Why I Am Not a Methodological Likelihoodist*

By Greg Gandenberger

August 11, 2014

Note: This paper is a work in progress. Please do not cite it without the author's permission. Comments are welcome at greg@gandenberger.org.

Abstract

Methodological likelihoodism is the view that it is possible to provide an adequate self-contained methodology for science on the basis of likelihood functions alone. I argue that this view is false by (1) arguing that an adequate self-contained methodology for science would provide a good norm of commitment; (2) arguing that a good norm of commitment based on likelihood functions alone would have a particular form; (3) articulating minimal requirements for a good norm of commitment; and (4) showing that no norm of the specified form satisfies those requirements. Given the Likelihood Principle, it follows that it is impossible to provide an adequate self-contained methodology for science on the basis of the evidential meaning of data alone.

*Thanks to Marcus Adams, Jake Chandler, Foad Diazadji-Bahmani, Branden Fitelson, Clark Glymour, Christopher Hitchcock, Satish Iyengar, Michael Lew, Conor Mayo-Wilson, Edouard Machery, John Norton, Elay Schech, Teddy Seidenfeld, Elliott Sober, Jan Sprenger, and Jim Woodward for discussions that contributed to the development of this paper.

1 Introduction

One of the leading ideas in the philosophy of induction is that "saving the phenomena is a mark of truth" (Norton, 2005, 11). In other words, an observation supports a hypothesis to the extent that the hypothesis predicts it. The Likelihood Principle and the Law of Likelihood are precise explications of evidential equivalence and evidential favoring, respectively, that accord with this idea. They are of interest for the philosophy of scientific method because frequentist methods violate them, while Bayesian and likelihoodist methods do not.

Methodological likelihoodists go beyond merely accepting the Likelihood Principle and Law of Likelihood: they claim that it is possible to use the Law of Likelihood to provide an adequate self-contained methodology for science. Their approach is appealing in that it combines major advantages of Bayesian and frequentist methodologies. Like Bayesian and unlike frequentist approaches, it conforms to the Likelihood Principle. Like frequentist and unlike Bayesian approaches, it avoids the use of prior probabilities.

Many Bayesians reject methodological likelihoodism because they maintain that an adequate self-contained methodology for science would provide guidance for belief and action in the form of posterior probability distributions, which requires appealing to prior probability distributions as well as likelihood functions (e.g. Berger and Wolpert 1988, 124–36). The impact of this argument is limited by the fact that non-Bayesians reject the claim that providing a posterior probability distribution is the only or even the best way to provide guidance for belief and action. Frequentists, for instance, favor the use of methods that do not give posterior probability distributions but are purportedly justified by the fact that they are in some sense guaranteed to perform well in repeated applications in the indefinite long run. Some frequentists also insist that their methods are guides for "inductive behavior" rather than "inductive inference" (e.g. Neyman 1957), and thus are guides for belief only when "belief" is understood behavioristically, in terms of what one would do under various circumstances.

One could give a stronger argument against methodological likelihoodism by showing that no norm of commitment based on likelihood functions alone satisfies generally accepted requirements for such a norm, where "commitment" can be understood in either a cognitive or a behavioral sense. I aim to give such an argument in this paper. I do so by (1) arguing that an adequate self-contained methodology for science would provide a good norm of commitment; (2) arguing that a good norm of commitment based on likelihood functions alone would have a particular form; (3) articulating minimal requirements for a good norm of commitment; and (4) showing that no norm of the specified form satisfies those requirements.

This argument has potentially far-reaching implications. If the Likelihood Principle is true and methodological likelihoodism is false, as I have argued, then it is impossible to provide an adequate self-contained methodology for science on the basis of the evidential meaning of the data alone. Something else is needed: either additional inputs such as prior probabilities or an approach such as the frequentist one that fails to respect the evidential meaning of the data but purports to be justified on other grounds.

I characterize the methodological likelihoodist position more precisely in Section 2. I then give an overview of my argument against methodological likelihoodism in Section 3 and develop it in detail in Sections 4–8.

2 Methodological Likelihoodism

Methodological likelihoodists attempt to use the Law of Likelihood to provide a methodology for science that conforms to the Likelihood Principle but does not use prior probabilities. I will first provide precise statements of those principles and then explain how methodological likelihoodists use them.

The Likelihood Principle says, roughly, that the degree to which evidence counts in favor of a hypothesis depends *only* on the degree to which that hypothesis predicts it.

