Should You Proportion Your Beliefs to Your Evidence?

Greg Gandenberger

December 3, 2015

Abstract

It is often said that a rational agent believes a proposition to the extent that his or her evidence supports it. However, this claim is incompatible with any adequate probabilistic explication of evidential support. Moreover, attempts to give an adequate non-probabilistic explication face serious obstacles. We should consider alternative views about the relationship between evidence and rational belief.

1 Introduction

It is often said that a rational agent believes a proposition to the extent that his or her evidence supports it. Call this view "evidence proportionalism." It goes back at least to Hume, who claims that "a wise man proportions his belief to his evidence" (EHU 10.1/SBN 87). Some contemporary epistemologists also endorse it. For instance, Engel (2000, 3) claims that "a belief is rational if it is proportioned to the degree of evidence that one has for its truth," and Tal and Comesaña (2015, 1) write that "rational subjects proportion their beliefs to the evidence." This idea could be regarded as a more specific version of Conee and Feldman's "evidentialist" thesis that whether a given doxastic attitude is epistemically justified for a particular agent at a particular time depends solely on whether it fits that agent's evidence at that time (1985, 15).

Evidence proportionalism seems plausible, but its precise meaning is unclear. To indicate just a few difficulties, epistemologists disagree about what kinds of things constitute evidence (Kelly, 2014); Bayesian confirmation theorists disagree about how to measure evidential support (Fitelson, 1999); and philosophers of science disagree about the adequacy of Bayesian principles (see e.g. Mayo, 1996; Achinstein, 2001; and Sober, 2008). It would be unfair to expect evidence proportionalists to resolve all of these disputes, but it is not unfair to expect them to make their central claim precise enough to have definite implications in simple cases.

Unfortunately, attempts to make evidence proportionalism precise face serious obstacles. The most common strategy for making claims about evidence and belief precise is to formalize them using probability theory. However, I will argue that any attempt to do so faces a trilemma. If it explicates the degree to which a body of evidence supports a hypothesis in terms of the prior probability of that hypothesis, then it fails to make judgments about evidential support epistemically prior to judgments about warranted belief, as evidence proportionalists seem to envision. If it violates a claim called the Likelihood Principle, then it has counterintuitive consequences. But I show that if it does neither of those things, then it violates accuracy- and coherence-related norms of belief that are more compelling than evidence proportionalism. This trilemma could be avoided by giving a non-probabilistic explication of evidence, but the alternative accounts currently on offer are too vague to give definite guidance even in simple cases.

Although evidence proportionalism seems highly plausible, there are at least two alternative views that also have some intuitive appeal. First, one could maintain that one should *update* one's beliefs in ways that are proportioned to one's evidence, but that one's initial belief states must be fixed in some other way. Second, one could maintain that what really matters is not whether the outputs of one's belief-formation processes accord with one's evidence in an appropriate way, but whether the processes themselves are reliable in some sense. Given the existence of these plausible alternatives, the correctness of evidence proportionalism should not be taken for granted.

The most substantial previous discussions of evidence proportionalism come from Goldman, who gives three reasons to reject it. His reasons have some merit, but they do not render the argument given here moot. First, Goldman denies that people have a wide range of "credal intensities" available to them (1986, 90). An evidence proportionalist could respond to this objection in a variety of ways, including granting it but denying its relevance by maintaining that evidence proportionalism is intended to apply to ideally rational agents.

Second, Goldman claims that the epistemic justification one has for believing a particular hypothesis in light of one's evidence depends on how one understands the relationship between that evidence and the hypothesis, and not just on the evidence and the hypothesis itself (1986, 90). It seems that if one simply guesses an answer to a question without consulting one's evidence, for instance, then one's belief in that answer is not justified even if it happens by luck to conform to one's evidence. This objection raises an important point but does not seem to be a decisive objection to evidence proportionalism. One possible response comes from Conee and Feldman, who concede that there is a notion of "well-foundedness" that is sensitive to the evidence one *uses* in forming an epistemic attitude, while maintaining that "justification" depends only on what evidence one *has* (Conee and Feldman, 2008). This issue is irrelevant to my objections. I aim to show that evidence proportionalism faces serious objections even when the agent in question fully understands the relevant evidential facts.

Finally, Goldman says that believing in accordance with one's evidence cannot be the sole fundamental aim of inquiry because it provides no incentive to gather new evidence (2002, 55–7). However, evidence proportionalists in my sense need not think that believing in accordance with one's evidence is the sole fundamental aim of inquiry. They might think, for instance, that the fundamental aim of inquiry is to have accurate beliefs, but that believing in accordance with one's evidence is a good means to this end. My argument supplements this one by showing that evidence proportionalists of this kind are in trouble, because proportioning one's beliefs to one's evidence in terms of the only suitable probabilistic account of evidence conflicts with accuracy- and coherence-related epistemic goals. Alternatively, an evidence proportionalist could respond to this objection from Goldman by claiming that believing in accordance with one's evidence is the sole fundamental aim of a class of "epistemic acts" that does not include acts of evidence-gathering (Greaves, 2013, 922). My argument addresses this response by raising difficulties for evidence proportionalists even within the purely epistemic realm.[1]

I proceed as follows. In Section 2, I spell out and motivate three requirements for an account of evidence that could be used in an explication of evidence proportionalism. In Section 3, I argue that those requirements pick out a single account of evidence called the Law of Likelihood. In Section 4, I show that the Law of Likelihood has disastrous consequences when combined with evidence proportionalism. In Section 5, I discuss difficulties for attempts to appeal to a non-probabilistic account of evidence and alternatives to evidence proportionalism.

2 Three Requirements for a Theory of Evidence

In this section, I articulate and motivate three requirements for an account of evidence. My claim is not that any account that violates these requirements must be incorrect, but that any attempt to combine such an account with evidence proportionalism would face difficult challenges.

Throughout this paper, I consider cases involving a definite set of epistemically possible worlds that is partitioned into a definite set of hypotheses (the "hypothesis space") and a definite set of possible observations (the "sample space"). Idealized scientific experiments have this structure. In such an experiment, a particular quantity is measured on a particular scale to a particular precision, which provides the sample space. The resulting measurements are brought to bear on, say, a set of hypotheses about the values of the parameters in a particular statistical model, which provides the hypothesis space. Even in science, the structure of real inquiry is typically not so well defined. However, it would be hard to maintain that we should trust a theory of evidence in more complicated cases if evidence proportionalism cannot handle cases that do have this simple structure.

Within this setting, I take it that any plausible account of evidence that can appropriately be called "probabilistic" would satisfy the following condition.

Probabilistic Explication. The evidential import of an observation from sam-

[1]Thanks to Richard Pettigrew for suggesting this point.

ple space **S** with respect to the elements of hypothesis space **H** depends only on a joint probability distribution $\Pr(S, H)$ over **S** and **H**, respectively.[2]

There are two ways to violate Probabilistic Explication. One is to explicate evidential notions in ways that cannot be reduced to probabilities. Some philosophers take this approach. For instance, many of them advocate an account of evidence that appeals to a non-probabilistic notion of explanatory goodness.[3] The primary challenge these accounts face is that our best accounts of explanatory goodness are notoriously vague. Explanatory goodness is often understood in terms of virtues such as accuracy, consistency, broad scope, simplicity, and fruitfulness, which can be assessed and aggregated in a variety of ways (Kuhn, 1977, 320–39). Philosophers of science have produced many insights about them, but they have not produced anything close to a precise explication of explanatory goodness that yields unambiguous judgments at least in simple cases. Some formal measures have been proposed for possibly relevant notions such as explanatory power, unification, and coherence. However, none of these measures are even intended to be complete measures of explanatory goodness; it is not clear how they could be combined with measures of other aspects of explanatory goodness to yield complete measures; and they are not by themselves good guides for belief (Glymour, 2015). At best, there is a great deal of work to be done to develop an adequate version of evidence proportionalism along explanationist lines. Other non-probabilistic accounts of evidence are similarly vague. Settling for an imprecise account is an option, but other approaches should also be explored.

The other way to violate Probabilistic Explication would be to explicate evidential notions in terms of probabilities over an algebra other than that of the elements of the sample and hypothesis spaces. This approach seems quite unmotivated. I will be assuming that one's total relevant evidence is being taken into account, so that the algebra over the sample space accounts for everything that is potentially evidentially relevant to **H**. A joint probability distribution over **S** and **H** entails marginal probability distributions $\Pr(S)$ and $\Pr(H)$ over **S** and **H** respectively, as well as conditional probability distributions $\Pr(S|H)$ over **S**

[2]If H or S is continuous, then we would consider a probability density function rather than a discrete probability distribution. I ignore this point in the main text for ease of exposition.

