Ambiguity, Value, and Resiliency

Lok C. Chan

1 Introduction

In this chapter I shall examine accounts of evidential weight by Keynes and others. I will examine four views in total: Keynes' early view of weight as the absolute amount of evidence, Ramsey's weight as the expected value of information, the resiliency account in formal epistemology, and, lastly, Keynes's refined view expressed in his *General Theory of Employment, Interest, and Money*. These views, with the exception of the last one aims to capture our intuition about the weight of evidence by explicating the notion in terms of some formal notions, such as conditional relevance or expected utility. While they can be, and often are, seen as competing accounts for evidential weight, my claim is that there is a conceptual unity in these views, and that they are quantitative measures that capture important aspects of the very notion of evidential weight.

This chapter is divided in four major sections, each devoted to an account of weight. In the first section, I shall focus on Keynes' original proposal that

evidential weight is the absolute amount of relevance evidence, and I shall show that Keynes' attempt to tie weight directly to relevance backfires. In section 2, I will try to extrapolate an account of evidential weight from a note by Ramsey, written in response to a puzzle about weight posed by Keynes. In section 3, I focus on the notion of resiliency mentioned by Jeffrey, formalized by Skyrms, and recently refined by Joyce. In the last section, I return to Keynes, by trying to reconstruct a different notion of weight by interpreting Keynes' disagreement with G. E. Moore on the force of expected value, in light of his later ideas about the role of Knightian uncertainty plays in economics.

2   Weight as Absolute Amount of Evidence

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.

2.1   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 being 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 probability, 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.

. . . 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.[6]

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 main section of this chapter.[7] –>

2.2   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

---

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

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

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
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

---

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

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 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:

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

[9]Keynes, *A Treatise on Probability*, 60.
[10]Keynes, 59.

2.3 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 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

---

[11]Keynes, 78.

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 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

---

[12]Keynes, 78.
[13]Keynes, 78–79.

9

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.

2.4   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.

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

---

[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

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 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.

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

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 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 "paradox of ideal evidence." We shall discuss this in more detail in a later section, but the basic is 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. It might work if the statistical report is a chain of conjunctions, that is, $E_1 \wedge ...E_i... \wedge E_{1mil}$ 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,

---

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

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.

As it turns out, however, this issue is moot, because there is a hidden implication of the strict definition that renders relevance trivial. Hence I said Keynes' definition also does too much: Carnap has proven that Keynes' strict definition has the consequence that almost any proposition can be made relevant to another, as long as the former is not a logical consequence of the latter,

2.5 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.[17] The argument

---

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

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, 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.[18] 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

---

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

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 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 $(\neg R \lor J)$, 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 $(\neg J \land K \land R)$ to be a contradiction (by Carnap's alternative definition), so $\neg(\neg J \land K \land R)$ has to be true. That is logically equivalent to $(\neg J \land K) \to \neg R$. 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.

3 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. However, sometimes he also speaks of the weight of evidence as a *degree of completeness of information.* This is evidently a different notion than the one tied to relevance, for as the completeness of information is a *relative* measure between how much information we currently have, and how much information there are to be gathered. In the this section, we will see that Keynes raises some puzzle about the rationality of evidence gathering, which sheds light on this different notion of weight. We will also examine a proposal given by Frank Ramsey.

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 of evidence. I will argue, however, that, even though the introduction of utility brings us closer to the heart of the problem, it overlooks the extrainductive role the weight of evidence is supposed to play.

3.1 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[19] If the answer for the former question is positive, and the latter

---

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

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.[20]

Aver, in response to Carnap, raises the Keynesian question: should "total evidence" include relevant evidence that I do not yet have in possession?[21] 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

---

[20]Carnap, *Logical Foundations of Probability*, 211.
[21]A. J. Ayer, *Probability and Evidence* (Macmillan, 1972), 56.

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.[22] 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 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.

3.2   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'

---

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

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.[23] 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, 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.[24] 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

[23]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.

[24]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

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:

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

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

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.[25]

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 $P(E'_b) = 0.7132$ 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:

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

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

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

---

[25]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.

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.

3.3 Defining Weight in Terms of 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."[26] 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 ad-

---

[26]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.

dress this issue by distinguishing what he calls Type I and Type II rationality.[27] 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 judgments."[28]

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.

4 Weight as the Resilience of Belief

In the last section, I examined the proposal that the weight of some evidence E could be understood in terms of the expected utility gained by acquiring E. We saw that, while such a notion has some attractive properties, it was unable to provide an answer for the problem Keynes poses: in light of the requirement of total evidence, how could one justified in refusing evidence? Good's idea of a rationality that aims for "a sufficient maturity of judgment" seems to be the

---

[27]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.

