

Peirce, Keynes, and the Weight of Evidence

Lok C. Chan

1 Weight Chapter

1.1 Keynes's Evidential Weight

1.1.1 Background

Even though the notion of the weight of evidence is often associated with Keynes, he is rather ambivalent about it in **A Treatise on Probability**. He thinks that while the question of weight is “somewhat novel” but he is unsure of its importance.¹ His discussion, nevertheless, is valuable for us, because the notion of evidential weight emerges directly out of Keynes' system of logical probability.

Probability, in Keynes's view, is defined as a probability relation between two propositions. These are objective and logical relationship that are independent of our subjective state, even though it can only be known through intuition. In order to render the numeric assignment of **some** probabilities

¹John Maynard Keynes, *A Treatise on Probability* (Macmillan; Co., Limited., 1921), 78.

possible, Keynes proposes a number of basic judgments about probability. We make **judgments** of indifference, for instance, when we judge that probabilities are uniformly distributed for a set of events.

More pertinent to us is the **judgment of irrelevance**. Keynes' observation is that we often can judge whether one proposition E counts as being relevant to another proposition H by considering whether the probability of H would change on the supposition that E is true. Keynes's example is that, in a typical urn example with some black and white balls, if we want to know the probability of a white ball being randomly chosen, the *color* of the ball would not change its probability of being chosen, so the idea is that a ball's probability of being chosen conditional on being (say) white is the same as the probability of the ball being chosen in general.² So, Keynes proposes that evidence E is irrelevant to the proposition H if and only if:

$$P(H|E) \neq P(H)$$

1.1.2 Weight vs Balance

Now that the notion of relevance has been introduced, we finally come to Keynes' idea of the weight of evidence. He first brings our attention to the fact that when we consider the conditional probability of the hypothesis in question under all relevant evidence, the resultant number constitutes the **balance** between favorable and unfavorable evidence.³ For instance, we may say that

²Keynes, 59.

³Keynes, 78.

when the probability is above 0.5, then the overall evidence is somewhat in favor of the hypothesis. Of course, the balance changes as we gather more relevant evidence, and it might go from favorable to unfavorable depending on the nature of the new evidence.

However, as Keynes points out, this is not the only way in which we think of how evidence can render a hypothesis acceptable or unacceptable, for we not only care about how much the current evidence favors the hypothesis, but we also concern ourselves with *amount* of information available, and Keynes calls this measure of this amount the *weight* of evidence. But, unlike the balance of the evidence, which can go either direction, the weight of evidence can only go up as we gather more relevant evidence. In Keynes' words, "New evidence will sometimes decrease the probability of an argument, but it will always increase its 'weight.'"⁴

A critical point Keynes makes in the just mentioned quote is that evidential weight forms a linear relationship with the amount of *relevant* evidence: bringing in new evidence *always* increases the weight of the evidence. As a matter of fact, Keynes suggests that as a matter of definition, to introduce new relevant evidence is simply another way to say that the weight of evidence has increased.⁵

Keynes, however, notices a technical difficulty. The problem is that sometimes complex evidence E might be irrelevant to H , but its logically simpler components E_1 and E_2 could be relevant to H . What Keynes has in mind seems to be this: suppose you have reasons to believe that the urn in

⁴Keynes, 78.

⁵Keynes, 78–79.

front of you contains 50 white balls and 50 black balls, so your probability for H , drawing a black ball, would be 0.5. Further suppose E_x says ‘a ball with the color x is drawn’, where $x = w$ for white and $x = b$ for black. Let $E = E_b \wedge E_w$. At this point, an argument could be made that your degree of belief for H should stay at 0.5, since sample mean is exactly that. In fact, this is a situation in which the use of Laplace’s “rule of succession” is justified..⁶ According to this rule, $P(H|E_b \wedge E_w)$ is

$$\frac{1 + 50}{2 + 50 + 50} = \frac{51}{102} = \frac{1}{2}$$

But suppose E_b and E_w are not gathered at the same time: say E_b was received first - this would sensibly nudge your opinion slightly toward the opinion that the proportion of black is higher than 50%. This could be verified by the rule of succession: $P(H|E_b) = \frac{51}{101} \neq P(H)$, so E_b is relevant to H . *Mutatis mutandis*, the same argument could be made for E_w . So we have a situation where getting E_w and E_b together does not increase the weight of the evidence, but getting them sequentially would lead to an increase. Though he, too, has to speculate Keynes’ precise meaning, Carnap [419] gives a similar but formal example. Let us call this the problem of partial relevance.

