Evidential Weight and the Expected Value of Knowledge

Lok C. Chan

1 Introduction

Keynes' discussion of the weight of evidence in his *Treatise on Probability* is perhaps the most referenced source in the topic. While he seems rather ambivalent about the subject, saying that he finds it "somewhat novel" but is "uncertain as to how much importance to attach to it," he has made many important contribution to ways in which evidential weight can be conceptualized.[1] One conception he develops formally is his idea that the weight of evidence is the *absolute amount* of the evidence we have.

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

2 Keynes' view in *A Treatise in Probability*

Let us begin, however, with a brief overview of the account provided in the *Treatise*, in which probability is defined as a probability relation between two propositions. Probability is *logical* relation, because Keynes sees it as the same relation as the entailment relation between the premises and conclusion in a deductive argument, only differing in degrees. Keynes says:

> Inasmuch as it is always assumed that we can sometimes judge directly that a conclusion follows from a premiss, it is no great extension of this assumption to suppose that we can sometimes recognise that a conclusion partially follows from, or stands in a relation of probability to, a premiss.[2]

Thus, in a derivation in deductive logic, the set of premises fully entails its conclusion. In probabilistic reasoning, the set of premises partially entail its conclusion, so in this view a probability is conceived as the degree of a partial entailment.[3] Keynes' view also provides an explanation on why probability is *normative*: it is rational to obey the rules of probability for the same reason that we should respect a deductive rule like *modus ponens*, so the degrees of a belief should correspond to degrees of entailment that the belief receives as a conclusion in an argument.

Since probability is logical relationship between propositions, it shares with the classical view of probability, i.e., Frequentism, that probability ought to be

---

[2]Keynes, 57.
[3]Keynes, 30.

objective. However, Keynes' view has an irreducible mind-dependent element in that probability can only be known to us through intuition and perception. This aspect of Keynes's thought is deeply rooted in the epistemological views of logical atomists who dominated Cambridge philosophy, such Bertrand Russell and G. E. Moore. In particular, Keynes accepts Russell's distinction between knowledge by acquaintance and by description.[4] Knowledge by acquaintance, a cornerstone of British empiricism, refers to knowledge that can only be acquired through direct perception, such as the qualitative experience of seeing yellow. So, as far as epistemology goes, Keynes in the *Treatise* is an orthodox empiricist in the British tradition. He, for instance, accepts that thesis of logical atomism that there are basic logical atoms knowable only through the senses:

> In all knowledge, therefore, there is some direct element; and logic can never be made purely mechanical. All it can do is so to arrange the reasoning that the logical relations, which have to be perceived directly, are made explicit and are of a simple kind.[5]

In other words, Keynes does not think all probability relations are knowable to us, but when they are knowable they are revealed through our intuition, so the pressing problem for him is to find a principled way through which we can determine our perception of certain probability relations are really genuine furniture of the logical space. Keynes' general strategy is to first appeal to our intuition about some basic properties of probabilities, and then to justify by delineating conditions under which these properties are genuine.

---

[4]Keynes, 11.

[5]Keynes, 14.

The identification of probability as some perceptible relation has introduced an irreducible element of subjectivity, and Keynes admits just as much.[6] Frank Ramsey sees a tension in Keynes' reliance on Russellian empiricism:

> ...there really do not seem to be any such things as the probability relations he describes. He supposes that, at any rate in certain cases, they can be perceived; but speaking for myself I feel confident that this is not true. I do not perceive them, and if I am to be persuaded that they exist it must be by argument; moreover I shrewdly suspect that others do not perceive them either, because they are able to come to so very little agreement as to which of them relates any two given propositions.[7]

In response, Ramsey sketches out a framework for a subjective interpretation of probability, in his groundbreaking paper "Truth and Probability". We will return to the Keynes-Ramsey connection in the second half of this chapter. It is worth pointing out some major differences between their views. First, Ramsey rejects Keynes' Russellian empiricism and the appeal to intuitable probability relations; instead, Ramsey draws from C.S. Peirce's pragmatism and argues for an operationalized and behavioral account of subjective probability. Second, to explicate the equality relation between probability, Ramsey abandons the use of the principle of indifference altogether. Instead, he proposes a procedure to elicit probability assignments based on the agent's betting behavior. Third, Ramsey rejects Keynes' commitment that not all probabilities are knowable. Instead, he adopts what comes to be the dominant Bayesian

---

[6]Keynes, 17.