[3]See e.g. (Lipton, 2001), (Achinstein, 2001), (Conee and Feldman, 2008), and (McCain, 2013).

for each element of **H** and vice versa, so those distributions are not being excluded. Given these facts, the idea that we simply need a richer or otherwise different algebra seems to be a nonstarter.

Probabilistic Explication has a strong pedigree. All standard Bayesian accounts of confirmation and favoring conform to it (Fitelson, 2001), as does the error-statistical account of evidence developed by Mayo (1996), and it is often stated as an axiom in formal explications of confirmation and similar notions.[4] Within the scope of probabilistic accounts of evidence, it seems to be completely innocuous.

Here is my second requirement.

> **Non-Circularity**: The evidential import of an observation from sample space **S** with respect to the elements of hypothesis space **H** does not depend on a marginal probability distribution $\Pr(H)$ on **H**.

As indicated by its name, Non-Circularity rules out accounts that make it true that one should proportion one's beliefs about **H** to one's evidence only by giving an account of evidence that is sensitive to what one already believes or should believe about **H**. It is not a compelling requirement for probabilistic accounts of evidence in general. Indeed, many plausible accounts violate it.[5] However, those accounts are simply unavailable to an evidence proportionalist, at least of the kind that I have in mind. This kind of evidence proportionalist aims to provide an account that tells you what you should believe about the hypotheses in some hypothesis space when you have some relevant evidence but are otherwise "starting from scratch." It claims that you should bring your beliefs about those hypotheses into alignment with your relevant evidence, the import of which is fixed (at least temporarily and for the purposes of that inquiry), rather than, say, bringing them into alignment through a process of mutual adjustment.

[4]For instance, Crupi (2015) requires Probabilistic Explication for accounts of confirmation, Sprenger (2015) requires it for corroboration, and Schupbach and Sprenger (2011) require it for explanatory power. (Schupbach and Sprenger's account of explanatory power is not a counterexample to my claim that we have no precise accounts of explanatory goodness, because Schupbach and Sprenger concede that explanatory power is only one aspect of explanatory goodness.)

[5]For instance, $\Pr(H|E)/\Pr(H)$ (or its logarithm) is a popular Bayesian measure of confirmation (Milne, 1996).

This is not to say that evidence proportionalists are committed to naïve views about evidence being "given by experience" without any reference to background beliefs. It is just to say that, at least in many cases, whatever background beliefs are involved in assessing the evidence are not beliefs about the hypotheses under consideration and can be held fixed for the purposes of that evaluation. When you draw a colored ball from an urn of unknown composition, for instance, your beliefs about the contents of the urn are not ordinarily implicated in your beliefs about the color of the ball (assuming good lighting, ample time to inspect the ball, and so on). Non-Circularity might be problematic when one is investigating the very processes that are involved in producing one's evidence about it, but my objections to evidence proportionalism do not appeal to cases of this kind.

Probabilistic Explication says that the evidential import of an observation from a sample space $\mathbf{S}$ for a hypothesis space $\mathbf{H}$ depends only on $\Pr(S, H)$, which can be decomposed in the the product $\Pr(S|H)\Pr(H)$. Non-Circularity says that it does not depend on $\Pr(H)$. Thus, it follows from the two requirements given so far that it can only depend on $\Pr(S|H)$.

My final requirement narrows down what evidential import could depend on still further.

> **The Likelihood Principle**: The evidential import of an observation E from sample space $\mathbf{S}$ with respect to the elements of hypothesis space $\mathbf{H}$ depends on $\Pr(S|H)$ only through $\Pr(E|H)$ as a function of H on $\mathbf{H}$, up to a constant of proportionality.

The Likelihood Principle rules out the possibility left open by Probabilistic Explication and Non-Circularity that the evidential import of E for $\mathbf{H}$ depends on the values that $\Pr(S|H)$ takes at points in $\mathbf{S}$ that correspond to observations other than E—observations that could have been made but were not. Stated as a positive claim, it says that the evidential import of an observation for a set of hypotheses depends only on what those hypotheses say about the probability of *that observation*.

$\Pr(E|H)$ considered as a function of H on $\mathbf{H}$ is called the *likelihood function* of E on $\mathbf{H}$. You might wonder why the Likelihood Principle says that the evidential import of E for $\mathbf{H}$ depends on the likelihood function of E on $\mathbf{H}$ only *up to a constant of proportionality*.[6]

[6]Some writers on this topic say that likelihood functions themselves are defined only up to a constant of

Consider the following set of examples. Suppose again that you wish to evaluate whether a particular coin is fair. In Case 1, you observe that it lands heads on a single flip. In Case 2, you observe that it lands heads on a single flip *and* that an unrelated fair die produced a six on a single roll. In Case 1, the probability of your observation is p, where p is the probability of heads on a given flip of that coin. In Case 2, the probability of your observation is $(1/6)p$. The Likelihood Principle yields the obvious judgment that these observations have the same evidential import for hypotheses about p only if it includes the phrase "up to a constant of proportionality" to account for the 1/6 factor that the irrelevant die roll introduces.

This example is not enough to show that evidential import depends on likelihood functions up to a constant of proportionality *in general*. After all, not all pairs of observations whose likelihood functions differ only by a constant factor are the same except for the addition in one of irrelevant information. However, the example can be used to illustrate a more general argument. Add to that example a Case 3 in which we observe something with the likelihood function $(1/6)p$ that does not differ from the observation in Case 1 merely by the addition of irrelevant information. Even without the phrase "up to a constant of proportionality," the Likelihood Principle would entail that the evidential import of this observation is the same as that of the observation in Case 2. Given our previous conclusion that the observations in Case 1 and Case 2 have the same evidential import, it follows that the evidential import of this observation is the same as that of the observation in Case 1. In general, whenever we have two observations that have the same likelihood function up to a constant of proportionality, we can imagine a third observation that is the same as one those observations except for the addition of irrelevant information and has exactly the same likelihood function as the other observation. We can then apply the Likelihood Principle restricted to hypotheses with the same likelihood function to conclude that this third observation has the same evidential import as the second observation, and the principle that we can ignore irrelevant evidence to conclude that it has the same evidential import as the first observation, and thus conclude by transitivity that the two original observations have the same evidential import. Therefore, if observations that have the same likelihood

proportionality, which would make the phrase "up to a constant of proportionality" in my formulation of the Likelihood Principle redundant. This difference is merely terminological.

function have the same evidential import, then observations that have merely proportional likelihood functions also have the same evidential import.

The Likelihood Principle seems intuitively plausible. Suppose, for instance, that you wish to evaluate whether a particular coin is fair. Your associate reports to you that she flipped it ten times and observed eight heads and two tails. It probably would not occur to you to ask her why should stopped after ten flips. Perhaps she planned all along to flip the coin ten times, for instance, or perhaps she planned to flip until she had observed two tails. Either way, the data consist of eight heads and two tails. The difference between these two scenarios is only a difference in your associate's intentions about whether or not she would have continued flipping if her observations had been different. It seems quite odd to claim that this difference in intentions, which made no difference to what actually happened, could make a difference to the evidential import of the data. Some statistical methods in common use violate this intuition, but the Likelihood Principle supports it: one can show that, outside of unusual cases, the likelihood functions will be proportional in cases of this kind, which produce the same data under different "stopping rules" (Raiffa and Schlaifer, 1961).

In addition to its immediate intuitive appeal, the Likelihood Principle is supported by proofs from principles that are even more compelling (Birnbaum, 1962; Berger and Wolpert, 1988; Gandenberger, 2015). However, it is contentious. Mayo (2014) has recently argued that the best-known proof (due to Birnbaum) is invalid, and it faces many purported counterexamples.[7] Given this controversy, one might think that it would be better to give up the Likelihood Principle than to give up Probabilistic Explication, Non-Circularity, or evidence proportionalism. Doing so might have some surprising consequences for our concept of evidence—for instance, that the evidential import of an observation can depend on the intentions of the person who ran the procedure that produced it—but avoiding such consequences is less important than preserving the role that we want the notion of evidence to

[7]For discussion of various purported counterexamples, see (Armitage, 1961); (Stein, 1962); (Birnbaum, 1964, 12–3) and (Royall, 1997); (Stone, 1976), (Fraser et al., 1985), (Evans et al., 1986), (Berger and Wolpert, 1988, 127–36), and (Hill, 1988, 161–74); (Sober, 1983, 354–6); (Leeds, 2004) and (Chandler, 2013, 133–4); (Sober, 2005, 128–9) and (Fitelson, 2007, 4–5); Forster (2006); (Sober, 2008, 37–8); (Fitelson, 2013); and (Gandenberger, MS).

play in guiding our beliefs.