> **The Likelihood Principle.** Data E_1 and E_2 are evidentially equivalent with respect to the set of hypotheses $\mathbf{H}$ if and only if they have the same likelihood function on $\mathbf{H}$ up to a constant of proportionality—that is, if and only if there is a constant c such that $\Pr(E_1|H) = c\Pr(E_2|H)$ for all H in $\mathbf{H}$.

Note that the *likelihood function* for a set of hypotheses $\mathbf{H}$ on datum E is $\Pr(E|H)$[1] as a function of H as it varies over $\mathbf{H}$.

Whereas the Likelihood Principle tells us *that* evidential meaning depends only on likelihood functions, the Law of Likelihood tells us something about *how* it depends on likelihood functions. It does so in a *contrastive* way, addressing the question of when and to what degree an observation *favors one hypothesis over another*.

> **The Law of Likelihood.** Datum E_1 favors hypothesis H_1 over mutually exclusive[2] hypothesis H_2 if and only if the likelihood ratio $\mathcal{L} = \Pr(E|H_1)/\Pr(E|H_2)$ is greater than one, with $\mathcal{L}$ measuring the degree of favoring.

I have argued elsewhere that there are good reasons to accept the Likelihood Principle and the Law of Likelihood as explications of evidential equivalence and evidential favoring, respectively (2014, manuscript).

Methodological likelihoodism is the view that it is possible to provide an adequate self-contained methodology for the post-data analysis of experiments

[1] TK continuous cases

[2] See Gandenberger manuscript for an argument that the Law of Likelihood applies only the mutually exclusive hypotheses.

in science on the basis of likelihood functions alone.[3] This statement calls for some unpacking. Methodological likelihoodists take their approach to be "adequate" in the sense that it provides good answers to intrinsically important scientific questions and "self-contained" in the sense that it does so directly, as opposed to doing so merely by providing approximations to some other type of method or inputs for further processing. One could use the Law of Likelihood as the basis for an approach to pre-data experimental design, but my interest in this paper is in its use in the post-data analysis of experimental outcomes. I am using "experiment" here in a broad sense, to refer to any observational situation with a definite hypothesis space and a definite set of possible outcomes, including "observational studies" in which no intervention is being performed on the system or population of interest.

Methodological likelihoodists differ from Bayesians who endorse reporting likelihood functions and likelihood ratios only so that individual Bayesian agents can use them to update their own prior probability distributions (e.g. Good 1985). Methodological likelihoodists generally recognize that the information they provide could be used for updating purposes, but what is distinctive about their view is that they regard the outputs of the Law of Likelihood as being of interest in themselves, and not merely as tools for Bayesian updating.

Methodological likelihoodists can be pluralists: they need not maintain that likelihoodist methods are appropriate for *all* genuine scientific problems, but only that they are *sufficient* for *some* of them. Methodological likelihoodists typically say, for instance, that Bayesian methods are appropriate when empirically well-grounded prior probabilities are available (e.g. Sober 2008, 32). They

[3]Prominent methodological likelihoodists such as Edwards, Royall, and Sober do not characterize their view in this way, but they are committed to methodological likelihoodism as I characterize because they present the practice of reporting facts about evidential favoring as explicated by the Law of Likelihood as a good genuine alternative to Bayesian and frequentist methodologies. It would not be a good approach if it were not adequate in my sense, and it would not be a genuine alternative if it were not a self-contained methodology for the post-data analysis of experiments in science.

are not full Bayesians because they wish to conform to the Likelihood Principle without appealing to prior probability distributions that are based on nothing more than ungrounded opinions or contentious formal rules.

Likelihoodist methodology consists primarily of reporting degrees of evidential favoring in accordance with the Law of Likelihood.[4] The following example illustrates how this process works.

Example 1. The survival rate for infants infected with Virus A is currently 50%. It is hoped that a new drug will raise that rate to 75%. In the first clinical trial, nine of the first twelve patients treated with the new drug survive.

Let H_p be the hypothesis that the chance of survival for those who take the new drug is p. The Law of Likelihood says that the degree to which E favors $H_{75\%}$ over $H_{50\%}$ is $\Pr(E|H_{75\%})/\Pr(E|H_{50\%}) = 4.8$. Royall (2000) suggests using 8 as the cutoff for declaring a piece of evidence to be "fairly strong." By this standard, E favors $H_{75\%}$ over $H_{50\%}$, but not "fairly strongly." By contrast, E favors $H_{75\%}$ over $H_{25\%}$ to the very large degree 729.

