INTRODUCTION

We may be interested in whether it will rain tomorrow. We may want to say that it would carry more weight if the prediction comes from a meteorologist than from a philosopher. In this sense, weight is essentially treated as a synonym for probability: the probability that it will rain tomorrow conditional on the fact that a meteorologist says it will is higher than being conditional on the fact that a philosopher says the same thing. This use of the term, while there is nothing wrong with it in and of itself, is not the sense with which we concern ourselves in this chapter.

Instead, imagine that you and I both know that meteorologist A says that the probability of the event R, that it will rain tomorrow, is 0.7. But unbeknownst to you, I also know that another meteorologist B independently says that the the probability of R is 0.7. Assuming neither of us have prior opinions about the chance of rain tomorrow and we both are willing the trust the experts, my additional knowledge does not mean my probability should be different than yours - as a matter of fact, we should both think that the probability of R based on our respectively information, should be 0.7. Yet it makes sense to say that my evidence has more weight than yours. It may be said that even though we would both assign the probability of R to be 0.7, my assignment is more certain, confident, trustworthy, reliable, etc. This is the sense of weight that concerns us - it is a measure that pertains to the *amount* of data available.

The following is a commonly cited quote from Keynes about this matter:

> 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.'

WEIGHT AND THE BALANCE OF REASONS

Peirce gave John Venn's *The Logic of Chance* a glowing review: 'a book which should be read by every thinking man.''(@vennlogicreview, 98) Venn earned Peirce's praise by advancing an objective interpretation of probability, which analyzes the concept as a relative frequency, against the conceptualist who holds that probabilities are degrees of credences. The conceptualists take their cues from Laplace, who in turns draws from Thomas Bayes himself, so conceptualism can be described the precursor of modern Bayesianism, both in epistemology and statistics.

The 1867 review marks the beginning of Peirce's public opposition to conceptualism He eventually gives his thoughts on conceptualism a full articulation about ten years later in 'Probability of Induction.' It is in this paper that the very idea of the weight of the evidence is first explicitly discussed in the context of probability. It is therefore appropriate to set the stage by giving some historical and conceptual background on this dispute between the early proponents of the two interpretations of probability - one that is still very much with us today.

THE RULE OF SUCCESSION

In a letter to Peter Carus, Peirce explains that one of the main targets he had in mind when attacking conceptualism was Laplace's view in *A Philosophical Essay in Probability*, which still has 'a great and deplorable influence.'(@cp, 8.220) Even though Peirce does not mention Laplace directly in 'Probability of Induction,' that he has Laplace in mind is not surprising, since Peirce spends a great deal to attack ideas originated from Laplace.

In the beginning chapter of *A Philosophical Essay on Probabilities*, Laplace proposes the so-called "rule of succession." This rule is adopted by the conceptualists as a model for belief revision based on new evidence. Laplace formulates this as follows:

> Thus we find that an event having occurred successively any number of times, the probability that it will happen again the next time is equal to this number increased by unity divided by the same number increased by two units.

In other words, for a sequence of independent trials, Laplace recommends that we ought to update our belief by using the following principle. Suppose H is the event of interest. After witnessing evidence E, we should update the prior probability $P(H)$ by calculating

$$P(H|E) = \frac{1+x}{2+n}$$

Where x is the number of successes observed out of n occurrences.

Laplace, using an example that "has had a catastrophic effect on Laplace's reputation," illustrates this by calculate the probability of the sun rising.(@jaynes, 564)He assumes that the sun has never ceased rising in the past and that the earth is about 5000 years, or 1,826,213 days old, so this means $x = n = 1,826,213$, so the probability of the sun rising is:

$$\frac{1+1{,}826{,}213}{2+1{,}826{,}213} = \frac{1{,}826{,}214}{1{,}826{,}215}$$

Even someone who is sympathetic to subjective interpretations of probability, this example seems off because this appears to say that if a bet against the sun rising at odds of $1826214:1$ were to be offered to someone, she would be irrational not to take it.