[28]Good, 29.

right direction, but he has not spelled out exactly what that entails.

In any case, formal epistemologists have proposed a notion of *resilience* that might fit that bill. This line of thought begins with Popper's so-called "paradox of ideal evidence." As we have seen earlier, the paradox of ideal evidence is supposed to show that evidential weight cannot be captured in terms of conditional relevance. We have already seen that Keynes' attempt to expand his definition of relevance to accommodate evidential weight has failed, since it has the unintended consequence of regarding any proposition not implied by the hypothesis, in conjunction with our background knowledge, as being relevant. As I do no think evidential weight ought to be defined merely in terms of conditional relevance, I have no intention to belabor that issue here.

However, Popper's paradox has prompted some formal epistemologists to purpose a different way to capture evidential weight, that is, in terms of the *stability.* This section aims to examine two accounts of weight as stability, one by Skyrm, and one by Joyce. In the first subsection, I shall return to the paradox of ideal evidence in greater details, in order to set the stage for discussion. I then examine in turn Skyrm and Joyce's account of stability.

4.1   The Paradox of Ideal Evidence

Popper has challenged Keynes' definition, which is adopted by Carnap and Bayesian epistemologists, by using what he calls the *paradox of ideal evidence.*[29] He asks us to consider a certain coin with an unknown bias: let N be the

---

[29]Popper, *The Logic of Scientific Discovery*, 425.

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 $P(N) = 0.5$.[30]

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. Now, given we have ideal evidence I, what is the probability of N? Popper claims that it would have to be 1/2. So

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

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

---

[30]It 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.

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.[31]

What is 'startling' about this? This returns to Keynes' initial observation: *something* is changed by the ideal evidence, but the probability fails to reflect this, so this serves as a counterexample to Keynes' claim that the increase of relevant evidence always increase its weight. The intuition is that our attitude toward N must have shifted by I somehow, so I must be relevant evidence, in a sense not captured by Keynes' definition. The weight of evidence is precisely what is missing.

4.2   The Concept of Resiliency

Skyrms credits Richard Jeffrey as the first who notices that Popper's paradox brings light to the very idea resiliency. To begin, Jeffrey points out that once we stop fixating on the probability of N, the next toss coming up head, we can see that our state of belief prior receiving the ideal evidence has a degree of malleability.[32] Consider, for instance, instead of asking only for the probability of N, we ask the probability of the next 5 tosses coming up heads. Once we think about how our belief responds to how these 5 tosses would act as

---

[31]Popper, 426.

[32]Richard Jeffrey, *Logic of Decision* (McGraw-Hill, 1965), 184.

potential evidence, given our posterior state of belief, we have very little choice but to believe that probability is $(0.5)^5$, but we would be a lot less compelled to do so with the prior state of belief. Jeffrey also makes another interesting point: since before any evidence we can be flexible with our choice of the prior, we will learn more from evidence by choosing a higher prior. This suggestion about choosing a strategic prior is especially interesting. Jeffrey has not spelled it out exactly, but we will do so in the next subsection. We will first consider Skyrms' idea about resiliency.

Skyrms tries to respond to Popper's paradox by suggesting that, even though evidential weight is not reflected by the probability, it is captured by its *stability*. The idea that there is a probabilistic representation for a *stable* state of belief can be illustrated as follows: if I have in front of me an urn U, with an unknown proportion of black and white balls. If I randomly draw 2 balls from it with replacement and find one ball for each color, my intuitive estimate of the proportion of black balls would sensibly be somewhere around 1/2. But my state of belief should be relatively *unstable*: it would be irrational for me to fixate on this estimate, especially new light of conflicting evidence. If I sample two more balls from the urn and they are both black, it would make sense for me to raise my estimate for the proportion of black balls to more than 1/2 - perhaps to 3/4. But suppose I continue to sample from for 996 more times. Out of the total 1000 draws, 500 are black. At this point a sensible would be back to around 1/2, but unlike my state of belief after only 2 samples, after 1000 samples my state of belief is stablized: suppose I sample again and I draw five black balls in a row. Now, even though drawing 5 black

30

balls in a row seems rather extraordinary, against the body of my evidence it would not warrant me to revise my belief in any significant measure.