[7]Frank P. Ramsey, *Philosophical Papers* (Cambridge: Harvard University Press, 1990), 57.

position that all probabilities can be given a precise numeric values.

3 Indifference and Irrelevance

Perhaps the most important relation is *judgments of indifference* between two propositions. The relation of indifference has its root in Laplace's "Principle of Non-Sufficient Reason":

> When the probability of a simple event is unknown, one may suppose that it is equally likely to take on any value from zero to one... the probability of each of these hypotheses, given the observed event, is a fraction whose numerator is the probability of the event under this hypothesis, and whose denominator is the sum of similar probabilities under each of the hypotheses.[8]

In other words, when we are in complete ignorance regarding the outcome of the event, the probability of each possible outcome is:

$$\frac{1}{\# \text{ of total possible hypotheses}}$$

Laplace's rule was harshly criticized on many grounds. One is the charge that it leads to absurdity, that tends to arise when the space of possible hypotheses is not clearly delineated. Peirce, for instance, points out that if we are in complete ignorance about whether some event A will occur, then principle

---

[8]Pierre-Simon Laplace, *Philosophical Essay on Probabilities* (Spring, 1995), 20.

will suggest equally distributing the probability to E and $\neg E$. However, imagine that there are inhabitants on Saturn but we have no idea what color of hair they have. Since we are in total ignorance about whether or not their hair color is red, the principle says that the probability of their hair being red is $1/2$. This would also goes for the probabilities of their hair being blue, green, yellow, etc. Assuming they can only have one hair color, this means that these are mutually exclusive events, but since they are all $1/2$, the sum of all of these probabilities would be more than 1, which violates the axioms of probability.

When Keynes wrote *A Treatise on Probability* many years later, he was keenly aware of these paradoxical results. However, he thinks that the paradoxes only suggest that the principle is to be restricted, not abandoned altogether. He argues that the reason that these paradoxes occur is, because (what he now calls) the Principle of Indifference should not be used when the alternatives under consideration can be further analyzed. Once all the alternatives are, in Keynes's words, *indivisible*, each of them should be assigned the probability of $1/n$, where n is the number of alternatives.[9] Only when the conditions for the proper application of the Principle of Indifference is satisfied, our judgment of indifference is *justified*.

In any case, more pertinent to the notion of evidential weight 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

---

[9]Keynes, *A Treatise on Probability*, 60.

6

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.[10] So, Keynes proposes that evidence E is irrelevant to the proposition H if and only if:

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

4   Keynes on Relevance and Weight

Now that the notion of relevance has been introduced, we come to Keynes' idea of the weight of evidence. We saw that Keynes was deeply concerned about whether uncertainty about the future could be captured by the precise magnitude of a probability. Keynes is especially troubled by the fact that the degree of a probability does not scale straightforwardly with the amount of the evidence we have at hand. In a well-known passage, Keynes says:

> As the relevant evidence at our disposal increases, the magnitude
> of the probability of the argument may either decrease or increase,
> according as the new knowledge strengthens the unfavourable or
> the favourable evidence; but something seems to have increased
> in either case,—we have a more substantial basis upon which to

---

[10]Keynes, 59.

rest our conclusion. I express this by saying that an accession of new evidence increases the weight of an argument. New evidence will sometimes decrease the probability of an argument, but it will always increase its 'weight.'