The danger here is that the more willing we are to revise our intuitions about evidence in order to make evidence proportionalism come out true, the more we are treating evidence proportionalism like an analytic truth, as opposed to a substantive claim that can guide philosophical theorizing. I claim that we would have to be willing to revise intuitions to quite a large extent in order to give up the Likelihood Principle, and thus that taking this approach comes at a high cost.

To justify this claim, I will briefly review a standard argument for the Likelihood Principle. Suppose you want to assess the probability that a particular coin lands heads when flipped. You flip it some predetermined number of times, generating a sequence of heads and tails. If you are willing to assume that the outcomes of the flips are independent and that the probability of heads is unchanging, then it seems that you should not have to worry about the *precise sequence* of outcomes: all you need to pay attention to is the number of flips and the number of heads.

More generally, statisticians have developed a notion of a *sufficient statistic*, which provides a coarse-grained description of an experimental outcome without discarding any information relevant to the hypothesis space that is under consideration. Formally speaking, a statistic is sufficient with respect to a hypothesis space if and only if for each value of that statistic, the probability distribution over the full sample space conditional on that value is independent of which element of the hypothesis space is true. For instance, a statistic that reports the number of heads in a sequence of ten coin tosses (where the number of tosses was fixed in advance) is sufficient with respect to hypotheses about the bias of the coin (assuming that the tosses are independent and the bias is fixed) because each value of that statistic corresponds to a set of sequences of heads and tails all of which have the same probability[8] conditional on that value of the statistic, regardless of the coin's true bias.

Birnbaum's proof of the Likelihood Principle appeals to the *Sufficiency Principle*. This principle says that if two outcomes of a single experiment (that is, a single observational

[8] In this case, the conditional probability distribution over the elements of the sample space that are consistent with a particular value of the sufficient statistic, given that value, is *uniform* under every element of the hypothesis space. However, the notion of sufficiency does not require that it be uniform under every hypothesis, but only that it be the *same* under every hypothesis.

situation with specified sample and hypothesis spaces) correspond to the same value of a statistic that is sufficient with respect to that experiment's hypothesis space, then they have the same evidential import for that hypothesis space. This claim seems entirely innocuous. In terms of the informal motivation for the notion of a sufficient statistic, it says that if two outcomes of the same experiment differ only with regard to information that is irrelevant to its hypothesis space, then they have the same evidential import for that hypothesis space. In terms of the formal characterization of sufficiency, it is hard to see how the information that a sufficient statistic omits could be evidentially relevant to the hypothesis space, given that each element of the hypothesis space implies the same probability distribution over that information conditional on the information the sufficient statistic does provide.

To introduce Birnbaum's second principle. suppose that you had made a random choice between two possible procedures. For instance, you flipped another coin that was unrelated to the coin under investigation. Because that coin landed heads, you flipped the coin under investigation ten times. However, if it had landed tails, then you would have flipped it until you had gotten two tails. It does not seem that you should change your assessment of the coin's bias for heads in light of this fact.

In general, it seems that facts about other experiments that could have been but were not performed are irrelevant to the evidential import of the outcome of the experiment that was performed, provided that the outcome of the process by which the choice among those experiments was made is not itself informative. This is Birnbaum's *Conditionality Principle*. Again, it seems entirely innocuous.

Birnbaum showed that the Likelihood Principle follows from the conjunction of the Sufficiency and Conditionality Principles (1962). Assuming that Birnbaum's proof is valid (*pace* Mayo)[9], giving up the Likelihood Principle requires giving up either the Sufficiency Principle or the Conditionality Principle. But again, those principles seem utterly compelling. Moreover, it is difficult to see how inquiry could proceed without assuming something like them. Because we cannot attend to or report every aspect of an observational situation (including the number of blades of grass visible through the laboratory window, for instance),

[9]Mayo (2014) claims that Birnbaum's proof is invalid, but Dawid (2014) and Gandenberger (2015) have responded to her arguments.

we have to rely on something like the Sufficiency Principle, at least implicitly, to tell us which aspects of that situation we can safely ignore. Because we cannot trace back all of the contingencies that might have resulted in a different experiment being performed (including, perhaps, the exact way the light was shining through the experimenter's window when inspiration struck), we have to assume something like the Conditionality Principle, which says that those contingencies can typically be ignored. Those who would deny the Likelihood Principle face the burden of providing plausible replacement principles that can do the same work without implying the Likelihood Principle.

On the other hand, evidence proportionalists who do not deny the Likelihood Principle owe us an account of evidence that accords with it and makes good on their central claim. In the next two sections, I argue that no such account is possible.

3 The Law of Likelihood Is the Only Plausible Account of Evidence That Satisfies the Three Constraints

In this section, I argue that an account of evidential favoring called the Law of Likelihood is the only plausible account of evidence that an evidence proportionalist might be able to use which satisfies the three requirements discussed in the previous section. I will then argue in the next section that this account cannot in fact be combined with evidence proportionalism.

Let us begin by considering Bayesian measures of confirmation. These measures satisfy Probabilistic Explication, and many of them satisfy the Likelihood Principle as stated in the previous section. There are many such measures and no consensus about which one is best. Chandler (2013) identifies the following six as "arguably the most popular."

(d) $\Pr(H|E) - \Pr(H)$

(c) $\Pr(H\&E) - \Pr(H)\Pr(E)$

(s) $\Pr(H|E) - \Pr(H|\neg E)$

(r) $\log\left[\frac{\Pr(H|E)}{\Pr(H)}\right]$

(n) $\Pr(E|H) - \Pr(E|\neg H)$

(l) $\log\left[\frac{\Pr(E|H)}{\Pr(E|\neg H)}\right]$

Unfortunately, none of these measures satisfy Non-Circularity: they can all be varied by varying $\Pr(H)$ over the hypothesis space $\mathbf{H}$ while $\Pr(S|H)$ over $\mathbf{H}$ and the sample space $\mathbf{S}$ is held fixed. This fact is obvious for $(d), (c), (s)$, and (r), all of which appeal explicitly to one or both of $\Pr(H)$ and $\Pr(H|E)$. It is also true for (n) and (l) because those measures appeal to $\Pr(E|\neg H)$, which typically varies with the probabilities that are assigned to the components of $\mathbf{H}$ that make up $\neg H$.[10]

One could try to avoid violating Non-Circularity by developing a measure of the degree to which E supports a particular hypothesis H that is a function of $\Pr(E|H)$ alone. However, no such measure can succeed. The basic problem is that evidential support sometimes increases with $\Pr(E|H)$ and sometimes decreases with $\Pr(E|H)$, in a way that depends on features of the problem other than $\Pr(E|H)$ itself. The following pair of examples illustrates this point. First, suppose a three-sided die is known either to be fair or to have bias probability .39 of yielding a one, .042 of yielding a two, and probability .568 of yielding a three. Call the second of these hypotheses H^*. In this situation, a roll that yields a one confirms H^* more than a roll that yields a two. (In fact, a roll that yields a two disconfirms H^*.) Thus, as one might expect, it can happen that a datum E for which $\Pr(E|H) = .042$ confirms H less than a datum for which $\Pr(E|H) = .39$. Next, suppose a coin is known either to be fair or to have bias 9/10 for heads. Call the second of these hypotheses $H^\dagger$. A sequence containing nine hundred heads in one thousand tosses confirms $H^\dagger$ more than the a sequence containing nine heads in ten tosses. But in the first case, $Pr(E|H^\dagger) = .042$, while in the second case $\Pr(E|H^\dagger) = .39$. Thus, perhaps surprisingly, it can sometimes happen that a datum E for which $\Pr(E|H) = .042$ confirms H more than a datum for which $\Pr(E|H) = .39$. From this pair of conclusions, it follows that no measure of favoring based on $\Pr(E|H)$ alone can rank instances of confirmation correctly in all cases, and thus that no such measure is adequate.

[10](n) and (l) do not vary with $\Pr(H)$ on $\mathbf{H}$ when $\Pr(E|H)$ is the same for all elements of $\mathbf{H}$ in $\neg H$. Thus, (n) and (l) satisfy Non-Circularity in this special case. However, this case is unusual enough that this point is of no great interest. Moreover, (n) is susceptible at least to the first objection and (l) to the second objection that I raise against the Law of Likelihood in the next section.

This problem for measures of support based on $\Pr(E|H)$ alone reveals that some kind of normalization is needed; what is relevant is not how probable H makes E in an absolute sense ($\Pr(E|H)$), but how probable H makes E *relative to alternatives to* H. This idea is implicit in Bayesian reaosning. For a Bayesian, $\Pr(H|E)$ is proportional to $\Pr(E|H)$ divided by a weighted sum over $\Pr(E|H)$ as a function of H on $\mathbf{H}$, with the weights given by a prior probability distribution over $\mathbf{H}$. Unfortunately, Non-Circularity requires doing without appealing prior probabilities. Indeed, together with Probabilistic Explication and the Likelihood Principle, it rules out any kind of any non-arbitrary differential weights over alternatives to the particular hypothesis H that is being considered. As a result, any measure that conforms to those principles and appeals to the full likelihood function over $\mathbf{H}$ has to be equally sensitive to highly plausible and highly implausible alternatives to H. It will also be very sensitive to how the sample space is carved up.