The intuition here is that the increase in the amount of evidence, expressed here in terms of the number of samples, corresponds to the increase of stability of the estimate. Skyrm has introduced a notion called the *resiliency* to capture this intuition sense of stability. Roughly speaking, resiliency is a function of the variability of one's degrees of belief in light of new evidence. The basic idea is similar to statistical robustness: a state of belief is resilient if it tends not to vacillate in light of new evidence.

This is an attractive way to capture to notion of evidential weight. Keynes' puzzlement about how relevance and weight could come apart is addressed; because, resilience seems to be negatively correlated with relevance. When a belief B is resilient, the conditional probability on some new evidence E should be approximately the same, that is, $P(B|E) \approx P(B)$, even if E would be highly relevant were B not resilient. If, for instance, a resilient belief is one where the weight of the evidence for it is high, then it is a logical consequence that evidence could make a belief more weighty without changing its degree, for the weight is in fact stablizing this particular value.

As a way to gauge resiliency, Skyrms suggests that we can reflect on what would happen to the state of our belief if some new evidence were to emerge. The resiliency of current belief would manifest itself as "a reluctance to change."[33] Regarding Popper's paradox, Skyrm argues that what the ideal

_____

[33]Brain Skyrms, "Resiliency, Propensities, and Causal Necessitye," *Journal of Philosophy* 74, no. 11 (1966): 707.

evidence changes is not the probability of heads, but the resilience of its probability at 0.5.

Skyrm has not spelled this out, but it is easy to demonstrate the difference in resiliency using basic statistical tools. For comparison, assume in the "some evidence" scenario we have tossed the coin 40 times in the past and exactly half of the landed of heads. For the "ideal evidence" scenario, it is the same exactly the coin was tossed one million times. We would expect that the weight of evidence would be higher for the ideal evidence scenario and the belief that the coin is fair ought to be fair resilient. This is clearly seen from figure 1.
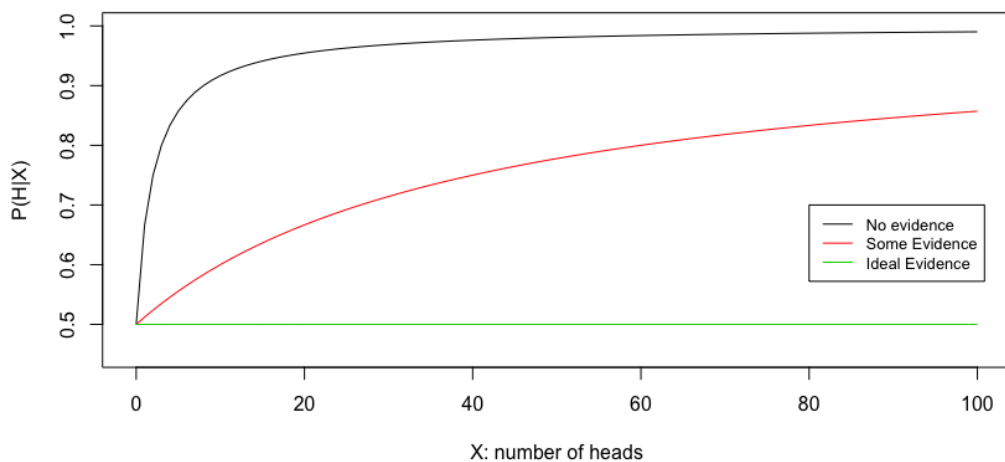

Figure 1: Comparison of Belief Resilience

We can see that if we already had ideal evidence for believing that the coin is fair, this belief would be quite resistant to new evidence. Even if we were to flip 100 consecutive heads, the probability would barely raise. In the case of no prior evidence, the situation is quite different: starting with no

evidence, would jump to after 5 flips, and rapidly approach certainty after 20 flips. Even with just some evidence, the belief would already exhibits some degree of resilience - it climbs steadily as the number of heads increases, in a rate much slower than the one with no evidence at all. So here we seem to have a response to Popper's paradox: even though that the expected value for these initial distributions are all at 0.5, they *respond* to new evidence differently.

The plot also provides an illustration of Jeffrey's point about learning. Instead of treating these data scenarios as the posteriors, we can instead suppose them to different choices of priors. Imagine that I approach the problem with an ideal-evidence like state of belief - I will learn next to nothing even if I see 100 heads in a row. On the other hand, with a no-evidence state of mind, my opinion will be swayed rather quickly.

Of course, there is nothing intrinsically irrational about each of these assumptions. For certain beliefs, there are reasons for them to be entrenched, but the problem is that it is unclear how this could be determined just by looking at resiliency. On what ground could we adjudicate a sufficiently mature judgment and a groundless prejudice?