To avoid this problem, Keynes proposes a stricter definition of irrelevance. E is irrelevant to H on background knowledge K if and only if there is no

⁶‘... Thus one finds that when an event has happened any number of times running, the probability that it will happen again next time is equal to this number increased by 1, divided by the same number increased by 2.’, Pierre-Simon Laplace, *Philosophical Essay on Probabilities* (Spring, 1995), 11

proposition ϕ , such that

1. ϕ is inferable from $(E \wedge K)$.
2. ϕ is not inferable from K .
3. $P(H|\phi \wedge K) \neq P(H|K)$

This roughly states that E is irrelevant to H on K if and only if there is no proposition implied by E and K together, but not K by itself, such that it would change the probability of H . So in the case where $E = E_b \wedge E_w$, E is *not* irrelevant to H , because E_b , which is inferable from $E \wedge H$, does raise the probability H . Also, this definition also means that any proposition that satisfies the three conditions is relevant. In other words, a piece of evidence is relevant so long as a part of it is relevant.

1.1.3 The Problem with Trivially Relevant Evidence

However, this definition has some peculiar implications. As pointed out by Carnap, this has the consequence that any proposition E not implied by background knowledge K would be relevant to hypothesis H conditional on K . [420] The idea is that Keynes's new definition has essentially rendered anything not implied by background knowledge relevant, because the problem is that any proposition that is not implied by K will always have implied components that are relevant to H , however weakly. The argument is roughly that, for any H on background knowledge K , if J is not implied by $H \wedge K$ and R is relevant evidence to H , J can always be shown to be relevant by drawing trivial inferences that that involves R , such as $J \vee \neg R$.

For a formal argument, we need to some minimal assumptions about relevance. (1) If E is relevant, then $\neg E$ is also relevant. (2) If a hypothesis is not implied by our background knowledge, then there is some proposition such that it is relevant to the hypothesis. It follows that:

1. Suppose J is not inferable from hypothesis H and background knowledge K , and R is relevant to H on K .
2. Let E be $(R \wedge \neg J)$: E is relevant to H , since R is relevant to H on K and E contains R .
3. Since E is relevant to H , $\neg E$ is also relevant to H . (By assumption (2))
4. But this means J is relevant to relevant H as well: $\neg E$ is equivalent to $(\neg R \vee J)$
5. Since an implication of J is relevant to H given K , this means J must be relevant to H .
6. Thus, if J is not inferable from K , then J is relevant to K .

So Keynes' attempt to solve the problem of partial relevance seems to prove too much, but recall that Keynes proposes the stricter definition of relevance in order to forge the direct connection between relevance and weight, and this is motivated by the assumption that weight always increases when relevant evidence is introduced. But if relevance can be trivially increased by simply introducing non-sequitur propositions into our body of evidence,

1.1.4 The Paradox of Ideal Evidence

Popper has illustrated a problem between Keynes's notion of relevance and weight by what he calls the *paradox of ideal evidence*.⁷ He asks us to consider a certain coin with an unknown bias: let N be the proposition "the next toss of the penny will yield heads". Now, what should $P(N)$ be? He suggests, either by appealing to intuition or the Principle of Indifference, it ought to be clear that $P(N) = 0.5$.⁸

Now let I be what he calls *the ideal statistical evidence* in favor of the idea that the coin in question is a fair one. Popper's example is to let I be the proposition that 'in a million tosses, the coin landed on heads roughly half a million times.' The exact number is not important, as long as the number of heads and tails would make it practically certain that the coin is fair - the same point could be made using 10 millions instead of a million. The crucial point here is that I is a basic summary of the relevant statistics: that is, I should *not* be construed as a chain of conjunctions of propositions $T_1 \wedge \dots T_i \dots \wedge T_n$, each of which indicates the result of a specific toss. Now, given we have ideal evidence I , what is the probability of N ? Popper claims that it would have to be $1/2$. So

$$P(N|I) = P(N) = \frac{1}{2}$$

⁷Karl Popper, *The Logic of Scientific Discovery* (Routledge, 2002), 425.