We seem to have encountered a dead end with measures of evidential support that conform to our three requirements. However, there is another option: we could consider measures of evidential *favoring* for one hypothesis relative to another. The most popular measure of this kind is given by the *Law of Likelihood*, which says that evidence E favors hypothesis H_1 over (mutually exclusive) hypothesis H_2 if and only if the *likelihood ratio* $\mathcal{L} = \Pr(E|H_1)/\Pr(E|H_2) > 1$, with $\mathcal{L}$ measuring the degree of favoring. Taking the ratio of the two likelihoods provides the normalization that measures based on $\Pr(E|H)$ alone lack, and focusing on relative judgments concerning just a pair of hypotheses obviates the need for weights over alternative hypotheses.

Are there genuine competitors to the likelihood ratio as a measure of evidential favoring? Perhaps the most obvious *prima facie* competitor is the likelihood difference $\Pr(E|H_1) - \Pr(E|H_2)$. However, this measure fails for the same reason as $\Pr(E|H)$ as a measure of support: it tends to decrease as the amount of relevant data increases, even if the degree of evidential support / favoring is obviously increasing. Another *prima facie* competitor is log-likelihood ratio $\log(\Pr(E|H_1)/\Pr(E|H_2))$, but this measure is not importantly different from the likelihood ratio because evidential favoring is generally taken to be defined only up to ordinal equivalence.

To my knowledge, the only other alternatives to the Law of Likelihood that have been proposed are derived from Bayesian measures of confirmation in accordance with the principle that E favors H_1 over H_2 if and only if it confirms H_1 more than H_2 (Fitelson, 2013).[11] Given this principle, each Bayesian measure of confirmation implies a criterion for evidential favoring (though not yet a measure of degrees of favoring). The problem for these measures in the present context is the same as the problem for the measures of confirmation on which they are based: they violate Non-Circularity. The one exception to this claim among the six measures of confirmation mentioned above is the measure $(r) : c(H, E) = \log\left[\frac{\Pr(H|E)}{\Pr(H)}\right]$. One can show that given this measure, E confirms H_1 more than H_2 if and only if $\Pr(E|H_1) > \Pr(E|H_2)$. Thus, this measure yields the same criterion for evidential favoring as the Law of Likelihood.

It is possible to go beyond evaluating proposed alternatives to the Law of Likelihood on a case-by-case basis and give a general argument that the Law of Likelihood provides the only plausible criterion or measure of evidential favoring that conforms to our three requirements. The Likelihood Principle says that the evidential import of E with respect to $\{H_1, H_2\}$ depends only on $\Pr(E|H_1)$ and $\Pr(E|H_2)$ up to a constant of proportionality. It already follows that the evidential import of E with respect to $\{H_1, H_2\}$ depends only on the likelihood ratio $\mathcal{L} = \Pr(E|H_1)/\Pr(E|H_2)$. A rule for assessing E as evidence with respect to H_1 and H_2 should be symmetric with respect to interchange of the labels H_1 and H_2. It follows that E must be evidentially neutral between H_1 and H_2 when $\mathcal{L} = 1$. Given that whether E favors H_1 over H_2 or vice versa or is neutral between them depends only on $\mathcal{L}$, it presumably cannot be the case that E favors H_2 over H_1 when $\mathcal{L} < 1$. There could be a constant $a > 0$ such that E favors H_1 over H_2 if and only if $\mathcal{L} > a$, but the choice of such a constant would be arbitrary. Thus, it seems we must say that E favors H_1 over H_2 (perhaps to an arbitrarily small degree) if and only if $\mathcal{L} > 1$.

We have now vindicated the *qualitative* part of the Law of Likelihood, which provides a *criterion* for evidential favoring. It remains to vindicate the *quantitative* part, according to which the degree to which E favors H_1 over H_2 is monotone increasing in $\mathcal{L}$. Suppose

[11] It may be necessary to require for favoring that H_1 and H_2 be mutually exclusive; see (Chandler, 2013) and (Gandenberger, MS) for discussion.

this claim were false. Then for an observation O of a string of heads as long as one likes on a sequence of independent and identically distributed coin tosses, it would be possible to construct pairs of hypotheses H and H' such that H posits a higher probability of heads on each toss than H', yet O favors H' over the hypothesis H_F that the coin is fair at least as strongly as it favors H over H_F.[12] I take it that this conclusion is unacceptable, and thus that the degree to which E favors H_1 over H_2 must be monotone increasing in $\mathcal{L}$. In other words, $\mathcal{L}$ measures the degree to which E favors H_1 over H_2 up to ordinal equivalence, exactly as the Law of Likelihood asserts.

There is a small gap in this argument. It assumes that the evidential import of E with respect to $\{H_1, H_2\}$ is the same regardless of what other hypotheses (if any) are being considered. This assumption seems plausible, but the contrary possibility that degrees of evidential favoring are sensitive to the theoretical context in which they are being evaluated is not obviously false. However, this point is not relevant in the present context. Many of the Bayesian criteria for favoring mentioned above have this "contextual" character, but they violate Non-Circularity. In general, given our requirements, a criterion or measure of favoring that considered alternatives to the pair of hypotheses under consideration would have to be based on their likelihoods only, but avoiding unreasonable sensitivity to implausible hypotheses would often require differentially weighting those likelihoods in a way that our requirements do not allow. Thus, "contextual" criteria and measures or favoring are not available to those who accept our three requirements. I conclude that the Law of Likelihood is the only plausible criterion or measure of evidential favoring that conforms to those requirements. There does not seem to be any evidential notion other than support or favoring that can serve evidence proportionalist purposes, so the Law of Likelihood is the only evidential measure of any kind that we need to consider.

[12] If the degree to which E favors H_1 over H_2 is not monotone increasing in $\mathcal{L}$, then there are an l_1 and an l_2 such that $l_1 \geq l_2$, yet for any $H_1, \ldots, H_4$ such that $l_1 = \Pr(E|H_1)/\Pr(E|H_2)$ and $l_2 = \Pr(E|H_3)/\Pr(E|H_4)$, E favors H_3 over H_4 at least as strongly as it favors H_1 over H_2. To construct hypotheses of the kind described in the main text, simply choose H and H' so that $\Pr(O; H)/\Pr(O; H_F) = l_1$ and $\Pr(O; H')/\Pr(O; H_F) = l_2$ for some l_1 and l_2 of this kind. Making this choice is possible because the length of O can be chosen to make $\Pr(O; H_F)$ arbitrarily small, and for a given O, H and H' can be chosen to make $\Pr(O; H)$ and $\Pr(O; H')$ arbitrarily close to 0 or 1.

4 Likelihoodist Proportionalism Has Unacceptable Consequences

In the preceding sections, I argued that there is no plausible probabilistic measure of the degree to which some evidence supports a single hypothesis considered in isolation that can do the work evidentialism proportionalism requires of it. However, there is a measure of the degree to which some evidence *favors one hypothesis over another* that meets all of the requirements we have laid down so far, namely the likelihood-ratio measure $\mathcal{L} = \Pr(E|H_1)/\Pr(E|H_2)$ that is provided by the Law of Likelihood. We can adapt the evidence proportionalism to this principle by regarding it as the claim that one should proportion one's *relative* belief in one hypothesis *over another* to the degree to which one's total evidence favors the former over the latter. I argue in this section that this idea does not work, and thus that evidence proportionalism cannot be combined with a probabilistic explication of evidential notions.

It might seem that the idea of combining evidence proportionalism with the Law of Likelihood is an *obvious* non-starter because of examples like the following (due to Sober, 2008, 10). Suppose you hear a banging sound coming from your attic. Let us say that plumbing problems in the attic would not usually produce such a sound, but a gremlin in the attic would. Thus, $\Pr(\text{sound}|\text{gremlins})/\Pr(\text{sound}|\text{plumbing problems}) \gg 1$, so the Law of Likelihood says that the sound favors the hypothesis of gremlins in the attic over the hypothesis of plumbing problems in the attic to a high degree. Nevertheless, the posterior probability of plumbing problems is presumably much higher than that of gremlins, simply because plumbing problems are (at least) much more common than gremlins.