4.3 Chance, Resilience, and the Probability of Probability

Indeed, while resiliency seems to be a workable as an intuitive answer for Popper's paradox, many important questions still need to be answered: what

exactly is being stabilized? Stabilized by what? More important, how does the resilience of a belief affect one's decision? Recall that Keynes' difficulty with evidential weight is that it seems like it could override the mandate of the expected value. It is not at all clear how resiliency is supposed help with that: when push comes to shove, all three priors have the same expected value, and therefore would recommend the same course of action.

Fleshing out the mechanical details of the paradox of ideal evidence will help answering these questions. This will also introduce the Lewisian chance/credence distinction, which will be useful when we discuss Joyce's suggestion.

4.3.1 Ideal Evidence Redux

Recall Popper begins by asking for our subjective probability for the coin landing on heads on the next toss, and, assuming that we give it a 50% chance, his argument is that this particular probability will be left unchanged even receiving the statistical summary that the coin were tossed a million times and half of them were heads.

Nevertheless, Popper's argument contains a sleight of hand that shifts between two ways of thinking about the probability of the coin toss - one is about our opinion about whether the next toss will be heads, and another one is about our opinion that the coin is *fair*. Once this is pointed out, the air of paradox dissipates. First, our subjective probability about the result of the next coin toss does not come from a vacuum; instead, it is informed by

our prior judgment about whether the coin is fair. That is, the only reason we say that the probability of the next coin toss being heads is because we are operating under the assumption that the coin is fair - it is a conditional probability. Second, our confidence is changed by the ideal evidence, but once the distinction is pointed out, it is clear that we become more confident not in our prediction that the next toss will be heads, but that the coin is fair. Imagine I offer you two bets:

1. Get $10 if the next toss is heads.
2. Get $10 if out of the next 100,000 tosses, half land on heads, with a reasonable degree of variability allowed (say, within 1 standard deviation).

After receiving the ideal evidence, bet 2 should look pretty appealing, as our state of belief is changed by the statistical report not with respect to whether the toss will be come up heads, but my confidence in the coin's fairness. Once this is pointed out, there is nothing strange about saying that the ideal evidence will not change the probability about the result of the next toss. This is the misdirection that Jeffrey pointed out.

It will be useful for our discussion to state these somewhat formally. We will use X_i to represent the result of the ith toss, and let

$$
X_i = \begin{cases} 1 & \text{the coin lands on heads on toss } i, \\ 0 & \text{otherwise(i.e., tails).} \end{cases}
$$

So $X_5 = 1$ would say "the coin lands on heads on toss 5." For the sake of

brevity, I will use X_{n+1} to refer the result of the next toss.

So, reformulating the line of thought above, we can summarize the ideal evidence as $X_{1:n} = 500,000$, where $n = 1,000,000$. In other words, within n tosses, 500,000 were heads, assuming that each toss was exchangeable, i.e., the order does not matter. So, we can understand the paradox of ideal evidence as

$$P(X_{n+1}) = P(X_{n+1}|X_{1:n}) = 1/2$$

Which, as I just suggested, is not the complete picture, because $P(X_{n+1})$ should be conditional on our belief about the bias of the coin. We will use θ to represent the bias of the coin, with $\theta = 1$ if the coin is guaranteed to land on heads, and 0 if guaranteed to land on tails. We can use a statistical model called the beta distribution to express our uncertainty about θ, so we can speak of $P(\theta = x)$, i.e., the probability of the coin's bias being x. So, my claim is that while it is true that $P(X_{n+1})$ is not changed by the ideal evidence $X_{1:n}$, the following is also true:

$$P(0.49 < \theta < 0.51) \neq P(0.49 < \theta < 0.51|X_{1:n})$$

We can use the cumulative distribution function of the beta distribution to find out the following probabilities:[34]

---

[34]This assumes the beta-bernoulli distribution as the model. No evidence: $\alpha = \beta = 1$. Ideal evidence:$\alpha = \beta = 500,000$

	$P(0.46 < \theta < 0.54)$	$P(0.49 < \theta < 0.51)$
No evidence	0.08	0.02
Ideal evidence	1	1

In other words, before getting the ideal evidence, the probability for the coin being fair is roughly 0.02, but after taking into account the ideal evidence, we are practically certain that it is fair, so there is no paradox, as long as we are being clear about what the ideal evidence is for. The ideal evidence is irrelevant to our opinion about the next toss *unless* we can using it to inform our opinion about the bias of the coin. Just as Jeffrey suggests, even though the probability of the next toss coming up heads on the *next* toss is 0.5 for both scenarios, one would be much more confidence in the fairness of the coin after receiving the ideal evidence.