The crucial idea here is that Keynes is taking the weight of evidence to be closely tied to the *absolute amount* of evidence. However, Keynes' idea of relevance is technical, so let us first explicate this notion. 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.[11] For instance, we may say that when the conditional probability $P(H|E) > 0.5$, then evidence E is somewhat in favor of the hypothesis. Of course, the balance changes as we gather more relevant evidence, and it might go from favorable from 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 the *amount* of evidence involved in calculating the balance of the evidence, and Keynes calls this measure 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

---

[11]Keynes, 78.

argument, but it will always increase its 'weight.' "[12]

A critical assumption Keynes makes in the just mentioned quote is that evidential weight forms a correlative 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.[13] The implication of this is that since relevance here is defined in terms of the difference between prior and posterior probabilities, the increase in the weight of evidence must be accompanied by a change in probability. However, this also contradicts directly with Keynes' earlier observation that sometimes can change the weight of the evidence but not its balance. Keynes is aware of this tension, and tries to resolve it by introducing a "strict" definition of relevance.

5   The Strict Definition of Relevance

The problem, as Keynes sees it, 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, so what we have a situation in which some evidence, when gathered altogether, is irrelevant to the hypothesis in question, but, were they gathered separately, they would yield a posterior probability different than the prior, and so they are would individually relevant.

---

[12]Keynes, 78.
[13]Keynes, 78–79.

What Keynes has in mind seems to be this: suppose, by a justified judgment of difference, you have reasons to believe that the urn in 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.

Now further suppose that we have received evidence E, where $E = E_b \wedge E_w$. 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..[14] According to this rule, $P(H|E_b \wedge E_w)$ is

$$\frac{1+1}{2+1+1} = \frac{2}{4} = \frac{1}{2}$$

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 balls is higher than 50%. This could be verified by the rule of succession: $P(H|E_b) = \frac{2}{3} \neq P(H)$, 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 gives a similar but

---

[14]'...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.', Laplace, *Philosophical Essay on Probabilities*, 11

formal example.[15]

To avoid this type of problems, Keynes proposes a 'strict' definition of irrelevance. While Keynes calls this the 'strict' definition, the purpose of the definition is to relax the requirement for relevance, so that evidence that is irrelevant "as a whole" may still be relevant due to its components. The strict definition states:

Definition of Relevance: E is relevant to H on background knowledge K if and only if there is at least one proposition R, such that

1. R is inferable from $(E \wedge K)$.
2. R is not inferable from K.
3. $P(H|R \wedge K) \neq P(H|K)$
4. If there is *no* proposition that satisfies conditions 1-3, then E is irrelevant to H.

In other words, the strict definition tries to capture the idea that a piece of evidence E is relevant H if and only if R, in conjunction with your existing body of knowledge, can imply some proposition R that is conditionally relevant to H. So in the case where $E = E_b \wedge E_w$, E is *not* irrelevant, i.e., relevant, to H, because E_b, which is inferable from $E \wedge K$, does raise the probability H.

Unfortunately, this definition does too little and too much at the same time. It does too little, because not all weighty but irrelevant evidence can be easily decomposed in the way Keynes describes. This is the point of Popper's

---

[15]Rudolf Carnap, *Logical Foundations of Probability* (Chicago]University of Chicago Press, 1950), 419.

"paradox of ideal evidence." We shall discuss this in more detail in a later chapter, but the basic idea is that sometimes the evidence we receive is in the form of a basic statistical summary, such as "in a million tosses of the coin c, heads occurs in $500,000 \pm 20$ cases."[16] So, assuming our prior for the probability of c landing on heads is $1/2$, conditionalizing on the basic statistical summary should not change this probability - both the prior and posterior probabilities should be $1/2$.

There is something suspicious about this argument, but there does seem to be some difficulty in analyzing the statistical summary in the way Keynes recommends.[17] It might work if the statistical report is a chain of conjunctions, that is, $E_1 \wedge ... E_i ... \wedge E_{1m}$ where the subscript be the index of the trial. *If we were to interpret it this way*, we might be able to infer E_1, which, for example, says that the coin lands on heads on the first toss, and conditional on E_1 we get a different posterior probability. But this won't do, for we do not know exactly in which of the tosses the coin lands on heads and which tosses tails. In fact, modeled as a Bernoulli trial, the idea is that this particular order does not matter in drawing statistical inferences. Further, it is clear we need more than a chain of conjunctions, because we are given an interval. It is possible that with some clever maneuver we could interpret the statistical report as a chain of disjunctions of conjunctions, but at this point, we must ask ourselves if we are analyzing the evidence or if we are in fact distorting it.