THE PRINCIPLE OF INDIFFERENCE

Nevertheless, Laplace seems to suggest that for most people, the rule of succession is not the right rule to use the estimate their posterior probability of the sun rising. He says that anyone "recognizing the totality of phenomena the principal regular of days and seasons, sees that nothing at the present moment can arrest the course of it" (laplace, 19). Presumably, this means that anyone who knows anything about planetary motion would not be rational to use the rule of succession to ascertain her personal probability for the sun rising tomorrow. Laplace is perhaps suggesting that only in complete ignorance about the event in question is one entitled to use the rule of succession. This places a restriction in the applicability of the rule.

Perhaps this is why the rule of succession does not receive any direct criticism from Peirce; instead, he focuses on another rule presupposed by it, that is, the principle of indifference, which is the idea that we should assign the probability of 0.5 to any totally unknown event. This can be seen by appealing the rule of succession in a case where no observation has been made, which means

$$\frac{1+0}{2+0} = \frac{1}{2}$$

This means that prior to any observation, the probability of any event of which we are know nothing should be 0.5, so the rule of succession presupposes the principle of indifference.

This principle is harshly criticized by Peirce, who thinks the principle of indifference has a host of issues. To begin, Peirce argues that assigning the probability of ${}^{1}/_{2}$ to unknown events can lead to paradoxical results. Imagine, he asks, that there are inhabitants on Saturn but we would like to the probability of a typical inhabitant's hair being red . Since we are in total ignorance about their physiology, the principle says that this probability should be ${}^{1}/_{2}$, and according to Laplace we can update this probability by adding the number of inhabitants we observed to the denominator and the number of red-haired inhabitants to the numerator. However, we run into a paradox when we consider other hair colors: since we are ignorant about, says, whether their hair is blue, its probability should also be ${}^{1}/_{2}$. Assuming they can only have one hair color, this means that these are mutually exclusive events, the sum of all of these probabilities would be more than 1, which contracts with the axioms of probability, creating a paradox.

THE BALANCE OF REASONS

This brings us to another important intuition about evidence. If we have a sufficient amount of evidence, we should be able to judge whether our overall evidence points in favor of or against the hypothesis in question. Both Keynes and Peirce make use of the metaphor of the balance of reasons. The idea is that we should be able to put all evidence on a metaphorical scale and see which side comes out on top: if our overall evidence is in general favorable to the hypothesis, then the balance of reasons should point to the same direction.

Peirce explicates the balance of the evidence as the *log-odds* of a belief. Take some arbitrary belief A, for instance. Suppose we start with $P(A) =^1 /_2$. The *prior odds* would be

$$\frac{P(A)}{1-P(A)} = \frac{{}^1/_2}{{}^1/_2} = 1:1$$

Suppose E is new evidence for A. Now, given a prior odd of 1:1 for A, the posterior odd for A, $\frac{P(A|E)}{P(\neg A|E)}$ can be algebraically reformulated as the likelihood ratio: $\frac{P(E|A)}{P(E|\neg A)}$. After taking the log of the odds as Peirce suggested, we can find the balance of reasons by calculating $log(P(E|A)) - log(P(E|\neg A))$. A positive value means the evidence is in favor of A and negative against it.

The log-odds also has a nice way of representing the intuition that independent reasons for a belief should be additive. The idea is that the strengths of two independent pieces of evidence should be combinable to increase the degree of the belief in question. This is also accounted for by Peirce's idea of log-odds. Suppose E_1 and E_2 are independent evidence for A. This can be captured by the idea of *conditional independence*:

$$P(E_1 \wedge E_2|A) = P(E_1|A)P(E_2|A)$$

Now, if we take the logarithm of the product, the product becomes $log(P(E_1|A)) + log(P(E_2|A))$. This captures the intuition that adding two pieces of independent evidence together should increase the intensity of our belief.