4.3.2 Chance and Probability

I have been freely using the phrase "the probability of the coin being fair" without definition, but it is potentially problematic if the meaning is not clarified. This appears to suggest that weight could be identified as a "higher order probability," i.e., the evidence has made $P(\theta = 0.5)$ - "the probability of the coin landing on heads is 1/2" - more probable. Unfortunately, this is not so simple.

De Morgan points out that sometimes it makes sense to speak of the

probability of a probability.[35] He asks us to imagine 100 urns, of which only one urn - call it M - has an equal proportion of black and white balls and others have various other proportions with the same two colors. If I were to randomly draw a ball from M, the probability of B, that a black ball is drawn, would be 0.5. But suppose the urns are indistinguishable from each other, and I can only randomly pick one without knowing if it is really M. We are uncertain if the probability of getting a black ball is 0.5, but we do know enough to quantify this uncertainty: since I am randomly picking one urn out of 100, the probability of *the probability of B is 0.5* is $1/100$.

L. J. Savage expresses a similar sentiment:

> . . . there seem to be some probability relations about which we feel relatively "sure" as compared with others. When our opinions, as reflected in real or envisaged action, are inconsistent, we sacrifice the unsure opinions to the sure ones.[36]

Nevertheless, Savage urges caution on introducing higher orders of probability, due to the technical challenges:

> The notion of "sure" and "unsure" introduced here is vague, and my complaint is precisely that neither the theory of personal probability, as it is developed in this book, nor any other device known to me renders the notion less vague. There is some temptation to introduce probabilities of a second order so that the person would find himself

---

[35]Augustus de Morgan, *An Essay on Probabilities* (Longman, Orme, Brown, Green, Longmans, 1838), 87.

[36]Leonard J. Savage, *The Foundations of Statistics* (Dover, 1954), 57–58.

saying such things as "the probability that B is more probable than

C is greater than the probability that F is more probable than G."

But such a program seems to meet insurmountable difficulties.[37]

Theoretical issues regarding higher order probabilities are beyond the scope of our discussion here. While Savage sounds pessimistic, higher order probability does have its supporters, such as Good and Skyrms. However, our focus here is evidential weight, and getting involved in this technical debate is unlikely to be helpful. Keynes also entertained this idea of weight as higher order probability, and quickly concluded that they cannot be the same thing, even though it appears to have more to do with his commitment that probability is a logical relation.[38]

In any case, I do not wish to saddle the weight of evidence with the commitment in the existence of higher order probabilities. Still, we must clarify what is exactly meant by "$\theta = 0.5$", because it *seems* to say that the probability of the coin landing heads is 0.5. Regarding this, I shall defer to the views of D. V. Lindley and David Lewis. Lindley argues that probability is a relation between the agent and the world, so when we say something about $P(\theta = 0.5)$, θ must be something about the world.[39] In our case, this has to be an objective feature of the coin, hence I have been careful in describing θ as the bias of the coin, which is a property in the world.ˆ

Of course, this does not fully answer the question: what is this objective

---

[37]Savage, 58.

[38]Keynes, *A Treatise on Probability*, ch. VI sec. 5.

[39]D. V. Lindley and L. D. Phillips, "Inference for a Bernoulli Process (a Bayesian View)," *The American Statistician* 30, no. 3 (1976): 115.

feature? Lindley's answer is that it is the *propensity* of the coin to land heads. Skyrms recommends a similar interpretation[40] This recommendation is compatible with, if not the same as, the influential view presented by David Lewis, who adopts Carnap's pluralistic stance on probability. Carnap thinks there are at least two concepts of probability: *probability*$_1$, which is an epistemic concept about degrees of confirmation and *probability*$_2$, which refers the empirical concept of long-run limiting frequencies.[41] Lewis suggests that we should instead interpret the epistemic concept as credence or degree of belief and the empirical concept as chance or propensity.[42] So, following Lewis, we can interpret $P(\theta = 0.5) = x$ to be "the degree for the belief that the chance of heads is 0.5 is x." For the sake of consistency, I will refer to subjective probability just as *probability*, and objective probability as *chance*. This also has the benefit of not deviating from Bayesian statistics: even though θ looks like a probability, in the Bayesian statistical framework it is just another parameter being modeled, not unlike μ or σ for normal distributions, so the bias of the coin is a objective feature of the world in a way no different than the fact that the average age of Duke students is an objective fact. Our degree of belief about them, however, is subjective.