There has been attempts to modify Keynes' definition of relevance slightly

---

[16]Karl Popper, *The Logic of Scientific Discovery* (Routledge, 2002), 425.

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

to accomodate Popper's paradox.[18] Roughly speaking, the proposal draws from Keynes' stipulation that if X to Y is relevant, then $\neg X$ is also relevant, and the solution would require adding another condition for irrelevance that requires both conditional probabilities of $P(Y|X)$ and $P(Y|\neg X)$.

As it turns out, however, this point is moot, because Carnap has proven that this assumption that supposedly saves Keynes' account of relevance from Popper's paradox renders relevance trivial. Hence I said Keynes' definition also does too much: Keynes' strict definition has the consequence of regarding almost any proposition as being relevant to another, as long as one is not a logical entailment of the another.

6 The Problematic Implication of the Strict Definition

Carnap's argument is that Keynes's new definition has essentially counted anything not implied by background knowledge as being 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.[19] 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 $J \vee \neg R$.

[18]O'Donnell, 50–51.

[19]Carnap, *Logical Foundations of Probability*, 420.

For a formal argument, first consider these two assumptions about relevance.

R1. If E is relevant, then $\neg E$ is also relevant. R2. If a hypothesis is not implied by our background knowledge, then there is some proposition such that it is relevant to the hypothesis.

I will not argue for these assumptions, but note that they are explicitly accepted by Keynes, so we do not have to find them plausible; we just have to accept them for the sake of this particular argument. In addition, Carnap has also proven that Keynes' definition of irrelevance is equivalent to the following:

Carnap's Definition of Irrelevance: E is irrelevant to H on K if and only if

1. $P(H|K) = P(H|K \wedge E)$ *or*,

2. $E \wedge K$ is a contradiction.

This alternative definition will be useful in the proof.

Carnap's proof is roughly as follows.[20] Suppose both evidence J and hypothesis H are contingent on background knowledge K. Also suppose a different piece of evidence R that is relevant to H on K - R must exist because of assumption R2 above. For an argument by *reductio ad absurdum*, suppose J is irrelevant to H. Now let E be $(R \wedge \neg J)$. E is either relevant or irrelevant to H, and from this we can make an argument by cases.

Suppose E is relevant: this means that E is relevant to H by virtue of Keynes' stricter definition; because, E contains R, which is relevant to H on

[20]Carnap, p.420. My presentation also partly follows Peter Gärdenfors, "On the Logic of Relevance," *Synthese* 37, no. 3 (1978): 351–67

K. By assumption R1, since E is relevant to H, $\neg E$ is also relevant to H. But this means J is relevant to H as well: $\neg E$ is equivalent to $(\neg R \vee J)$, which is inferable from J. Since an implication of J is relevant to H given K, this means J must be relevant to H.

Now suppose E is irrelevant. Since R is relevant to H on K, the only way for E to be irrelevant is for $(\neg J \wedge K \wedge R)$ to be a contradiction (by Carnap's alternative definition), so $\neg(\neg J \wedge K \wedge R)$ has to be true. That is logically equivalent to $(\neg J \wedge K) \to \neg R$. Now, by assumption R1, $\neg R$ is relevant, because R is relevant. The above formula means that from $\neg J$ plus K, we can infer $\neg R$, so this means $\neg J$ is relevant, which entails J is relevant.

So, we see that both possibilities implies that J is relevant. By *reductio*, then, J is relevant. Thus, if J is not inferable from K, then J is relevant to H.