WEIGHT AND BALANCE

Peirce was reportedly the first to points out that while degrees of belief can capture the intuitions about evidential balance and quite successfully, it cannot directly capture evidential weight, which refers to a measure of the amount of relevant data and tends to be independent from the probability itself.

Peirce, who is recognized as the first thinker who makes explicit reference to the notion of evidential weight, mobilizes the notion in order to attack conceptualists' use of the principle of indifference and the rule of succession. His argument is that the combination of precisionism and Laplacean updating entails that conceptualism has no room for a notion of evidential weight. The core idea is that by using Laplace's rules of reasoning, conceptualists are always required to assign precise numerical probability assignment to a proposition or event, regardless of the amount of relevant evidence available to the reasoner. The result is the counter-intuitive scenario where two beliefs can have identical degrees of belief, even though the amounts of evidence involved are vastly different.

Peirce illustrates this point by using the following example: consider an urn with 100 balls. Each ball is either white or black. A ball will be selected randomly and your job is to guess its color. Before the ball is drawn, however, you are allowed to randomly draw and look at i balls with replacement before making the guess. Furthermore, consider two possibilities: 1) $i = 2$: two balls, one white and one black, are drawn, and 2) $i = 1000$, 500 white balls and 500 black balls are drawn. In each of these scenarios, what should your subjective degrees of belief for the proposition k that the $i + 1$ ball is black?

The conceptualist answer is that for each case, the degree of belief should be $^1/_2$. Before drawing any ball from the urn, we have no knowledge about its content, so according to Laplace we are then allowed to the rule of succession, which yields identical results for both cases:

$$\frac{1+1}{2+2} = \frac{1+500}{2+1000} = \frac{1}{2}$$

Yet, Peirce points out that it makes intuitive sense to suggest that there is something different about your degrees of belief for k in these two cases, because of the vast difference in the *amount* of evidence available to you, that your belief $P(k) =^1/_2$ has more *weight* when $i = 1000$.

WEIGHT AND RELEVANCE OF EVIDENCE

One goal of probabilism is to account for many intuitions about evidence in a precise manner. To do so, it has to satisfy certain basic intuitions about how evidence relates to the hypothesis it purports to support. Two evidential relations, relevance and balance, are particularly well represented by the machinery of probabilism.

THE INDIVISIBILITY REQUIREMENT

When Keynes wrote *A Treatise on Probability*, he was keenly aware of these paradoxical results. However, he thinks that the paradoxes only suggest that the principle of indifference is to be restricted, not abandoned altogether. He argues that the reason that these paradoxes occur is because the principle of indifference should not be used when the alternatives under consideration can be further analyzed, and 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.(@keynes, 60)

Keynes's solution is influenced by Russell's logical atomism and Moore's intuitionism, the dominant views at Cambridge when he was thinking about issues in probability.(@gilliesbook, 33) Probability, in Keynes's view, is defined as a logical relation between a premise and a conclusion. Probability relations are logical, because this relation belongs to the same conceptual category as the entailment relation between the premises and conclusion in a deductive argument. The difference here is one of degree: 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. (@keynes, 30) This assumption about probability relations is the basis of a rule of rationality that governs that degrees if belief: the degrees of a belief should correspond to degrees of entailment that the belief receives as a conclusion in an argument. (@keynes, 3)

More important, these logical probability relations are Platonic entities that are acquired through intuition - not unlike Moore's non-natural normatively properties. Russell's influence manifests itself through Keynes's appeal to the distinction knowledge by acquaintance and by description. (@keynes, 11) To begin, Keynes explains that there are two kinds of judgments an agent could make through direct acquaintance of logical relations. The first is the *judgment of indifference*, which is simply what the Principle of Indifference aims to justify.(@keyes, 60) In other words, Keynes holds that we can judge that the two probability relations are equal when we perceive that