However, this example simply exploits the fact that the evidential import of one piece of evidence can be quite different from the evidential import of a total body of evidence that contains it. Presumably, your total body of evidence contain many credible reports of plumbing problems but none of gremlins. Thus, even if the sound from the attic favors the gremlin hypothesis over the plumbing-problems hypothesis, your total evidence as a whole might point quite strongly in the opposite direction. If so, then proportioning your relative beliefs to your *total* evidence would yield the sensible result of believing the plumbing-problems hypothesis over the gremlin hypothesis. Thus, examples of this kind are not enough to show that proportioning one's relative beliefs to one's total evidence in accordance with the Law of Likelihood will yield bad outcomes.

However, combining evidence proportionalism with the Law of Likelihood does lead to other, less obvious problems. In fact, problems arise even from a very weak claim that I will call *minimal likelihoodist proportionalism* (MLP). Rather than saying that one should believe one hypothesis over another hypothesis exactly to the degree to which one's total evidence favors the one over the other, this claim says simply that there is *some* degree of favoring c such that, for any pair of hypotheses H_1 and H_2, one should believe H_1 over H_2 *to some degree* if one's total evidence T favors H_1 over H_2 to degree c or greater; that is, if $\Pr(T|H_1)/\Pr(T|H_2) > c$. The order of the quantifiers here is important: an MLPer is committed to there being *some* degree of favoring that suffices for a degree of relative belief for *any* pair of hypotheses. The degree of relative belief can be arbitrarily small and can vary across pairs of hypotheses, but the degree of favoring must be fixed. A Bayesian would reject this claim: the likelihood ratio that is required to induce a Bayesian to believe a given hypothesis H_1 over an given alternative H_2 to some degree (in the sense that $\Pr(H_1|T)/\Pr(H_2|T) > 1$, where those probabilities represent his or her degrees of belief) is simply $\Pr(H_2)/\Pr(H_1)$, which can be arbitarily small. However, an evidence proportionalist cannot reject it. To do so, he or she would have to appeal to something other than evidence to decide what degree of favoring to require in a given case, such as initial probabilities $\Pr(H_1)$ and $\Pr(H_2)$, but to do so would be contrary to the aims of evidence proportionalism.

I will now present two problems for MLP. Consider the following example.

> **Example 1.** Suppose you observe ten radioactive isotopes labeled i_1 to i_{10} of a particular species for a period equal to their half-life. You are sure that each of those isotopes decays or not independently of the others, but you are not sure whether radioactive decay is deterministic. Thus, you consider the hypothesis $H_{50\%}$ that each of the ten isotopes has a 50% chance of decaying independently

of the others and the 2^{10} hypotheses each of which says that exactly some subset of the isotopes is bound to decay. You have no evidence about those hypotheses prior to the experiment.

Suppose you observe E_{23469} : only the isotopes i_2, i_3, i_4, i_6, and i_9 decay. How does this outcome bear on the deterministic hypothesis H_{23469} that exactly those isotopes were bound to decay, against the indeterministic hypothesis $H_{50\%}$? According to the Law of Likelihood, it favors the former over the latter to the substantial degree 1024.[13]

This result might look bad for the Law of Likelihood. After all, E_{23469} does not seem to make H_{23469} substantially more believable than $H_{50\%}$. The standard response on behalf of the Law of Likelihood is that the Law of Likelihood itself does not address the question of which hypothesis is more believable in light of the data, but only the question of which hypothesis the data favor and by how much (Royall, 1997, 13–5). If you want to know what to believe in light of the data, advocates of the Law of Likelihood say, then you should use a Bayesian approach. However, this response is not available to someone who accepts MLP.

An MLPer can set the likelihood ratio threshold that always suffices for believing one hypothesis over another as high as they like, including to values higher than 1024. But however high they set it, one can describe an experiment like the one described in Example 1 the result of which will inevitably favor some deterministic hypothesis over $H_{50\%}$ to some degree greater than that threshold: an experiment like the one described in Example 1 will inevitably yield a result that favors the deterministic hypothesis that correctly predicts the data over $H_{50\%}$ to the degree 2^n, where n is the number of isotopes observed, so one simply needs to increase n until 2^n is greater than the threshold. Thus, for any MLPer, one can describe an experiment like the one described in Example 1 (though possibly larger) the result of which will inevitably lead him or her to believe some deterministic hypothesis over $H_{50\%}$.

To make matters worse, given the result of the experiment, H^* and $H_{50\%}$ are respectively equivalent to the hypotheses H_d that the experiment is deterministic and the hypothesis

[13] H_{23469} gives E_{23469} probability 1, while $H_{50\%}$ gives it probability $1/2^{10}$, so $\Pr(E|H_{23469})/\Pr(E|H_{50\%})) = 1/(1/2^{10}) = 1024$.

H_{ind} that it is indeterministic. Thus, assuming that relative belief is closed under known single-premise entailment conditional on one's evidence, it follows that the experiment will inevitably lead an MLPer who knows that H^* is equivalent to H_d and $H_{50\%}$ to H_{ind} to believe (to some degree) H_d over H_{ind}. However, the result of the experiment seems completely irrelevant to the issue of determinism.[14] To make matters worse still, because the outcome of a sufficiently large radioactive decay experiment will inevitably lead an MLPer to believe (to some degree) H_d over H_{ind}, it seems that he or she should not have to wait for the outcome of the experiment at all—in accordance with an analogue of van Fraassen's reflection principle (1995),[15] he or she should simply believe (to some degree) determinism over indeterminism from the beginning.

These plainly mistaken conclusions follow inexorably from MLP, given the Law of Likelihood. But MLP and the Law of Likelihood are required by evidence proportionalism, if we attempt to understand the notion of evidence probabilistically. Thus, we must either abandon evidence proportionalism or give a non-probabilistic account of evidence.

An MLPer gets into trouble in this case because MLP requires that some particular likelihood ratio d be sufficient for (some degree of) relative belief in *all* cases. For a Bayesian, by contrast, the likelihood ratio that is required to make the posterior probability of one hypothesis larger than that of another depends on his or her initial degrees of belief in the hypotheses and has no upper bound. If he or she assigns the same probability to $H_{50\%}$ in the radioactive decay case regardless of the number of isotopes observed and distributes the remaining $1 - p$ probability equally among the deterministic hypotheses, then the prior probabilities for those hypotheses diminish as 2^n as the number of isotopes increases, exactly compensating for the increase in the likelihood ratio for H^* against $H_{50\%}$. These constraints

[14]For a Bayesian, the result of the experiment is irrelevant to the issue of determinism if all of the deterministic hypotheses have equal prior probabilities. If some are more probable than others, then a result that agrees with one of the more probable deterministic hypotheses will raise the overall probability of determinism, while a result that agrees with one of the less probable deterministic hypotheses will lower the probability of determinism. The intuition that the result is completely irrelevant plausibly arises from the fact that one has no reason to prefer one deterministic hypothesis over another prior to the experiment. In any case, a Bayesian would not approve of believing H_d over H_{ind} regardless the the experimental outcome if one were neutral between H_d and H_{ind} before the experiment began.

[15]Van Fraassen's *General Reflection Principle* says that for any future time t, one's current opinion about a hypothesis must lie in the span of what one currently regards as the opinions that one might hold about that hypothesis at t.

on the prior are not *ad hoc*; they are required to represent the beliefs of a Bayesian who thinks that $H_{50\%}$ and the 2^n possible deterministic hypotheses are the only possibilities, regards all possible observations as equally probable, and regards the size of the experiment and the truth of $H_{50\%}$ as probabilistically independent. Roughly speaking, a Bayesian who is uncertain about what will be observed will naturally penalize hypotheses like the deterministic ones in this case that make very definite predictions with his or her prior probabilities, but an evidence proportionalist has no way to do likewise.

Thus, MLPers get into trouble in the example just discussed because they cannot use prior probabilities to penalize hypotheses for making very definite predictions, as a Bayesian who is uncertain about what will be observed would do automatically. A second example shows that they also cannot use prior probabilities to penalize hypotheses for logical strength, for instance by requiring a larger likelihood ratio to believe (to some degree) H_1 over $\neg H_1$ than to believe (to some degree) H_2 over $\neg H_2$ when H_1 is strictly logically stronger than H_2. As a result, MLP can produce a paradoxical situation in which H_1 entails H_2, yet one believes (to some degree) both H_1 over $\neg H_1$ and $\neg H_2$ over H_2.

As an illustration of this phenomenon, consider the following example.