---

[40]Skyrms, "Resiliency, Propensities, and Causal Necessitye," 707.

[41]Carnap, *Logical Foundations of Probability*, 517.

[42]David Lewis, "A Subjectivist's Guide to Objective Chance," in *Philosophy of Probability: Contemporary Readings*, ed. Richard C. Jeffrey (University of California Press, 1980), 83–132.

4.3.3 Expected Loss and Resiliency

With the distinction between chance and probability in mind, let us return to the question of what exactly resiliency is, and how evidential weight figures in this. It seems that resiliency is what we have when a probability is stable, but it's not yet clear what that means.

To begin, we have to clarify *which* probability is being stabilized by *which* evidence. For the ideal evidence case, the stabilizing agent is the statistical report given to us - this much is clear. Now, Jeffrey's suggestion seems to be that what's being stabilized is $P(X_{n+1} = 1)$, that is, the probability that the next toss coming up heads.

let us return to Skyrms' proposal that weight is, or at least manifest itself in, the resilience of the probability

Joyce agrees with Skyrms that evidence expresses its weight in terms of the stability of the belief it supports, but he disagrees that it is the probability that is being stablized. His suggestion is that the basic quantiity being stablized

It is helpful to motivate the idea of expected loss intuitively. Let us recycle an old example from an earlier section: we are trying to guess the proportion of black balls, θ in an urn, and we know that either θ is either 0.1, 0.5, or 0.9. Let X_i be the result of the ith draw, and $X_i = 1$ if the ball is black and 0 otherwise. Recall that, we started by assigning the same subjective probability to the three hypothesis, and based on these assumption our initial subjective probability for drawing a ball $P(X = 1) = 0.5$. We then drew one sample from

the urn, and found one black ball, and then we update our beliefs as follows:

$$P(\theta = 0.9 | X_1) = 0.6$$

$$P(\theta = 0.5 | X_1) = 0.333$$

$$P(\theta = 0.1 | X_1) = 0.067$$

$$P(X_2 = 1 | X_1 = 1) = 0.7132$$

In section 2, we stipulated that we will get \$1 for guessing the correct hypothesis and nothing otherwise. Now, instead, suppose the rules are different: we do not win anything but we lose nothing if we guess correctly, and if we make a wrong guess, our loss is proportion to how far our guess is from the actual state of affair. For instance, given a, we should lose more if the proportion turns out to be 0.1 than 0.5. There are different *loss function* with which this loss could be calculated; we adopt Joyce's choice of the so-called the squared loss function: $(\theta - a)^2$, so if it turns out that $\theta = 0.5$, then our loss would be $(0.5 - 0.9)^2 = 0.16$. Intuitively, we can think of the squared loss as a way to formalize "how wrong we are" in terms the distance between our choice and the actual state of affair. The idea of the expected loss is to consider the probability of each potential loss. We know that there are only three possible proportions, so

1. There is a probability of 0.6 that $\theta = 0.9$, in which case we lose nothing.
2. There is a probability of 0.333 that $\theta = 0.5$, in which case we lose 0.16.

3. There is a probability of 0.067 that $\theta = 0.1$, in which case we lose 0.64.

We can see that our decision was a good one, since we are most probably not going to lose anything. It also makes sense that we would lose the most if it turns out that $\theta = 0.1$. The individual expected loss for each possible state of affair for a is respectively $0, 0.053, .043$, and the *Bayesian expected loss* is the sum of these values, which is .096. Following the statistical decision theory literature, we will use $L(\theta, a)$ to refer to loss inccured by action a, given the state of the world is θ, and $E(L(\theta, a))$ to refer to the overall expected loss.

Joyce's claim is that the "basic" resilient quantity is $L(\theta = x, a)P(\theta = x)$, which is what was referred to as "individual expected loss" above. In words, we can say that it is the loss we incur, if we were to decide that θ is a but θ is in fact x, discounted by our subjective probability that θ is actually x. *This*, instead of

Joyce's proposed measure of weight, w, is the difference between the prior and the posterior expected loss - the smaller w is, the weightier the evidence.