3 An Overview of My Argument Against Methodological Likelihoodism

The Likelihood Principle and the Law of Likelihood have many virtues. They are intuitively plausible and have strong axiomatic bases (Gandenberger, 2014). They cohere well with Bayesian approaches and are useful for diagnosing what has going wrong when frequentist approaches yield intuitively unreasonable results (Berger and Wolpert, 1988, 65ff.). The likelihood functions on which they

[4] A likelihoodist might also report an entire likelihood function, to be interpreted in accordance with the Law of Likelihood. Their toolkit also includes several admittedly *ad hoc* techniques for dealing with "nuisance parameters," that is, parameters that affect the probability of the data but are not of interest to the experimenters (Royall, 1997, Ch. 7).

are based are often objective, and when they are not objective they are often easier to assess than prior probabilities (Sober, 2008). The Law of Likelihood is useful for resolving disputes about the significance of data for a pair of theories: Sober, for instance, uses it to good effect in discussing disputes about evidence that purportedly favors the theory of intelligent design over the theory of evolution or vice versa (2008).

What, then, could be wrong with the methodological likelihoodist position? Simply the idea that likelihood functions are sufficient for a methodology that can stand on its own. Using the Law of Likelihood to clarify the evidential meaning of data is fine, but the goals of science require that we then be able to say something about what we should believe or do in light the data. Purely likelihood-based methods do not address questions about what we should believe or do explicitly, and, I will argue, attempts to use them to answer such questions cannot succeed.

My argument against methodological likelihoodism begins in Section 4 with the claim that an adequate, self-contained methodology for the post-data analysis of experimental outcomes would provide a good norm of commitment. I deny that it is enough to provide characterizations of data as evidence without a way to use those characterizations to guide one's beliefs or actions. I then argue in Section 5 that if there are good norms of commitment based on likelihood functions alone, then they include a norm that directs one to accept one hypothesis over another to some function of the degree that one's total evidence favors the former over the latter, where that function satisfies a few minimal and intuitively reasonable constraints. In each of Sections 6–8 I show that any norm of this kind is incompatible with a minimal and intuitive requirements for a good norm of commitment. If one accepts the central claims of Section 4, Section 5, and *any of* Sections 6–8, then one is committed to rejecting methodological likelihoodism.

A few points of clarification are in order before I proceed. “Acceptance,” like “commitment,” is intended to be ambiguous among purely cognitive notions of belief, purely behavioral belief-like notions (such as dispositions to act in certain ways), and anything in between. Thus, my argument should be acceptable both to those who accept a notion of inductive inference and to those who reject it in favor of a notion of inductive behavior. I assume that degrees of acceptance form a continuum ranging from full acceptance of one hypothesis over another, through successively weaker degrees of acceptance in the same direction, to a state of neutrality, through successively stronger degrees of acceptance in the opposite direction, to full acceptance in that direction. “Preferring” one hypothesis over another means accepting the one over the other to some degree greater than the degree associated with neutrality.

I assume that there is no need to consider norms of action separately. This assumption is warranted if norms of action are reducible to norms of commitment because, for instance, commitments just are dispositions to act in certain ways, or if they are conceptually posterior to norms of commitment because, for instance, choosing the right action is a matter of appropriately integrating one’s beliefs and values (e.g. by maximizing expected utility) and one’s evidence bears only on the belief component of this process.

4 Claim 1: An adequate self-contained methodology for science provides a good norm of commitment

Bayesians provide a methodology for updating degrees of belief and using them to make decisions. Frequentists provide a methodology for evaluating hypotheses and making decisions in ways that are in some sense guaranteed to perform well in repeated applications with varying data in the indefinite long run. Likelihoodism is supposed to be a genuine alternative to these approaches, but it does not explicitly address questions about how we should evaluate hypotheses or what we should do. It only explicitly addresses questions about the evidential meaning of the data.

Methodological likelihoodists fully acknowledge this point (see e.g. Royall 1997, 3–4, Sober 2008, 32). They are explicitly advocating that we "change the subject" when prior probabilities are not available. But they fail to answer the following question, which they must answer in order to vindicate their approach: what value does a characterization of data as evidence have apart from the possibility of using it to guide one's commitments?

There are two kinds of answers that a methodological likelihoodist could give: (1) a characterization of data as evidence has value for some purpose beyond itself that does not require that it give guidance for one's commitments, or (2) a characterization of data as evidence has value in itself. Neither kind of answer has much plausibility.