In addition, if J is relevant to H on K, then J is not inferable from K, since one of the conditions of the strict definition of relevance is that relevant evidence cannot be a proposition that is a logical consequence of our background knowledge. So, this means that J is relevant to H on K if and only if J is not inferable from K.

In other words, any proposition not implied by our background knowledge is by definition relevant, so Carnap's result shows that the new definition is unhelpful: his suggestion is that we should simply stick with the simpler definition. Modern Bayesian have followed Carnap in accepting Keynes' basic definition of relevance as a minimal requirement for any confirmation measure. The conclusion to be drawn, I think, is that evidential weight clearly has

something to do with how much evidence we have, but it cannot be the heart of the matter.

7   Weight as Potential Net Gain in Expected Utility

In the last section, I concerned myself with Keynes' official slogan that evidential weight is the absolute amount of evidence. We saw that Keynes explicitly grounds the concept of evidential weight in his formal notion of relevance. Keynes remains ambivalent about his account of weight in the *Treatise*, and he concludes his discussion by raising "a very confusing problem" about the rationality of the demand for more evidence. Frank Ramsey responds to Keynes in an unpublished note, in which he hints at a definition of the weight of evidence E as the gain in expected value from getting E. This essentially provides a way to understand evidential weight as a sort of utility that evidence brings in. I will argue, however, that, even though the introduction of utility brings us closer to the heart of the problem, it still does not addressing the fundamental Keynes has in mind.

7.1   Keynes' puzzle about the demand for new evidence

Keynes was evidently not satisfied with his account of evidential weight in the *Treatise*. One particular vexing issue is that he could not seem to explain the

practical import that the weight of evidence seems to have. He is explicitly conflicted:

> 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. . . I do not feel sure that the theory
> of 'evidential weight' has much practical significance.

His diffidence is further amplified by a "a very confusing problem": Bernoulli, he points out, suggests that inductive inference requires the utilization of all evidence available to us. Keynes points out that this seems to imply that it's rational to get more evidence, but then it raises another critical problem about whether or not one could ever be rational in refusing new evidence[21] If the answer for the former question is positive, and the latter question negative, then we have to conclude that rationality dictates us that we should never stop looking for more evidence.

This problem is essentially repeated many years later in an exchange between Carnap and Ayer. In his *Logical Foundation of Probability*, Carnap crystalizes Bernoulli's claim as "the requirement of total evidence".

> *Requirement of total evidence*: in the application of inductive logic
> to a given knowledge situation, the total evidence available must
> be taken as basis for determining the degree of confirmation.[22]

Aver, in response to Carnap, raises the Keynesian question: should "total

---

[21]Keynes, *A Treatise on Probability*, 84–85.
[22]Carnap, *Logical Foundations of Probability*, 211.

evidence" include relevant evidence that I do not yet have in possession?[23] The answer must be "yes", Ayer argues. If finding the truth value of some proposition P could potentially sway the balance of my evidence, then I should definitely acquire it. Thus the principle of total evidence seems to suggest that I am also rationally compelled to consider some evidence I do not yet have.

But Ayer, as does Keynes, points out that this cannot be the whole picture: taken as a rule of rationality, this means we should never stop acquiring unless we are certain that we have acquired all available evidence. This, however, assumes that we know what evidence is available, but it could often be unrealistic to expect to know how much evidence we *do not* currently have.

Keynes provides a critical clue for a solution to this problem by introducing the dimension of utility to the notion of weight: he suggests that sometimes getting more evidence is clear the action that will "probably produce the greatest amount of good," especially when our evidence is slight and we know more is available.[24] Thus, for a given hypothesis of interest H, the same evidence E generates different amount of "good" relative to the amount of information we already have for H. If an agent has almost no information about H, gathering more information would generate the most utility, but for the same evidence, the demand might to low, because the agent might already have enough information about H, so getting more evidence would yield very little to no utility. What seems to be the lesson here is that the weight of evidence should be a measure of the *sufficiency* of the evidence, by informing

---

[23]A. J. Ayer, *Probability and Evidence* (Macmillan, 1972), 56.

