsing, and it really ends up being a problem for lots of versions of the PoI. It might be, as I once argued,

that (c) is the villain here, so this isn't actually a problem for PoI. But it's a worry.

General Irrelevance

On page 60 Keynes makes a bad slip, and it's going to have significant consequences for us down the line. He is asking when the addition of evidence h_1 is irrelevant to x/h. He first gives the familiar contemporary answer: it's when $x/h = x/hh_1$. But then he says something stronger.

> h_1 is irrelevant to x on evidence h, if there is no proposition, inferrable from h_1h but not from h, such that its addition to evidence h affects the probability of x.

And this rules out too much[8]. On this definition, practically will be irrelevant. The problem is that the disjunction $h_1 \lor x$ is (typically) inferrable from hh_1 but not from h alone and (almost always), $x/h < x/h(h_1 \lor x)$.

[8] Carnap, for instance, says this in *Logical Foundations of Probability*.

As I said, I think this is, and similar examples, are going to land Keynes in all sorts of hot water. Some of the problems will come up later in this chapter.

I said last week that the sets of probability function model is, I once argued, a pretty good formalisation of Keynes's intuitions. This point is relevant to that idea, in a roundabout way. On the sets of probability functions approach, you really don't get anything like Keynes's general notion of irrelevance. But that's a *good* thing, because we should want to be rid of it, and any part of his theory that relies on it. What parts are those? Well, that's a question we'll need to keep an eye on.

The points at the top of 61, which are nowadays commonplaces, are important observations. Irrelevance (in either Keynes's defective strict sense, or our more familiar loose sense), is closed under negation, and is symmetric. These are sometimes useful things to remember.

Keynes's Positive Version of the PoI

After so many attacks on the PoI, it is somewhat surprising, to me at least, that Keynes ends the chapter by offering his own version of the PoI. It is very heavily qualified, as we're about to see more qualified than Keynes realises, but it is a version of the PoI. When I got to this part there were two problems I expected to find, which in retrospect I'm not sure were there. But there's a much bigger problem that relates to the previous point about irrelevance.

Countable Additivity

First, I thought that the PoI would lead to some kind of problem with countable additivity. Keynes is careful to say that he only intends the PoI to apply to finite cases. But it's very hard to get a philosophical principle here that doesn't lead to countable additivity violations. Here's the basic idea.[9]

The cash value of Keynes's version of the PoI is the following idea. Let $\phi(x)$ be some propositional radical such that $\exists x : x \in S \wedge \phi(x)$ is known (and is more or less trivial) for some set S, and so is $\forall x \forall y (\phi(x) \wedge \phi(y)) \rightarrow x = y$ is also trivial. In words, exactly one thing is ϕ, and it is in S. And let the evidence h be symmetric with respect to any two substitution instances $\phi(a), \phi(b)$, at least where $a, b \in S$. Then what Keynes really wants is that $\phi(a)/h = \phi(b)/h$. And that sounds kind of plausible enough, perhaps even after everything that we've seen previously in chapters 3 and 4.

That principle works fine when S is finite; the probability of any one thing being the ϕ is $1/|S|$. And it works fine when S is uncountably infinite. The probability of any one thing being the ϕ is 0, but that's probably what we want. And you can't deduce any annoying things about cube volumes or substance density from these 0 values.[10]

But it very much does not work when S is countably infinite. In that case we get that the probability of any one thing being the ϕ is 0, but the probability of one of these things being the ϕ is 1. So the probability of this countably infinite disjunction, i.e., the first one is the ϕ, or the second one is, or the third one is, or ..., is not equal to the sum of the probability of the disjuncts.

[9] Jake Ross has a couple of papers on Sleeping Beauty, one in *Phil Review* and one in *Phil Studies* that make a similar point. And I make a similar complaint, less clearly, in my objection to Adam Elga's version of the PoI in an old PPR paper.

[10] Does everyone understand why? Maybe we should pause over this if not.