It is worth probing our intuition behind this account. We would agree that, in general, we like scenarios in which we incur great loss to be highly improbable, and scenarios in which we incur no loss to be the most probable. This can be accomplished by carrying out a decision analysis, and we would choose the option that minimizes our loss. For instance, in the urn example, after drawing one black ball, we find out that the expected loss would be 0.096 had we chosen $\theta = 0.9$, 0.11 had we chosen $\theta = 0.5$, and 0.44 had we chosen $\theta = 0.1$. This result should resonate with our intuition - if the only evidence

we have is the draw of a black ball, then it stands to reason that the best hypothesis to pick would be one that has the highest chance of yielding a black ball. However, this is clearly not enough: since we do not have much evidence, the evidential weight for belief is low. There are two clues for this: the first is that the expected loss for choosing $\theta = 0.9$ is not that much lower than choosing $\theta = 0.5$. Another clue is the marginal probability for drawing a black ball. After taking our hypotheses and the evidence into account, we saw that the probability of drawing a blackball again is 0.7132, and this is almost at an equal distance between 0.9 and 0.5, erring ever so slightly to 0.9, so we see that while our best guess is $\theta = 0.9$, it is overwhemling favored over $\theta = 0.5$.

The reason that weight does not merely stablize probabilities, Joyce suggests, is that sometimes an agent might assign makes the same probability assignment after seeing some evidence E, even though

With the technical details out of the way, it is still useful to

5 Weight as a Counter to Immeasurable Uncertainty

Expected Value and Moore's Argument in the *Principia Ethica*

In section 1, we saw that Keynes' epistemology was squarely in the Moore-Russell tradition. However, while Keynes' explicit appeal to our logical intuition is clearly influenced by Moore intuitionism about ethics, he finds the

role of probability played in Moore's ethics problematic. On one hand, Keynes was alarmed Moore's anti-utiliarian argument that probability cannot be the very guide of life, because facts about events in the remote future are often underdetermined by the information we currently have. On another hand, he was skeptical of what Daniel Ellsberg later calls "the Bernoulli proposition," which asserts that for a reasonable decision maker can infer what the right course of action is by choosing the option that maximizes her expected utility.[43]

According to Keynes, his interest in probability was prompted by a problematic application of probability in Moore's *Principia Ethica.* In the chapter "Ethics in Relation to Conduct," Moore argues against the consequentialist/utilitarian view that the right action is one that has the highest probability of producing the best outcome.[44] Moore's argument is that when we analyze the the utility generated by an action, we often only take into account the probability of the *immediate* utility that would follow from the action, but to take utilitarianism seriously, Moore claims, we must consider the effects the action in the long the run. We have to know, for instance, whatever good our action produces in the immediate future will not be negated by the negative effects it has in the long run. However, Moore argues that, because so much about the future is currently unknown to us, we simply lack the ground to make good inferences about what effects each single action will probably case in the long run.

Keynes thinks accepting Moore's argument amounts to the abandonment

---

[43]Daniel Ellsberg, *Risk, Ambiguity, and Decision* (Garland Publishing, 2001), 49–50.
[44]???

of the very idea that probability can guide our actions.[45] In particular, Keynes thinks that Moore's argument is based on the mistaken assumption that we need to have certain information about the future in order to say anything about the probable effects of our current actions, for the very idea of probabilistic reasoning is predicated on the assumption that we can act rationally under uncertainty. He points out that Moore is operating under an empirical conception of probability: to *know* the probability of an event, we must have a direct acquaintance of it by observation. But if probability is strictly defined as something that expresses itself only in long term behavior, we cannot speak of the probability of an event with any credibility unless we can study the event in a controlled and repeatable environment. This is clearly not something possible with most of our actions, since they often only occur once. In other words, probability cannot be the very guide in life if there is no such a thing as a single-case probability.

In contrast, Keynes' conception of probability as a *logical* relation, instead of an empirical frequency relaxes the requirement needed for the probability of a single case to be at least conceptually intelligible. The price of this is the principle of indifference: the reason Keynes, even though fully recognizes its flaws, finds it necessary to adopt the Principle of Indifference is that it makes it possible to make a rational decision between two actions in comparing their expected utilities:

> If good is additive, if we have reason to think that of two actions
> one produces more good than the other in the near future, and if
> we have no means of discriminating between their results in the

---

[45]Keynes, *A Treatise on Probability*, Ch. XXIV, sec.4.

distant future, then by what seems a legitimate application of the Principle of Indifference we may suppose that there is a probability in favour of the former action.[46]

So, while Keynes appreciates the force of Moore's argument that we in general have no clue about the probable effect of our action in the remote future, he thinks that we can think analyze our rationality of our action probabilistically by appealing to the principle.