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

us if the evidence we have at hand is *enough.*

Note, however, this question can take two forms: the first is a question about finding a threshold of sufficiency: we increase the weight of our evidence until it reaches this threshold, and the second is to *define* sufficiency *in terms* of weight: our body of evidence is sufficient, when relevant evidence no longer increases its weight. The conceptual connection between the sufficiency and weight is tighter in the second form than the first, and this is the route Ramsey takes in his attempt to answer Keynes' question.

7.2 Ramsey on the Value of Evidence

Keynes' implicit suggestion that weight can be a measure of the worth of evidence is picked up by Ramsey, perhaps the most famous critic of Keynes' theory of probability. Interestingly, however, he has proven essentially the opposite conclusion reached by Ayer and Keynes: Ramsey shows that we should always look more more evidence, because we can never be worse off from doing it. How can this be?

Ramsey's argument is roughly that, *if* we assume an agent to be a perfect Bayesian and that new information does not cost anything, then she will never be no worse off getting new evidence.[25] A perfect Bayesian agent is someone who studiously updates her opinions based on Bayes' rule and then act by choosing the action that maximize her expected utility. As Skyrms points out,

---

[25]F. P. Ramsey, "Weight or the Value of Knowledge," *British Journal for the Philosophy of Science* 41, no. 1 (1990): 1–4, also see I. J. Good, "On the Principle of Total Evidence," *British Journal for the Philosophy of Science* 17, no. 4 (1966): 319–21.

this also implies that the agent knows that she will always *stays* being perfectly Bayesian in the future.

The basic idea of the proof is that the ideal Bayesian will always get better expected utility from more evidence, until there is no more evidence, but even in that case the agent will not be worse off. An intuitive example is perhaps more helpful than reproducing the proof.[26] Suppose we have three hypotheses about the content of an urn in front of us:

1. H_b: 90 black balls and 10 white balls

2. H_w: 10 white balls and 90 black balls

3. H_n: 50 white balls and 50 black balls.

We then start by assuming $P(H_b) = P(H_w) = P(H_n) = 1/3$. Suppose we win \$1 by picking the correct hypothesis. Our expected payoff for choosing each hypothesis would be the same at $1/3$. Nevertheless, we are allow to sample with replacement as many times as we wish. Should we get more evidence? To begin, at this point, the probability of getting a black ball is the same as getting a white ball. Let E_b be "a black ball is drawn" and E_w for white balls. So:

$$P(E_b) = P(H_b)P(E_b|H_b) + P(H_w)P(E_b|H_w) + P(H_n)P(E_b|H_n)$$

$$1/3(0.9) + 1/3(0.1) + 1/3(0.5) = 0.5$$

[26]This example is adapted from Isaac Levi, "The Weight of Argument," in *Fundamental Uncertainty: Rationality and Plausible Reasoning*, ed. Silva Marzetti Dall'Aste Brandolini and Roberto Sczzieri (Palgrave MacMilan, 2011), 39–58

And $P(E_w) = 1 - P(E_b) = 0.5$. So, in the event of drawing a black ball from the urn, we would update our belief like so:

$$P(H_b|E_b) = \frac{P(H_b)P(E_b|H_b)}{P(E_b)} = \frac{1/3(0.9)}{0.5} = 0.6$$

Similarly, applying the calculation on the other hypotheses, we get:

$$P(H_w|E_b) = 0.067$$

$$P(H_n|E_b) = 0.333$$

Similar argument can be made assuming E_w, that is, a white ball is chosen. In that case $P(H_w|E_w) = 0.6$. If we were an ideal Bayesian agent, we should pick H_b if E_b, and pick H_w if E_w. Since an ideal Bayesian would choose the option that minimizes our expected utility, in either case the expected value after drawing from the urn once is 0.6, which is an improvement, since before drawing our expected utility is 1/3 for all options. The net gain in expected utility would be $0.6 - 0.33 = 0.27$, is referred to as _the value of information* in the decision theory literature.[27]

It turns out that we would be even better off if we were to draw from the urn again. Suppose the first draw yields a black ball. Let us refer to our state of belief after the first draw as $H_b', E_b', ..$ and so on. For instance, $P(H_b') = P(H_b|E_b)$ and $P(E_b') = P(E_b'|E_b)$. One notable change is that

---

[27]Howard Raiffa and Robert Schlaifer, *Applied Statistical Decision Theory* (Harvard, 1964), 89–90. For a more digestible presentation see Robert Winkler, *An Introduction to Bayesian Inference and Decision* (Probabilistic Publishing, 2010) sec.6.3.