⁸It should be noted that Popper is not attacking the principle of indifference in this context. That is, for this argument he is willing to grant that Bayesians have some way of arriving at $P(N)$ - it could be by indifference, through elicitation, etc.

However, as discussed earlier, evidence E is irrelevant to the hypothesis H if and only if

$$P(N|I) = P(N)$$

If $P(N|I) = P(N) = 1/2$, this means that the ideal evidence is also irrelevant evidence. Popper then concludes:

Now this is a little startling; for it means, more explicitly, that our so-called ‘degree of rational belief’ in the hypothesis, $[N]$, ought to be completely unaffected by the accumulated evidential knowledge, $[I]$; that the absence of any statistical evidence concerning [the hypothesis that the coin is fair] justifies precisely the same ‘degree of rational belief’ as the weighty evidence of millions of observations which, *prima facie*, support or confirm or strengthen our belief.⁹

What is ‘startling’ about this? This returns to Keynes’ initial observation: *something* is changed by the ideal evidence, but it is not reflected in the probability. Now, ignoring the problem with trivial evidence, one may think that a problem like this is exactly what Keynes’ restricted definition is supposed to solve.

However, recall that I was stipulated to be a basic report of the statistical result of the ten million tosses. The reason for that is to preemptively block Keynes’ strict definition of relevance.¹⁰ Since what we have is simply a basic

⁹Popper, 426.

¹⁰Rod O’Donnell, “Keynes’s Weight of Argument and Popper’s Paradox of Ideal Evidence,”

statement of summary statistics, there is no way to decompose the I such that its component would change the probability of N , except in the vacuous sense spelled out in Carnap's example.

So now it seems that Keynes' restricted definition does not solve the problem it was meant to address. Even with the restricted definition, there is still What our discussion suggests I think, is that we ought to give up the tight connection between relevance and weight on which Keynes insists. The problem with trivially relevant evidence is caused by Keynes' attempt to account for evidential weight in terms of the absolute amount of relevant evidence. Carnap's proposal is that we ought to stick with the simple definition of conditional relevance.

Once we let go of that assumption, the paradox of ideal evidence has a straightforward answer: the so-called ideal evidence is ideal not because it is ideally relevant, but because it increases our weight considerably. This response, again, would not be possible if we insist on the necessary connection between relevance and weight. Nevertheless, while the force of Popper's argument is weakened, it is not entirely eliminated. By separating relevance and weight, it is true, the paradox seems to vanish, but to address it adequately we must be able to state what weight could be.

But if not Keynesian weight, then what? In the next section, we shall turn to Peirce's view on the matter.

Philosophy of Science 59, no. 1 (1992): 47.

1.2 Peirce's Evidential Weight

1.2.1 Background

While the idea of evidential weight is closely associated with Keynes' discussion in **A Treatise**, Carnap [554] and Popper point out that Peirce was really the first person who gave a substantial treatment of the issue. Nevertheless, discussions of Peirce's notion of weight are often What I think has gone under the radar is how evidential weight figures in Peirce's other philosophical views, especially his view on abduction, belief-doubt epistemology, and pragmatism.

There are three possible reasons for which the importance of Peirce's view on the matter has gone relatively unnoticed. One of the reasons is that Peirce's explicit discussion of evidential weight is embedded in his criticisms of subjective interpretation of probability, known as "conceptualism", in his time. Like Keynes, Peirce recognizes that there is a sense of confidence that is not captured by degrees of belief, and Peirce argues that this is a fatal flaw of conceptualism. Since evidential weight only seems to occur in a critical context, it perhaps gives the impression that Peirce lacks a positive view on the matter.