> **Example 2.** The El Niño/Southern Oscillation (ENSO) is a cyclical pattern of temperature changes in the central and east-central equatorial Pacific Ocean. "El Niño" refers to warm phases in this cycle, "La Niña" to cool phases. One complete cycle typically lasts about three to five years. ENSO is associated with many weather patterns around the world. For instance, hurricane activity in the East Pacific is generally greater during an El Niño that during other phases of the cycle (NOAA, 2012).
>
> Suppose that in 2015 you are put in charge of planning hurricane relief efforts for the Peruvian government. You are told that the area has been experiencing greater than normal hurricane activity but not whether we are in an El Niño. A major event is planned for the summer of 2017, and you are concerned about the possibility that it will be disrupted by a hurricane. Thus, you wish to evaluate both the hypothesis H_1 of an El Niño in both 2015 and 2017 and the hypothesis

H_2 of an El Niño in 2017.

Unfortunately, MLP can exhibit anomalous behavior when used to evaluate those hypotheses against their negations. In particular, given no additional evidence, it could lead one to believe (to some degree) both the hypotheses of an El Niño in both 2015 and 2017 over its negation and the hypothesis of no El Niño in 2017 over its negation. But this combination of beliefs makes no sense because an El Niño in 2015 and 2017 entails an El Niño in 2017.

How does this problem arise? The period of the ENSO cycle is typically three to five years, so if there will be an El Niño in 2017, as H_2 posits, then it is likely that there is a La Niña in 2015. A La Niña in 2015 would give a low probability to more hurricane activity than normal in 2015. Thus, the evidence E of more hurricane activity than normal in 2015 speaks against H_2 relative to $\neg H_2$. By contrast, increased hurricane activity would be rather likely given H_1, which posits the unlikely but possible event of El Niños occurring only two years apart, in 2015 and 2017. Thus, increased hurricane activity in 2015 favors H_1 over $\neg H_1$ and $\neg H_2$ over H_2. Depending on the details, it could do so to a degree that would cause an MLPer to believe (to some degree) H_1 over $\neg H_1$ and $\neg H_2$ over H_2.

The details of this concrete example should not be taken too seriously: the relevant likelihoods lack clear objective values, and one presumably has relevant evidence other than E. It is simply an illustration of the relevant mathematical result, which is that for any d, there are possible hypotheses H_1 and H_2 and possible data E such that H_1 entails H_2, yet $\Pr(E|H_1)/\Pr(E|\neg H_1)$ and $\Pr(E|\neg H_2)/\Pr(E|H_2)$ both exceed d. See the appendix for a proof of this claim. This result is no problem for a Bayesian, who will automatically assign at least as much posterior probability to H_2 as to H_1 upon learning E in such a situation simply by following the rules of probability.[16] It is a problem for an MLPer, however, because it shows that no matter what likelihood ratio threshold for relative belief he or she uses, he or she could be led to believe an H_1 over $\neg H_1$ and $\neg H_2$ over H_2 in a case in which H_1 entails H_2.[17] As noted above, this pattern of beliefs seems to make no sense.

[16]It is an easy consequence of the axioms of probability that if H_1 entails H_2, then $\Pr(H_1)/\Pr(\neg H_1) \leq \Pr(H_2)/\Pr(\neg H_2)$, and likewise when those probabilities are conditioned on some common event E.

[17]Note that MLP is not being applied to the comparison between H_1 and H_2 directly, but to the com-

One can think of these problems as illustrating conflicts between believing in accordance with one's evidence and achieving other epistemic goals, given a proportionalist understanding of what it means to believe in accordance with one's evidence and a commitment to probabilistic accounts of evidential notions. The first problem illustrates a conflict between believing in accordance with one's evidence and forming one's beliefs in a reliable manner. An MLPer will be led to believe hypotheses that make strong predictions that turn out to be correct, which would result in inaccurate beliefs in cases involving indeterministic phenomena. MLP is thus at odds with a probabilistic version *sensitivity*, where a belief is sensitive if and only if one probably would not have formed it if it were false (Roush, 2007). An even greater conflict arises if we assume conformity to the coherence-related goal of having beliefs that are closed at least under single-premise entailment: we get a global belief in determinism from any sufficiently large set of observations, regardless of whether or not it is true.

The second problem illustrates a straightforward conflict between believing in accordance with one's evidence (again, understood in a proportionalist way using a probabilistic account of evidence) and the coherence-related goal of not believing a hypothesis H_1 over $\neg H_1$ and $\neg H_2$ over H_2 when H_1 entails H_2. Violations of these accuracy- and coherence-related goals are unacceptable, so we must either give up the goal of believing in conformity with our evidence or provide an alternative account of evidence. I discuss these options in the next section.

5 Alternatives to Probabilistic Evidence Proportionalism

If the conclusions I have reached so far are correct, then either evidence proportionalism is false or the notion of evidence cannot be understood probabilistically. In this section, I provide a tentative argument for giving up evidence proportionalism by indicating some

parison of each of these hypotheses against its negation, consistent with the restriction of MLP to mutually exclusive hypotheses.

difficulties for the most obvious ways to give a non-probabilistic account of evidence and suggesting that there are alternatives to evidence proportionalism that have some plausibility.

As noted above (p. 5), our best nonformal accounts of evidence are at present too vague to provide much guidance. We might consider formal tools outside of standard probability theory, such as sets of probability measures, Dempster-Shafer belief functions, possibility measures, plausibility measures, and ranking functions. However, it is not clear that these frameworks can help. I have not assumed a probabilistic understanding of *beliefs*, but only of evidence. Thus, these frameworks would need to provide a different account of evidence. What is needed is an alternative account of evidence. Sets of probability measures, Dempster-Shafer belief functions, and plausibility measures are all generalizations of probability theory. Thus, whatever explication of evidence one formulated in terms of those kinds of representations would presumably reduce to one of the probabilistic explications discussed in Section 2 in cases with the same formal structure as those described in Section 4 in which the likelihoods are perfectly objective and determinate. They would thus either run afoul of one of the three requirements discussed in Section 2 or encounter the same problems as the Law of Likelihood. Possibility measures and ranking functions are not straightforward generalizations of probability theory, so it might be worth investigating the possibility of using them to formulate evidence proportionalism. A discussion of those frameworks would take us too far afield here, so I leave it as a possible direction for future research.

One possible alternative to evidence proportionalism is the view that one should *update* one's beliefs in a way that is "proportioned" to one's evidence, but that one's "initial" beliefs must be fixed in some other way. Call this view "dynamic evidence proportionalism." Bayesian epistemology is naturally interpreted in this way. It even makes it possible to interpret "proportioned" literally: if you interpret the degree to which one changes one's beliefs in H_1 against H_2 in response to E as the ratio of one's posterior odds $\Pr(H_1|E)/\Pr(H_2|E)$ to one's prior odds $\Pr(H_1)/\Pr(H_2)$ for those hypotheses, then Bayes's rule entails that this degree should exactly equal one's likelihood ratio $\Pr(E|H_1)/\Pr(E|H_2)$.

The standard main objection to Bayesian epistemology is also an objection to dynamic evidence proportionalism: without supplementary constraints on initial beliefs, it does very little to constrain the beliefs that are permitted in light of a given body of data. Many responses to this objection have been given in the Bayesian literature. I will not discuss them here, except to note that some of them are broadly "evidentialist" in spirit and thus might appeal to those who find evidence proportionalism attractive. For instance, "reference Bayesians" endorse the use of "modest" initial probability distributions that maximize a measure of the degree to which the evidence from some prospective experiment is expected to influence the resulting posterior beliefs (Bernardo, 1985).

One could try to rescue evidence proportionalism within a broadly Bayesian framework by claiming that conforming one's initial beliefs to the prior probability distributions that are provided by the MaxEnt framework or some other objective Bayesian approach constitutes proportioning one's beliefs to one's evidence when one's evidence is the empty set. However, the result of updating such a distribution by Bayesian conditioning will not in general constitute believing in proportion to one's evidence on any notion of evidence that conforms to the requirements discussed in Section 2. One could say that proportioning one's beliefs to one's evidence just means believing in accordance with the recommendations of the relevant objective Bayesian framework, but only by discarding the idea that evidence proportionalism is making a substantial claim about the relationship between rational belief and evidence in some intuitive sense of "evidence."

Another possible alternative view is that what really matters is not proportioning one's beliefs to one's evidence, but forming one's beliefs in reliable ways. "Reliabilist" views of this kind have been influential in epistemology (Goldman, 1986), statistics (Neyman and Pearson, 1933; Neyman, 1957), and philosophy of science (Hacking, 1980; Giere, 1997; Kelly, 1995). They face objections, but along with dynamic proportionalist views, they are sufficiently plausible to undermine the idea that evidence proportionalism is so pretheoretically compelling that we must defend it at all costs.

6 Conclusion

The evidence proportionalist claim that one should proportion one's beliefs to one's evidence has a great deal of intuitive appeal. However, it is incompatible with any plausible accounts of evidence that is formulated in terms of probability theory, and we currently lack plausible and precise non-probabistic accounts. It is possible that such an account will eventually be found, but the existence of plausible alternatives to evidence proportionalism indicates that its ultimate vindication should not be taken for granted.