$P(E'_b) = 0.7132$ and $P(E'_w) = 0.2868$. If we draw again and get a black ball, this means:

$$P(H'_b|E'_b) = 0.757$$

$$P(H'_w|E'_b) = 0.009$$

$$P(H'_n|E'_b) = 0.233$$

If a white ball were to be drawn:

$$P(H'_b|E'_w) = 0.21$$

$$P(H'_w|E'_w) = 0.21$$

$$P(H'_n|E'_w) = 0.58$$

Thus, if for the second sample we get a black ball, we would choose b since it has the maximum expected utility at 0.757, and if we get a white ball, we choose n with the expected value at 0.58. So, the expected utility, if we were to draw from the urn again, is: $0.7132(0.757) + 0.2867(0.58) = 0.706$, which is an improvement over just drawing once. The net gain is $0.706 - 0.6 = 0.106$. Ramsey's proof shows that we can keep on getting more evidence and we will never be worse off.

7.3 Weight as the Difference in Prior and Posterior Expected Utility

What should we make of Ramsey's proof? There are two issues involved here. The first is whether Ramsey's proof provides an account of evidential weight, and the other is whether it addresses Keynes' puzzle about gathering evidence. The two are related; let us address them in turn.

Ramsey's note is tantalizing, because he never clearly explains what he thinks the weight of evidence is, even though it's in the title. But assuming that he is interpreting Keynes' question of the weight of evidence as whether in relation to the value or worth of evidence, it is not far fetched to think that he is thinking evidential weight in terms for the expected utility that new evidence will generate. Ramsey might have something like the following in mind: the weight of new evidence E for hypothesis H is the difference between the prior expected utility $EU(H)$ and the posterior $EU(H|E)$.

For instance, we saw that in the example above, the posterior expected utility of the first draw was 0.27 higher than our prior expected utility, and we saw a net gain of 0.106 in expected utility if we were to draw again after drawing a black ball, so we see that the first piece of evidence has a higher Ramseyian weight than the second one. What Ramsey's proof demonstrates is that new evidence has a diminishing return - I get a "bigger bang for the buck" for my evidence gathering endeavor when I have less evidence.

This way of thinking about evidential weight explains how weight cannot

be measured just in terms of the amount of relevant evidence at hand: if we are almost always better off getting more evidence, we should incorporate and acquire as much new evidence as possible, but since evidence has a diminishing return, at a certain that new evidence will no longer raise our expected utility in a meaningful way (even though it will also never decrease it). This also means that at that point the new evidence will no longer have any weight, since the posterior expected utility will stay the same, even though it would have been weighty if we have no prior evidence.

I think that Ramseyan weight captures aspect of the weight of evidence, but it cannot be the whole picture. To be sure, I do not question that given some assumptions, Ramsey's result will necessarily follow: the same result is proven by both Good and Leonard Savage, so there is no doubt that the result will holds if the assumptions are granted, but that's a big *if*.

Perhaps the most crucial assumption here is that new information has no cost. It is assumed in the example that it costs us neither money nor time to draw from the urn. Suppose it costs us 25 cents for each sample. This means that we would be gaining only $0.27 - 0.25 = 0.02$ in expected payoff for the first draw, and the second draw would definitely not be worth the additional 25 cents. Or suppose that one dollar is not worthy any endeavor that lasts longer than 15 seconds, and it takes 30 seconds to draw from the urn. Cost might also enter into consideration in different forms, e.g., computational cost or memory. As such as minimally realistic assumptions are introduced, Ramsey's result no longer holds.