The other reason is terminological. Peirce uses both the terms 'balance' and 'weight' to describe evidence, but from the context it is clear that in both cases he's referring to idea that probability can act as an indicator of whether a belief is favored or disfavored by the overall evidence - what Keynes calls the balance of the evidence. Peirce in fact never gives a name to what Keynes, and what came to be known in the literature as, the weight of evidence. He simply

refers to that as the number in addition to probability that is required to properly express our doxastic state.

Lastly, Peirce’s own mathematical expression of the balance of evidence adopted by various Bayesians, which further overshadows his notion of weight. Roughly speaking, Peirce’s idea is that to find the balance of the evidence E for the proposition H , we should take the odds of the probability of H conditional on E versus its probability conditional on $\neg E$. In other words,

$$\frac{P(H|E)}{P(\neg H|E)}$$

Using the conceptualist’s principle of indifference, which renders $\frac{H}{\neg H} = 1$, the expression above then becomes $\frac{P(E|H)}{P(E|\neg H)}$, which is an expression of what we now call the Bayes Factor. I. J. Good, who refers to Peirce as a precursor to his own Bayesian view, confusingly calls this formula the weight of the evidence, which, while consistent with Peirce’s use of the term, is not the Keynesian sense that come to be the standard.¹¹ Branden Fitelson takes Peirce’s idea of taking the logarithm of posterior odds and develops a Bayesian account of independent evidence.¹²

¹¹James M. Joyce, “How Probabilities Reflect Evidence,” *Philosophical Perspectives* 19, no. 1 (2005): 165.

¹²Branden Fitelson, “A Bayesian Account of Independent Evidence with Applications,” *Proceedings of the Philosophy of Science Association* 2001, no. 3 (2001): S123.

1.2.2 The Concept of Weight in “Probability of Induction”.

Peirce’s illustration of weight is roughly as follows: imagine two urns A and B with unknown proportions of black and white balls. Suppose you sample (with replacement) 100 balls from the urn A and find 50 black balls and 50 white balls. Justifiably, you infer that the proportion of black balls in A - call it θ_A is about 0.5. You then decide to sample from B , but this time you only manage to draw 4 samples, 3 of which are black balls. Your best estimate for θ_B is 0.75. At this point, I offer you another chance to draw from one of the urns, and if you manage to draw a black ball from that urn, you get \$100. Which urn would you pick?

Clearly, $\theta_B > \theta_A$, but it is not clear that B is obviously the better choice, because the amount of evidence you have for $\theta_A = 0.5$ is higher than for $\theta_B = 0.75$. This is a problem for conceptualism, because it states that the option with a higher degree of belief would be the better one, and, in terms of just comparing the probabilities alone, picking urn B has a higher probability of winning; however, all the facts in the situation are different than what the probability lets on, so the conceptualist approach overlooks some crucial information. To belabor the implication a bit, further consider another urn C , from which you draw 2 balls, and one of them is white, so your best estimate would be $\theta_c = 0.5$. If degrees of belief are just probabilities, then it must mean that your doxastic attitude toward A and C ought to be the same, but Peirce insists that this cannot be the case:

In short, to express the proper state of our belief, not one number

but two are requisite, the first depending on the inferred probability, the second on the amount of knowledge on which that probability is based.

In a footnote, Peirce says that the weight of the evidence is in some way dependent on the “probable error” of the evidence.¹³ “Probable error” can be thought of as the precursor of the concept of the confidence interval in Frequentist statistics.¹⁴ So Peirce’s idea is that evidential weight has something to do with the accuracy of the probability, expressed in something like confidence interval.

However, it would be premature to interpret Peirce as suggesting that evidential weight *is* the spread of a distribution. While the idea that confidence interval becomes smaller as our evidence gets weightier has a strong appeal, the two can come apart. This is recognized by Keynes.¹⁵ An intuitive example would be to consider receiving two conflicting testimonies from two experts.¹⁶ Suppose one of them predicts, say, the probability of rain tomorrow is 0.01 and the other 0.99. The weight of our evidence has increased, but we are no more closer to a narrower distribution. Or suppose we find in our dataset extreme outliers. In such a situation, it is often advisable to disregard the outliers - presumably decreasing the weight of the evidence at least in Keynes’ view - for the sake of the integrity of the statistical model, especially if the chosen model

¹³Charles S. Peirce, “The Probability of Induction,” in *Writings of Charles S. Peirce: A Chronological Edition, Volume 3: 1872–1878* (Indiana University Press, 1966), 265.