However, it is clear to Keynes that the Principle of Indifference cannot fully address the necessarily incomplete nature of our knowledge of future events and how it should affect the credibility of our probability statement. Keynes' own analysis of the Principle of Indifference reveals that unless it is applied under the right conditions, our judgements of indifference can often lead to contradictions. Keynes', as we shall see, tries to find a middle way between Moore and the expected utility theorists by appealing to evidential weight and its relation to so-called "Knightian uncertainty" or "ambiguity."

5.1   Ambiguity and the Weight of Evidence

The concept of evidential weight provides the framework Keynes needed to address Moore's concerns without falling back to the typical Bernoullian position. Keynes' response to Moore is that the severe uncertainty about the future does not prevent us from speaking of their probabilities - in Keynes word, they are "genuine" probabilities even if our knowledge about the remote

---

[46]Keynes, 354.

future is slight. However, what distinguishes the probability of an immediate or near future event from one of event in the remote future is the degree of completeness of information involved. In other words, probability about the near future tends to be weightier evidentially than events in the remote future.

Many of the issues below hinge on different ways in which our ignorance about a state of affair could be characterized, so in order to avoid some confusion, I will define some terms here. I shall use "uncertainty" to refer to the epistemic attitude toward any event that involves some degree of ignorance and/or randomness. In other words, it is used as the broadest umbrella term that covers a cluster of similar epistemic attitudes; so, for instance, ambiguity and risk (as defined below) are different kinds of uncertainty. My propose usage of "risk" will follow Keynes' - the "risk" of an event is the definite and precise quantity, i.e., a probability, that captures our degree of belief toward it. I shall use "ambiguity" to describe an event to which we cannot assign a *precise* probability, but we can say *something* about it, either in terms of an interval or by comparison. Keynes also think that there is also a kind of uncertainty about which we cannot say anything *at all*. This is what Frank Knight calls "uncertainty" in *Risk, Uncertainty, and Profit*, in which he defines "uncertainty" as being immeasurable uncertainty. For our purpose, we can call this "Knightian uncertainty". The choices here are purely pragmatic.

He is resistant to utilitarians' claim that the rational course of action can be wholly determined by calculating the precise "mathematical expectation", that is, what we called *expected utility*.[47]

---

[47]Keynes, ch. XXVI sec. 5-6.

This is where the concept of evidential weight is important, and where he disagrees with the expected utility theorists. To begin, the price of

It plays an important role in his ideas about economics and probability; thus, One major reason for his skepticism is that expected utility does not reflect the amount of evidence involved in its calculation, which, as we shall see, piques Keynes' interest in the very notion of the weight of evidence. The other major reason is that Keynes holds that some probability relations cannot be determined precisely, and sometimes we can only give it an interval estimate, and sometimes it is completely indeterminate.

For the sake of clarity, we must first have a brief discussion on terminology.

As we shall see, in *The General Theory*, Keynes has adopted the distinction between uncertainty and risk to capture his thoughts about the weight of evidence. While Keynes was clearly aware of this distinction in *A Treatise*, this terminology seems to be more closely associated with The distin

For the sake of clarity, I shall use the term "ambiguity" to refer to

In contrast to Knight, I try to avoid making out the distinction to be one about measurable vs unmeasurable, since I find that distinction itself unclear. For instance, Knight suggests that whether Britain would go to war unmeasurable, but I prefer to conceive the situation as being measurable, just with an extremely high degree of imprecision, since we can still say that the probability of Britain going to war is somewhere between 0 and 1.

These definitions are chosen simply because I find they they most easily

reconciled with the relevant literature and what I find to be the most natural.

general theory

gamble device

secondary lit

refernce frame

Ellsberg

Reference

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

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

Ellsberg, Daniel. *Risk, Ambiguity, and Decision.* Garland Publishing, 2001.

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.

Jeffrey, Richard. *Logic of Decision.* McGraw-Hill, 1965.

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.

Lewis, David. "A Subjectivist's Guide to Objective Chance." In *Philosophy of Probability: Contemporary Readings*, edited by Richard C. Jeffrey, 83–132. University of California Press, 1980.

Lindley, D. V., and L. D. Phillips. "Inference for a Bernoulli Process (a Bayesian View)." *The American Statistician* 30, no. 3 (1976): 112–19.

Morgan, Augustus de. *An Essay on Probabilities.* Longman, Orme, Brown, Green, Longmans, 1838.

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.

Savage, Leonard J. *The Foundations of Statistics.* Dover, 1954.

Skyrms, Brain. "Resiliency, Propensities, and Causal Necessitye." *Journal of Philosophy* 74, no. 11 (1966): 704–13.

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