References

Achinstein, Peter. 2001. *The Book of Evidence.* Oxford Studies in the Philosophy of Science. Oxford University Press, USA.

Armitage, Peter. 1961. "Contribution to "Consistency in Statistical Inference and Decision"." *Journal of the Royal Statistical Society. Series B (Methodological)* 23:30–1.

Berger, James and Wolpert, Robert. 1988. *The Likelihood Principle*, volume 6 of *Lecture Notes—Monograph Series.* Beachwood, OH: Institute of Mathematical Statistics, 2nd edition.

Bernardo, J.M. (ed.). 1985. *Bayesian statistics 2: proceedings of the Second Valencia International Meeting, September 6/10, 1983.* Amsterdam: North-Holland.

Birnbaum, Allan. 1962. "On the Foundations of Statistical Inference." *Journal of the American Statistical Association* 57:269–306.

—. 1964. "The Anomalous Concept of Statistical Evidence: Axioms, Interpretations, and Elementary Exposition." Technical Report IMM-NYU 332, New York University Courant Institute of Mathematical Sciences.

Chandler, Jake. 2013. "Contrastive Confirmation: Some Competing Accounts." *Synthese* 190:129–38.

Conee, Earl and Feldman, Richard. 2008. "Evidence." In Quentin Smith (ed.), *Epistemology: New Essays*. Oxford University Press.

Crupi, Vincenzo. 2015. "Confirmation." In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Fall 2015 edition.

Dawid, A. P. 2014. "Discussion of On the Birnbaum Argument for the Strong Likelihood Principle." *Statist. Sci.* 29:240–241.

Engel, P. 2000. *Believing and Accepting*. Philosophical Studies Series. Springer Netherlands.

Evans, Michael, et al. 1986. "On Principles and Arguments to Likelihood." *Canadian Journal of Statistics* 14:181–94.

Feldman, Richard and Conee, Earl. 1985. "Evidentialism." *Philosophical Studies* 48:15–34.

Fitelson, Branden. 1999. "The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity." *Philosophy of Science* 66:S362–S378.

—. 2001. *Studies in Bayesian Confirmation Theory*. Ph.D. thesis, University of Wisconsin.

—. 2007. "Likelihoodism, Bayesianism, and Relational Confirmation." *Synthese* 156:473–89.

—. 2013. "Contrastive Bayesianism." In Martijn Blaauw (ed.), *Contrastivism in Philosophy*, Routledge Studies in Contemporary Philosophy, chapter 3, 64–87. Routledge.

Forster, Malcolm. 2006. "Counterexamples to a Likelihood Theory of Evidence." *Minds and Machines* 16:319–338.

Fraser, DAS, et al. 1985. "Marginalization, likelihood and structured models." In *Multivariate analysis–VI: proceedings of the Sixth International Symposium on Multivariate Analysis*, volume 6, 209. North Holland.

Gandenberger, Greg. 2015. "A New Proof of the Likelihood Principle." *The British Journal for the Philosophy of Science* 66:475–503.

—. MS. "New Responses to Three Counterexamples to the Likelihood Principle."

Giere, Ronald N. 1997. "Scientific Inference: Two Points of View." *Philosophy of Science* 64:pp. S180–S184.

Glymour, Clark. 2015. "Probability and the Explanatory Virtues." *The British Journal for the Philosophy of Science* 66:591–604.

Goldman, A.I. 1986. *Epistemology and Cognition.* Harvard University Press.

Goldman, Alvin. 2002. ""The Unity of the Epistemic Virtues"." In *Pathways to Knowledge*, 51–72. New York: Oxford University Press.

Greaves, Hilary. 2013. "Epistemic Decision Theory." *Mind* 122:915–952.

Hacking, Ian. 1980. "The Theory of Probable Inference: Neyman, Pearson, and Braithwaite." In *Science, Belief and Behaviour: Essays in Honour of R. B. Braithwaite*, 141–60. Cambridge University Press.

Hill, Bruce. 1988. *The Likelihood Principle*, volume 6 of *Lecture Notes—Monograph Series*, chapter Discussion by Bruce M. Hill, 161–74.4. Beachwood, OH: Institute of Mathematical Statistics, 2nd edition.

Hume, David. 1999. *Enquiries Concerning the Human Understanding.* Oxford University Press. Cited by section and paragraph, and including page references to the third Selby-Bigge edition, as revised by P.H. Nidditch (Clarendon Press, 1975).

Kelly, K.T. 1995. *KELLY LOGIC OF RELIABLE INQUIRY P.* Logic and Computation in Philosophy. Oxford University Press, USA.

Kelly, Thomas. 2014. "Evidence." In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Fall 2014 edition.

Kuhn, Thomas. 1977. *The Essential Tension: Selected Studies in Scientific Tradition and Change.* University of Chicago Press.

Leeds, Stephen. 2004. "Other Minds, Support, and Likelihoods." Unpublished manuscript. Available at <http://philsci-archive.pitt.edu/5472/>.

Lipton, Peter. 2001. "Inference to the Best Explanation." In W.H. Newton-Smith (ed.), *A Companion to the Philosophy of Science*, Blackwell Companions to Philosophy. Wiley.

Mayo, Deborah. 1996. *Error and the Growth of Experimental Knowledge*. University of Chicago Press.

—. 2014. "On the Birnbaum Argument for the Strong Likelihood Principle." *Statist. Sci.* 29:227–39.

McCain, Kevin. 2013. "Explanationist Evidentialism." *Episteme* 10:299–315.

Milne, Peter. 1996. "log[P(h/eb)/P(h/b)] Is the One True Measure of Confirmation." *Philosophy of Science* 63:21–26.

Neyman, Jerzy. 1957. ""Inductive Behavior" as a Basic Concept of Philosophy of Science." *Review of the International Statistical Institute* 25:7–22.

Neyman, Jerzy and Pearson, E. S. 1933. "On the Problem of the Most Efficient Tests of Statistical Hypotheses." *Philosophical Transactions of the Royal Society of London, Series A* 231:289–337.

NOAA Center for Weather and Climate Prediction. Apr. 26, 2012. "Frequently Asked Questions About El Niño and La Niña." `http://www.cpc.noaa.gov/products/analysis_monitoring/ensostuff/ensofaq.shtml#HURRICANES`. Accessed: Aug. 26, 2015.

Raiffa, H. and Schlaifer, R. 1961. *Applied Statistical Decision Theory*. Studies in managerial economics. Division of Research, Graduate School of Business Adminitration, Harvard University.

Roush, S. 2007. *Tracking Truth: Knowledge, Evidence, and Science*. Clarendon Press.

Royall, Richard. 1997. *Statistical Evidence: A Likelihood Paradigm*. London: Chapman & Hall.

Schupbach, Jonah N. and Sprenger, Jan. 2011. "The Logic of Explanatory Power." *Philosophy of Science* 78:105–127.

Seidenfeld, Teddy. 1985. "Comments on 'Weight of Evidence: A Brief Survey'." In *Bayesian Statistics 2: Proceedings of the Second Valencia International Meeting, September 6/10, 1983*, 264–6. Amsterdam: North-Holland.

Sober, Elliott. 1983. "Parsimony in Systematics: Philosophical Issues." *Annual Review of Ecology and Systematics* 14:335–357.

—. 2005. "Is Drift a Serious Alternative to Natural Selection as an Explanation of Complex Adaptive Traits?" *Royal Institute of Philosophy Supplements* 56:10–11.

—. 2008. *Evidence and Evolution: The Logic Behind the Science.* Cambridge University Press.

Sprenger, Jan. 2015. "Two Impossibility Results for Measures of Corroboration."

Stein, Charles. 1962. "A Remark on the Likelihood Principle." *Journal of the Royal Statistical Society. Series A (General)* 125:565–568.

Stone, Mervyn. 1976. "Strong Inconsistency from Uniform Priors." *Journal of the American Statistical Association* 71:114–116.

Tal, Eyal and Comesaña, Juan. 2015. "Is Evidence of Evidence Evidence?" *Noûs,* early view. Doi: 10.1111/nous.12101.

van Fraassen, BasC. 1995. "Belief and the problem of Ulysses and the sirens." *Philosophical Studies* 77:7–37.

A Proof from Section 4

Theorem 1. For any d, there is a possible joint probability distribution over an E, H_1, and H_2, where H_1 entails H_2, such that $\Pr(E|H_1)/\Pr(E|\neg H_1), \Pr(E|\neg H_2)/\Pr(E|H_2) > d$.[18]

[18] Seidenfeld (1985) gives an example of this phenomenon, claims that it is problematic for the use of likelihoodist characterizations of data as evidence in fixing beliefs, and suggests that it can be generalized. Here I provide the relevant generalization.