$$P(H_1|E) = P(H_2|E)$$

Note that this judgment is not used to justify the Principle of Indifference. Keynes intends it to be the other way around: the judgment of indifference is correctly applied only if the conditions for the Principle of Indifference are satisfied. The other kind of logical relations a competent perceiver can know through acquittance is the *judgement of relevance*, which is the perception about a premise's evidential relevance to its conclusion. This is critical for Keynes's defense of the principle of indifference, because it gives intelligibility to his idea of indivisibility. The intuition Keynes wants to capture is that, when we ask if the alternatives are divisible, we do not really mean conceptually physically divisible, but probabilistically: an indivisible alternative is one where no other facts can be perceived as having an effect on its probability. 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, it does not really concern us whether or not the ball is made of iron of or tin. (@keynes, 59) In other words, the material of the ball is irrelevant, and he thinks that this is something we can intuitively grasp. We will discuss the technical details of evidential relevance in the next section.

So, According to Keynes, the use of the principle in Peirce's example is not legitimate, because the probabilities should be $^{1}/_{n}$, where n is number of possible colors, not $^{1}/_{2}$. In this case of their hair color being red or not, we can see that being not-red can be analyzed into being blue, being yellow, etc., because we can see that being blue is evidentially relevant to not being red.

RELEVANT AND IRRELEVANT EVIDENCE

Some evidence is more relevant than others. A probabilistic explication of the concept of evidential relevance was spelled out by Keynes and has subsequently been accepted by modern Bayesian. His 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, it does not really concern us whether or not the ball is made of iron of or tin. (@keynes, 59) In other words, the material of the ball is irrelevant, so this is an example of judgment of irrelevant in action.

So, the intuition is that *only relevant evidence should change the probability of proposition in question*, and Keynes goes one step further by explicating the notion of relevance as how the probability of the proposition changes conditional on the evidence. Evidence E is relevant to the proposition H if and only if:

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

The idea is that if E is relevant to H in any way, when we consider them together H probably should be different than the probability of H considered alone. So, more precisely, the probability of H given E should be different the unconditional probability H. It should also be obvious that E is irrelevant to H if and only if

$$P(H|E) = P(H)$$

Note that this means E can be relevant to H in two different ways, because $P(H|E)$ can be greater or smaller than $P(H)$. This makes intuitively sense, because if H becomes less probable with E in the background, then it means E disconfirms, and therefore is relevant to, H.

WEIGHT AND RELEVANCE

Keynes recognizes that his idea of evidence relevance leads to some implausible consequence when evidence weight is involved. The increase of weight should correlate the accumulation of *relevant* evidence - ideally adding irrelevant information to our body of evidence should not increase its weight. Keynes points out that they sometimes come apart.

1. Either the

Popper credits Peirce as the first person who notes the problem of the weight of evidence and further sharpen it by formulating it as *the paradox of ideal evidence'. (@popperlogic, 425) He asks us consider a certain penny and let A be the proposition ``the nth toss of the penny will yield heads''. The proponents of subjective probability, whom he is arguing against, would suggest that the prior probability should be $^1/_2$, $P(A) = 0.5$. Now let E be what he calls *the ideal statistical evidence* in favor or the idea that the penny in question is a fair one. The actual evidence itself is not of huge importance - it just has to be some sort of statistic that would leave very little doubt that the penny is fair. Popper's example is to let B be the proposition that "out of 1,000,000 tosses of the penny, 500,000, plus or minus 20, turned up heads."

Now, Popper asks, given we have ideal evidence B, what is the probability of A? He claims that it would have to be $^1/_2$, but this would also mean:

$$P(A|B) = P(A) = 0.5$$

Before discussing exactly what is wrong with this picture, 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 $P(A)$ - it could be by indifference, through elicitation, etc. So Popper's argument, unlike Peirce's, does not hinge on whether the use of the principle of indifference is permissible. I take it to be a strength of Popper's argument.