¹⁴Hacking in fact suggests that there is a direct link between confidence interval and Peirce.

¹⁵Keynes, *A Treatise on Probability*, 84.

¹⁶Seamus Bradley, “Imprecise Probabilities,” in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Winter 2016 (<https://plato.stanford.edu/archives/win2016/entries/imprecise-probabilities/>; Metaphysics Research Lab, Stanford University, 2016).

is not robust against outliers.

In fact, what Peirce has in mind is subtler than simply equaling the weight of evidence as the distance between the lower and upper bounds of the confidence interval, as evidenced by his characterization of evidential weight as a measure of “liability to be changed by further experience”¹⁷ Consider this slightly modified example of his: returning to the urn with black and white balls, suppose you have randomly sampled with replacement from the urn 2000 times, and you find 1000 of them to be black, and 1000 to be white. At this point, it is reasonable to infer that the proportion of white to black balls is equal. Now, suppose 20 more balls are drawn, but they are all black. At this point, Peirce asks, should we conclude that the urn has more black balls than white? Peirce says that the 20 consecutive black balls should be regarded as an anomaly, and we should continue believing that the proportion is equal.

This example illustrates two important ideas. The first has to do with the behavior of conditional probability. The posterior probability of a black ball being drawn conditioning on the 2000 samples, will not be changed drastically by 20 additional samples, even if they are all black balls. This again can be conveniently demonstrated using the rule of succession: the posterior probability generated by updating a prior distribution of 1000 successes and 1000 failures with 20 additional successes is $\frac{1000+20}{2000+20} \approx 0.505$, so it barely made a dent in the prior probability 0.5.

But Peirce has a more general epistemological point as well, that is, we

¹⁷Peirce, “The Probability of Induction,” 296.

can make higher order judgment about the relevance of new experience based on the weight of evidence. Since the balance of the evidence is in favor of some hypothesis H if and only if $P(H) > 0.5$, it would appear that the 20 black balls should tip the balance from justifiably indifference between black and white balls, to slightly in favor of the urn having more black balls. This can be interpreted as what formal epistemologists calls an “insensitivity to mild evidential sweetening,” which describes the epistemic situation in which one begins with an agnostic attitude toward some proposition P and its negation $\neg P$, and continue doing even in light of some new evidence. Miriam Schoenfield,¹⁸ 200 argues that such an insensitivity can serve as a ground for accepting the position of *imprecise probability*, which states that degrees of beliefs cannot be represented by a single probability distribution or a single numeric value. Imprecise probability thus does away the classical Bayesian assumption that each belief can always be represented by a unique number. [citeRamsey] Much of Peirce’s criticism of the conceptualists, in fact, can be boiled down to their assumption about precision. It is also not far-fetched to suggest that the precision or sharpness of a probability assignment probability is what Peirce has in mind regarding evidential weight.¹⁹ In X, Peirce expresses his admiration of the method of deducing indeterminate probabilities by Boole, which is seen as a precursor of modern imprecise probability..²⁰

¹⁸“Chilling Out on Epistemic Rationality,” *Philosophical Studies* 158, no. 2 (2012): 197–219.

¹⁹Cite Levi

²⁰Peter Walley, *Statistical Reasoning with Imprecise Probabilities*, 1991.

1.3 Total Evidence, Reference Class, and Weight

At this point, it is helpful returning to a broader philosophical context. Reflections on Peirce's view hints at a richer conception of weight that is not defined merely as the amount of relevant evidence. Clearly, the amount of evidence does influence the weight of evidence, but it does not have to be the full picture. To begin, as Peirce points out, the whole motivation of evidential weight stems from the intuition that even on epistemological ground alone, we can make a higher order decision to "veto" the balance of the evidence.