El Niño 2015	El Niño 2017	Hurricanes 2015	Probability
T	T	T	(2d-1)c
T	T	F	c/2
T	F	T	(2d-1)c/2
T	F	F	c/4
F	T	T	2dc
F	T	F	2d(4d-2)c
F	F	T	c/4
F	F	F	(2d-1)c/2

Table 1: A joint probability distribution over possible states that yields the result $\Pr(R|H_1)/\Pr(R|\neg H_1), \Pr(R|\neg H_2)/\Pr(R|H_2) > d$, where R indicates greater than normal hurricane activity in 2015, H_1 indicates El Niño in 2015 and 2017, and H_2 indicates el Niño in 2017.

This result is significant because (1) MCP requires a threshold d such that one will believe (to some degree) H_1 over $\neg H_1$ if E is one's total relevant evidence and $\Pr(E|H_1)/\Pr(E|\neg H_1) > d$, and likewise for $\neg H_2$ and H_2; and (2) MDP requires thresholds d_1 and d_2 such that one will proceed as if one believed H_1 rather than $\neg H_1$ if E is one's total relevant evidence and $\Pr(E|H_1)/\Pr(E|\neg H_1) > d_1$, and likewise for $\neg H_2$, H_2, and d_2. Thus, in the situation described in the theorem, learning E as one's total evidence would lead an MCPer to believe (to some degree) H_1 over $\neg H_1$ and the $\neg H_2$ over H_2; and in the case $d = \max(d_1, d_2)$, it would lead an MDPer to proceed as he or she believed H_1 over its $\neg H_1$ and $\neg H_2$ over H_2. However, these patterns of belief and action make no sense when H_1 entails H_2.

I will refer to the example in the main text, in which H_1 asserts that an El Niño occurs in both 2015 and 2017, H_2 asserts that one occurs in 2017, and E asserts that there is greater than normal hurricane activity in the East Pacific in 2015 (a likely consequence of an El Niǹo in 2015).

The proof is by construction. Fix a likelihood-ratio threshold for (some degree of) relative belief d, in accordance with MCP. Let $c = 1/[8d^2 + 2d - 1]$, and assign probabilities to states according to table 1. For ease of calculation, table 2 provides the marginal probabilities that this distribution implies, which are found simply by summing the corresponding probabilities in table 1 over the values of the omitted variable(s).

I will first confirm that table 1 gives a probability distribution. Because the states

El Niño 2015	El Niño 2017	Hurricanes 2015	Probability
T	T		$(4d-1)c/2$
T	F		$(4d-1)c/4$
F	T		$(8d^2-2d)c$
F	F		$(4d-1)c/4$
T		T	$3(2d-1)c/2$
T		F	$3c/4$
F		T	$(8d+1)c/4$
F		F	$(16d^2-6d-1)c/2$
	T	T	$(4d-1)c$
	T	F	$(16d^2-8d+1)c/2$
	F	T	$(4d-1)c/4$
	F	F	$(4d-1)c/4$
T			$3(4d-1)c/4$
F			$(32d^2-4d-1)c/4$
	T		$(16d^2-1)c/2$
	F		$(4d-1)c/2$
		T	$(20d-5)c/4$
		F	$(32d^2-12d+1)c/4$

Table 2: The marginal probabilities that are implied by the distribution given in table 1.

in the table are mutually exclusive and exhaustive, it is necessary and sufficient that the entries in the table be non-negative and sum to one. $d > 1$, so going down the table one finds $0 < (2d-1)c < 1/9$, $0 < c/2 < 1/18$, $0 < (2d-1)c/2 < 1/18$, $0 < c/4 < 1/36$, $0 < 2dc < 2/9$, and $4/9 < 2d(4d-2)c < 1$ for the first six rows; the last two rows are repeats. Thus, the entries in the table are all non-negative. One can show that they sum to one using elementary algebra. Therefore, the table gives a probability distribution.

Although it is not necessary for the theorem, for the sake of motivation I will next confirm that the distribution captures the following key features of the ENSO example: (1) $\Pr(EN7|EN5) < \Pr(EN7)$, where $EN5$ indicates El Niño in 2015 and $EN7$ indicates El Niño in 2017, and (2) $\Pr(R|EN5) > \Pr(R)$, where R indicates increased hurricane activity in 2015.

First,

$$\begin{aligned}
\Pr(EN7|EN5) &< \Pr(EN7) \\
\frac{\Pr(EN7 \ \&\ EN5)}{\Pr(EN5)} &< \Pr(EN7) \\
\frac{(4d-1)c/2}{3(4d-1)c/4} &< (16d^2-1)c/2 \\
\frac{2}{3} &< \frac{16d^2-1}{2(8d^2+2d-1)} \\
\frac{2}{3} &< \frac{(4d-1)(4d+1)}{2(4d-1)(2d+1)} \\
\frac{2}{3} &< \frac{4d+1}{4d+2} \\
8d+4 &< 12d+3 \\
1 &< 4d \\
d &> 1/4
\end{aligned}$$

$d > 1$, so the final inequality is satisfied.

Next,

$$\begin{aligned}
\Pr(R|EN5) &> \Pr(R) \\
\frac{\Pr(R \ \&\ EN5)}{EN5} &> \Pr(R) \\
\frac{3(2d-1)c/2}{3(4d-1)c/4} &> (20d-5)c/4 \\
\frac{4d-2}{4d-1} &> \frac{20d-5}{4(8d^2+2d-1)} \\
\frac{4d-2}{4d-1} &> \frac{20d-5}{2(4d-1)(4d+2)} \\
2(4d-2)(4d+2) &> 20d-5 \\
32d^2-8 &> 20d-5 \\
32d^2-20d-3 &> 0
\end{aligned}$$

Again, $d > 1$ ensures that the final inequality is satisfied.

Finally, I will show that $\Pr(E|H_1)/\Pr(E|\neg H_1), \Pr(E|\neg H_2)/\Pr(E|H_2) > d$ for $E = R$, $H_1 = EN5 \mathbin{\&} EN7$, and $H_2 = EN7$.

$$\begin{aligned}
\frac{\Pr(E|H_1)}{\Pr(E|\neg H_1)} &> d\\
\frac{\Pr(R|EN5 \mathbin{\&} EN7)}{\Pr(R|\neg(EN5 \mathbin{\&} EN7))} &> d\\
\frac{\Pr(R \mathbin{\&} EN5 \mathbin{\&} EN7)}{\Pr(EN5 \mathbin{\&} EN7)}\frac{\Pr(\neg(EN5 \mathbin{\&} EN7))}{\Pr(R \mathbin{\&} \neg(EN5 \mathbin{\&} EN7))} &> d\\
\frac{\Pr(EN5 \mathbin{\&} EN7 \mathbin{\&} R)}{\Pr(EN5 \mathbin{\&} EN7)}\frac{\Pr(\neg EN5) + \Pr(EN5 \mathbin{\&} \neg EN7)}{\Pr(\neg EN5 \mathbin{\&} R) + \Pr(EN5 \mathbin{\&} \neg EN7 \mathbin{\&} R)} &> d\\
\frac{(2d-1)c}{(4d-1)c/2}\frac{(32d^2-4d-1)c/4 + (4d-1)c/4}{(8d+1)c/4 + (2d-1)c/2} &> d\\
\frac{4d-2}{4d-1}\frac{32d^2-4d-1+4d-1}{8d+1+4d-2} &> d\\
\frac{4d-2}{4d-1}\frac{32d^2-2}{12d-1} &> d\\
\frac{128d^3-8d-64d^2+4}{48d^2-4d-12d+1} &> d\\
128d^3-64d^2-8d+4 &> 48d^3-16d^2+d\\
80d^3-48d^2-9d+4 &> 0
\end{aligned}$$

Again, $d > 1$ ensures that the final inequality is satisfied.

$$\begin{aligned}
\frac{\Pr(E|\neg H_2)}{\Pr(E|H_2)} &> d \\
\frac{\Pr(R|\neg EN7)}{\Pr(R|EN7)} &> d \\
\frac{\Pr(\neg EN7 \ \&\ R)}{\Pr(\neg EN7)}\frac{\Pr(EN7)}{\Pr(EN7 \ \&\ R)} &> d \\
\frac{(4d-1)c/4}{(4d-1)c/2}\frac{(16d^2-1)c/2}{(4d-1)c} &> d \\
\frac{1}{2}\frac{(4d-1)(4d+1)}{2(4d-1)} &> d \\
d + 1/4 &> d \\
1/4 &> 0
\end{aligned}$$

which is true for any d.