Of course, Ramsey probably understood that information was rarely free. It is clear that he intended his note to be an response to Keynes' remarks on weight of evidence, as the title of his note is "Weight or the Value of Knowledge," which suggests that he interprets the problem Keynes poses as a question why evidence is valuable. Good interprets Ayer's remarks in the same way. According to Good, Ayer is questioning "why... we should bother to make new observations."[28] So, both Ramsey and Good seem to think what is needed is a justification for getting new evidence *in general.* With respect to this version of the problem, the proof makes perfect, since it demonstrates that all things being equal we usually end up with better expected utility by considering more evidence. But Keynes' question was about reconciling the general duty to get more evidence and the intuition that evidence gathering for a belief of interest is not always a worthy endeavor.

Good, who proved the same result independently of Ramsey, tries to address this issue by distinguishing what he calls Type I and Type II rationality.[29] Type I rationality is that of the ideal Bayesian agent, one who lives her life by abiding to the principle of maximizing expected utility. Good recognizes, however, that type I rationality provides no guidance in regard to when an investigation should be concluded. This is where type II rationality comes in: it is principle of maximizing expected utility *plus* the consideration of "the cost of theorizing." More important, the goal of type II is "a sufficient maturity of

[28]Irving J. Good, "Subjective Probability as the Measure of a Non-Measurable Set," in *Good Thinking: The Foundations of Probability and Its Applications* (Univ Minnesota Pr, 1983), 178.

[29]Good, p. 29-30 As far as I could tell, this has nothing to do with the distinction between Type I and Type II error in Frequentist statistics.

judgments."[30]

There are two important points worth keeping in mind in anticipation of the next chapter. First, I think Good's distinction, and the admission that there are aspects of evidential practice that cannot be captured by the inductive ideal expressed by Type I rationality, is the crux of the problem. In other words, "a sufficient maturity of judgments" cannot just be a matter of maximizing our expected utility. Intuitively, maturity implies a sort of stability - a sort of imperturbability against confounding experience. We can interpret Keynes' puzzle in this light: perhaps sometimes it is rational to refuse evidence, because my belief has reached a degree of maturity such that new evidence is either just noise or not significant enough to change my mind. When there is cost involved, it makes perfect sense not to gather such evidence.

Second, how ought we to distinguish between a mature state of degree and just stubbornness? The latter is often irrational, and an agent who is simply refusing evidence out of stubbornness should be considered as irrational. This, I think, is one of the normative roles played the weight of evidence: it ought to provide a way to justify the stability of certain judgments.

---

[30]Good, 29.

26

Reference

Ayer, A. J. *Probability and Evidence.* Macmillan, 1972.

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

Gärdenfors, Peter. "On the Logic of Relevance." *Synthese* 37, no. 3 (1978): 351–67.

Good, I. J. "On the Principle of Total Evidence." *British Journal for the Philosophy of Science* 17, no. 4 (1966): 319–21.

Good, Irving J. "Subjective Probability as the Measure of a Non-Measurable Set." In *Good Thinking: The Foundations of Probability and Its Applications.* Univ Minnesota Pr, 1983.

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

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

Levi, Isaac. "The Weight of Argument." In *Fundamental Uncertainty: Rationality and Plausible Reasoning*, edited by Silva Marzetti Dall'Aste Brandolini and Roberto Sczzieri, 39–58. Palgrave MacMilan, 2011.

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

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

Raiffa, Howard, and Robert Schlaifer. *Applied Statistical Decision Theory.* Harvard, 1964.

Ramsey, F. P. "Weight or the Value of Knowledge." *British Journal for the Philosophy of Science* 41, no. 1 (1990): 1–4.

Ramsey, Frank P. *Philosophical Papers.* Cambridge: Harvard University Press, 1990.

Winkler, Robert. *An Introduction to Bayesian Inference and Decision.* Probabilistic Publishing, 2010.