Now it's a *very big question* in the last 100 years of philosophy of probability whether countable additivity is a requirement. It's somewhat disappointing that Keynes doesn't really address it. Indeed, he just seems to treat the only salient division among sample spaces to be finite vs infinite, whereas the modern tendency is to treat the big division as finite vs countably infinite vs uncountably infinite. Plenty of people think that it's obviously true that you can have an equiprobable distribution over a countable set, so finite additivity must fail. And there are other reasons people have had for rejecting countable additivity.[11] Still, I think countable additivity is a good constraint, and if the PoI violates it, all the worse for the PoI.[12]

Syntax and Semantics

Keynes's version of the PoI makes heavy use of a *syntactic* constraint. It's all about whether the competing claims, which are meant to be equiprobable, can be put into the form $\phi(a), \phi(b)$. The contrast is with *semantic* constraints on indifference principles. Adam Elga, in his Dr Evil paper, has such a principle[13]. Two propositions are equiprobable iff there is the right kind of bijection between the sets of centred worlds the two propositions represent.[14] And it's very natural nowadays to have semantic versions of these principles.

The syntax/semantics split here recurs in a bunch of related fields. It is kind of the same as the debate between proof-theoretic and model-theoretic conceptions of validity in philosophy of logic. And it is kind of the same as the debate between syntactic and semantic conceptions of laws in philosophy of science. And there are probably debates in metaphysics that have the same structure; or, if not, it's only because the semantic side so comprehensively won.

And outside of logic, the semantic side has basically been on a 75 year winning streak, for a fairly simple reason. A very widespread view (which I am definitely sympathetic to) is that syntactic approaches can't help with anything to do with probability/epistemology because of the grue paradox. Maybe a syntactic approach with a restriction to a special subset of the language can work, but it would take some proving that it did work.[15]

Keynes has a syntactic approach for a few reasons. One is that it's the early 20th century, and nobody was using semantic approaches;

[11] Dmitri Gallow, formerly of this parish, just posted a note arguing against countable additivity on grounds that should appeal to a Keynesian, because it appeals to the idea that conditional distributions are primitive. I think it's a variant on the idea that if X is a flat distribution over the reals in [0, 1], then conditionalising on the evidence that X is rational will give you a violation of countable additivity. But that's simplifying Dmitri's views a little bit I suspect.

[12] If we have time, I'll go over the conglomerability argument for countable additivity, and maybe why not everyone buys it.

[13] The paper is called "Defeating Dr Evil with Self-Locating Belief", and it's in *Philosophy and Phenomenological Research*.

[14] There's a bunch of jargon in that last sentence, and I'm more than happy to go over it. But if you don't want to know the details, the big picture point I'm trying to make here is that Elga's version doesn't look at how the two propositions are textually represented, but at the ways in which they are made true.

[15] This is related to the points I was making about logicality.

everything was syntactic. Another was that the grue paradox hadn't been discovered. And a third, perhaps most important, was that Keynes was really committed to the idea that he was extending *logic* here, and everyone (even the model theorists) agrees that logic has *something* to do with syntax.

Still, when I read this my first thought was, Ah, Keynes is bringing syntactic weapons to a semantic fight, and this is bound to fail. So I started trying to come up with grue-like counterexamples to his theory. And I didn't quite find them. Because there's a more fundamental problem with the view.

Disjunctive Parts

Let's look a little more closely at precisely what Keynes says is the constraint on the PoI.[16]

> The Principle of Indifference is applicable to the alternatives $\phi(a)$ and $\phi(b)$, when the evidence h is so constituted that, if $f(a)$ is an independent part of h (see §14) which is relevant to $\phi(a)$, and does not contain any independent parts which are irrelevant to $\phi(a)$, then h includes $f(b)$ also.

What is an 'independent part'? When we look back at 14, we find this.

> It will be convenient to define also two other phrases. h_1 and h_2 are independent and complementary parts of the evidence, if between them they make up h and neither can be inferred from the other.

But now we have a problem. For any h, ϕ, we can define an f as follows

$$f(x) =_{df} (h \lor \phi(a)) \land (h \to (a = x))$$

If $x = a$ that is entailed by h, and together with $h \lor \neg\phi(a)$ entails h. So those two are complementary and independent, and hence $f(a)$ is an independent part of h. But $f(b)$ is false given h, so clearly not

[16] Carnap, in *The Logical Foundations of Probability* is very critical of Keynes on just this point. He doesn't quite have the example I'm giving below, but essentially the criticism I'm making here is in Carnap.

part of h. So for any h, ϕ, a, b there will be an f such that h includes $f(a)$ and not $f(b)$, and $f(a)$ is an independent part of h.

Does $f(a)$ satisfy the other clause Keynes gives that it "does not contain any independent parts which are irrelevant to $\phi(a)$"? I'm not sure. This has gone on long enough, and maybe I'll leave that one as a question for the readers.