Keynes is aware of this, but he seems to just throw up his hands, saying:

For in deciding on a course of action, it seems plausible to suppose that we ought to take account of the weight as well as the probability of different expectations. But it is different to think of any clear example of this, and I do not feel sure that the theory of 'evidential weight' has much practical significance.²¹

However, in the paragraph that immediately follows the passage above, I think Keynes has put his finger on the crux of the problem. He points out that probability theorists, such as Bernoulli, often assumes what Carnap later calls "requirement of total evidence": when making judgments about probability, we ought to consider all of our background knowledge.²², also see Rudolf Carnap,²³ p.211]. What this seems to imply, Keynes points out, is that we ought to find as much information as we can - the weightier the evidence, the better. But

²¹Keynes, *A Treatise on Probability*, 84.

²²Keynes, 84.

²³*Logical Foundations of Probability* (Chicago]University of Chicago Press, 1950).

Keynes demurs:

There clearly comes a point when it is no longer worth while to spend trouble, before acting, in the acquisition of further information, and there is no evident principle by which to determine how far we ought to carry our maxim of strengthening the weight of our argument.²⁴

What Keynes is articulating is a crucial intuition about why evidential weight is philosophically interesting: evidence is, in a sense, like money - it has a diminishing return. This, I think, provides a good rationale for thinking that the weight of evidence cannot simply a straightforward measure of the absolute amount of evidence: the increase of relevant evidence does not always entail the increase of weight, because, after we have reached a critical mass, new relevant evidence loses its ability increase the overall weight. This is why it does not appear rational to keep on looking more new evidence endlessly.

In light of this, Keynes' suggestion of how evidential weight would work in the language of Frequentism is especially telling, for, he says, "the weight is increased if we are able to employ as the class of reference a class which is contained in the original class of reference."²⁵ By class of reference, Keynes here means that the class of events we use as a reference point to determine the relative frequency of the sub-event of interest. In other words, Keynes is suggesting that as evidential weight increases, a narrower reference class would be used to determine the relevant relative frequency. But based on his criticism

²⁴Keynes, *A Treatise on Probability*, 84–85.

²⁵Keynes, 85.

of Frequentism, this cannot be right. The idea is that a narrower reference, even if it is a more accurate description of the event of question, does not always mean that we make better inference - as a matter of fact, Keynes points out that

If the process of narrowing the class were to be carried to its furthest point, we should generally be left with a class whose only member is the proposition in question, for we generally know something about it which is true of no other proposition.²⁶

In other words, there must be some point at which it becomes irrational to continue narrowing the reference. So Keynes' criticism of Frequentism can be applied to his notion of weight.

Bradley, Seamus. "Imprecise Probabilities." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2016. <https://plato.stanford.edu/archives/win2016/entries/imprecise-probabilities/>; Metaphysics Research Lab, Stanford University, 2016.

Carnap, Rudolf. *Logical Foundations of Probability*. Chicago: University of Chicago Press, 1950.

Fitelson, Branden. "A Bayesian Account of Independent Evidence with Applications." *Proceedings of the Philosophy of Science Association* 2001, no. 3 (2001): S123.

Joyce, James M. "How Probabilities Reflect Evidence." *Philosophical*

²⁶Keynes, 115.

Perspectives 19, no. 1 (2005): 153–78.

Keynes, John Maynard. *A Treatise on Probability*. Macmillan; Co., Limited., 1921.

Laplace, Pierre-Simon. *Philosophical Essay on Probabilities*. Spring, 1995.

O'Donnell, Rod. “Keynes’s Weight of Argument and Popper’s Paradox of Ideal Evidence.” *Philosophy of Science* 59, no. 1 (1992): 44–52.

Peirce, Charles S. “The Probability of Induction.” In *Writings of Charles S. Peirce: A Chronological Edition, Volume 3: 1872–1878*. Indiana University Press, 1966.

Popper, Karl. *The Logic of Scientific Discovery*. Routledge, 2002.

Schoenfield, Miriam. “Chilling Out on Epistemic Rationality.” *Philosophical Studies* 158, no. 2 (2012): 197–219.

Walley, Peter. *Statistical Reasoning with Imprecise Probabilities*, 1991.