No likelihoodist to my knowledge has given the first kind of answer. A characterization of data as evidence might have value for deciding which theory to use for purposes of explanation or prediction, but only insofar as it guides one's commitments with respect to those hypotheses, or at least with respect to claims about their suitability for those purposes. If a methodological likelihoodist wishes to claim that his or her characterizations of data as evidence have value for some purpose beyond themselves that does not involve using them to guide one's commitments, then the burden is on him or her to explain what that purpose is and how likelihoodist methods can be used to achieve it.

The main proponents of methodological likelihoodism such as Royall and

Sober seem to hold that a characterization of data as evidence has value in itself. That view is at least coherent: it is always possible simply to insist that something has intrinsic value, provided that it is not part of the concept of that thing that it lacks intrinsic value. But a characterization of data as evidence is not obviously valuable in itself, as many take happiness, virtue, and substantive knowledge to be. Nor do people generally act as if they regard characterizations of data as evidence as valuable in themselves. No one gathers evidence simply for the sake of gathering evidence. They gather evidence in order to inform their beliefs and/or decisions. It would be difficult to justify allocating time and tax dollars to a scientific research project that we expected to yield only characterizations of data as evidence and not act least approximate truth or predictive accuracy.

Likelihoodists have not argued that their characterizations of data as evidence are valuable for any purpose beyond themselves that does not require using them to guide commitments, and our behavior and attitudes seem to indicate that we do not regard them as valuable in themselves. For this reason, an adequate self-contained methodology for science would not stop at characterizing data as evidence, but would provide a good norm of commitment.

5 Claim 2: A good purely likelihood-based norm of commitment would fall within the scope of a generalization of Hume's dictum

5.1 Generalizing Hume's Dictum

I argued in the previous section that an adequate self-contained methodology for science would provide a good norm of commitment. If that claim is correct, then methodological likelihoodists need to maintain that it is possible to provide a good norm of commitment on the basis of likelihood functions alone. In this section I argue that such a norm would have to have a particular form. My goal is to restrict the set of possible norms under consideration enough to be able to prove things about that set without excluding any norm that might be worth considering.

An attractive starting point is Hume's dictum that a wise person proportions his or her belief to his or her (total) evidence (1825, 111, paraphrased). Likelihoodists provide only a contrastive measure of evidential favoring, so we need a variant on Hume's dictum which says that a wise person proportions his or her commitment in one proposition *relative to another* to the degree that his or her total evidence favors the one over the other.

This claim is attractive but too specific for present purposes. I would not want to rule out, for instance, being more cautious than it prescribes by remaining neutral when one's total evidence is non-neutral but weak. It does seem safe, on the other hand, to require that the relationship between evidential favoring and relative commitment be *nondecreasing*, so that an increase in favoring never leads to a decrease in relative commitment, and *neutrality-calibrated*, in the sense that neutral evidence leads to neutrality of commitment. It also seems safe to rule out as uninteresting norms that prescribe neutrality regardless of the evidence.

If we require proportioning relative acceptance to a *function of* the degree of evidential favoring that satisfies those requirements, then we get the following class of norms.

> **P**roportion **R**elative **A**cceptance to a **F**unction of the **E**vidence (**PRAFE**): Accept H_1 over H_2 to the degree $f(\mathcal{L}) = f(\Pr(E_T|H_1)/\Pr(E_T|H_2))$, where f is some nondecreasing function such that $f(1) = 1$ and

$f(a) > 1$ for some a, and E_T is one's total relevant evidence.

As a matter of convention, accepting H_1 over H_2 to degree 1 means being neutral between H_1 and H_2; to a degree greater than 1, preferring H_1 to H_2; and less than 1, preferring H_2 to H_1.[5]

This class of norms is very large. Because likelihoodists are committed to $\Pr(E_T|H_1)/\Pr(E_T|H_2)$ as their measure the degree to which one's total evidence favors H_1 over H_2, and the restrictions PRAFE places on f are well-motivated, PRAFE seems to include every purely likelihood-based norm that a methodological likelihoodist would want to consider.