Nevertheless, Popper is zeroing in on a different Bayesian idea that the degree of ***relevance* of evidence can be measured in terms of conditional probability. As discussed earlier, evidence E is irrelevant to the hypothesis H if and only if

$$P(H|E) = P(H)$$

Now the problem with ideal evidence becomes clear. If $P(A|B) = P(A) =^1 /_2$, this means that the ideal evidence is also irrelevant evidence. Spelling out Popper's argument exactly, it should look as follows:

1. If E is ideal evidence for the belief H, it is also relevant evidence for H.
2. E is relevant for H if and only if $P(H|E) \neq P(H)$
3. Let B some ideal evidence for A, so by premise 1 B is also relevant evidence for A .

4. Let A be the proposition that ``the nth toss of the penny will yield heads'' and $P(A) =^1 /_2$
5. But $P(A|B) = P(A) =^1 /_2$
6. $\therefore B$ is irrelevant to A.

We have then arrived at a contradiction, so these propositions cannot be all true. But it is unclear which of the premises can be rejected by a probabilist. Rejecting premise 1 would be biting the bullet - the notion of 'ideal evidence' is imported by Popper, but I also find it to have substantial intuitive appeal. Premise 2 is a core Bayesian thesis about probabilistic relevant. Premise 3 simply follows from 1. Rejecting premise 4 is the most promising and this has to be done by rejecting $P(A) = 0.5$, but how?

Without the principle of indifference, a probabilist may say that $P(A)$ can be anything as long as it is between 0 and 1, and if it is something other than 0.5, then premises 5 would be false. But it is hard to see how this can be justified. To begin, 0.5 seems to me the most reasonable prior probability one can assign to A. For the sake of the argument, let's say it begs to the question to assume 0.5 is the most reasonable prior, but it would be *ad hoc* to suggest that it cannot be 0.5 simply to avoid the paradox.

Peirce's suggestion would be to reject 4, but not by assigning $P(A)$ to some value other than 0.5, but to say $P(A)$ before any evidence should be *indefinite*. To anticipate, my proposal of using imprecise probability agrees with Peirce's assessment - the idea is that at the beginning our degrees of belief for A should be extremely vague about 0.5, but, assuming the penny is fair, the evidence will increase precision and the bounds will approach 0.5 as the weight of the evidence gets higher. The ideal evidence B, with maximum weight, will essentially bring $P(A|B)$ to precisely 0.5. Since $P(A)$ is not 0.5, premise 5 will also be false. The paradox will then be dissolved.

Peirce points out that the increase in evidential weight is associated with a certain *stability* of the degree of belief of the hypothesis in question. Intuitively, if I have only drawn one white and one black ball so far, the color of the third ball will greatly change my opinion be it white or black, but if I have already drawn 1000 balls, drawing one additional ball would have little impact on my belief. Thus, Peirce suggests that when we have limited information, knowing evidential weight becomes more important than knowing the probability itself. So when $i = 2$, the weight of evidence is so low that the assignment of probability is so unreliable that Peirce thinks that we simply refrain from doing so, but this conflicts with the precisionist thesis of conceptualism.

INDETERMINATE AND UNKNOWN PROBABILITIES

However, the indivisibility requirement rules out the use of the principle of indifference in many cases, since many probabilities are infinitely divisible. As Keynes himself recognizes, the indivisibility requirement "is fatal to the practical utility of the Principle of Indifference" when there is no *ultimate* alternatives could be found, which is the case with continuous distributions (@keynes, 68). Since the principle of indifference provides the prior distributions needed for the precise calculations of probabilities of many events, a restriction of its use entails that many probabilities are not measurable.

This has led Keynes to accept that there are probabilities that cannot be given a numeric value.

Even though Keynes makes use of the principle of indifference, his acceptance of indeterminate probability in fact brings his view a lot closer to Peirce. In his criticism of the principle of indifference, Peirce often suggests that in those cases where conceptualists are tempted to use the principle, they really should say the probability cannot be determined due to the lack of evidence.

BOOLE AND THE USE OF ARBITRARY CONSTANTS IN BAYES'S RULE

BAYESIAN SOLUTIONS