5.2 Objection and Reply

Objection: A good norm of commitment could allow withholding Judgment

Perhaps the most questionable feature of PRAFE is that it does not allow for withholding judgment about a pair of hypotheses, as opposed to being neutral between them. One might think that if one has no evidence bearing on a pair of hypothesis, or only neutral evidence, then one should have no definite state of commitment at all, as opposed to a definite state of neutrality (Norton, 2008). Similarly, one might think that we should at least in some cases allow degrees of relative acceptance to be interval-valued rather than point-valued.[6]

[5] The logarithmic scale is perhaps more natural: if we let the degree of relative acceptance be $f(\log[\Pr(E_T|H_1)/\Pr(E_T|H_2)])$, then a degree of 0 means being neutral, a positive degree means preferring H_1 to H_2, and a negative degree means the opposite. Nothing substantive hangs on this choice of scales.

[6] The assumption that degrees of belief should always have sharp real values strikes many as implausibly strong. For instance, if I know only that the probability that a particular coin lands heads when flipped is between .25 and .75, inclusive, then perhaps the degree of belief I shoul dhave that it will land heads in a given flip is the interval $[.25, .75]$, rather than to any particular number in that interval. Similarly, one might think that degrees of relative acceptance should be interval-valued in some cases.

Reply

I am willing to grant the somewhat plausible idea that one should withhold judgment in the absence of relevant (non-neutral) evidence. I am open to the possibility that degrees of relative acceptance should be interval-valued, but I will not consider it in this paper. Allowing for interval-valued degrees would make methodological likelihoodism harder to refute, but hardly any more plausible. Imprecise Bayesians allow probability functions to be interval-valued, but in doing so they are generalizing a successful theory. By contrast, a methodological likelihoodist who found my arguments against PRAFE-rules (that is, rules of the form given by PRAFE) persuasive would be allowing for interval-valued degrees of relative acceptance in an attempt to rescue a failed theory. The burden is on methodological likelihoodists who think that this approach seems promising to develop it and argue that it works.

I will not grant that one should withhold evidence given multiple pieces of evidence that are individually non-neutral but collectively neutral. That proposal is not very attractive on its face, and it has unattractive consequences. For instance, suppose that E_1 favors H_1 over H_2 to some degree $c > 1$, while E_2 favors H_2 over H_1 to the same degree, where E_1 and E_2 are independent. Then the conjunction of E_1 and E_2 is neutral between H_1 and H_2. It is implausible that if one learns first E_1 and then E_2, with no other relevant evidence, then one should accept H_1 over H_2 after learning E_1 but have no definite opinion about H_1 and H_2 after learning E_2, rather than being neutral between them. After all, if E_2 had been stronger evidence by any degree ϵ as small as one likes, then one would not have withheld judgment, but would have accepted H_2 over H_1 to some degree no less than one. This discontinuity in the prescribed response to E_2 as its strength varies seems unnatural and ill-motivated.

I cannot see any other way in which PRAFE might be ruling out a purely likelihood-based norm of commitment that is worth considering. The burden is on anyone who would claim otherwise to provide an example and to show that the norm in question avoids the kinds of problems discussed below.

6 Claim 3: A good norm of commitment would allow one to prefer chance hypotheses to maximally likely alternatives

6.1 The Problem of Maximally Likely Hypotheses

Example 2. Suppose you take a deck of cards at random from a collection of standard decks and "anomalous" decks consisting of fifty-two copies of the same card, in unknown proportions. You shuffle the deck and flip over the top card. It is the two of clubs. What should your state of relative acceptance be with respect to the hypothesis H_s that the deck is standard and the hypothesis $H_{2\clubsuit}$ that it is a deck consisting of only twos of clubs?

The Law of Likelihood says that the observation E of the two of clubs favors $H_{2\clubsuit}$ over H_s to the degree $\Pr(E|H_{2\clubsuit})/\Pr(E|H_s) = \frac{1}{1/52} = 52$. We can stipulate that this observation is the only evidence one has about $H_{2\clubsuit}$ and H_s. Thus, any PRAFE–rule says to accept $H_{2\clubsuit}$ over H_s to the degree $f(52)$, where f is some nondecreasing function such that $f(1) = 1$ and $f(a) > 1$ for some $a > 1$.

Now, some PRAFE-rules would not direct one to prefer $H_{2\clubsuit}$ to H_s on a likelihood ratio of fifty-two: PRAFE permits "stingy" rules that have $f(52) = 0$. However, we can make the degree to which the observation of a two of hearts favors an analogue of $H_{2\clubsuit}$ over an analogue of H_s as large as we like simply by increasing the sizes of the decks. For instance, we could construct an analogous example using thousand-card decks (with four suits running 1–250, for instance) to generate a likelihood ratio of 1000. For any PRAFE-rule, we can, simply by making the decks large enough, produce an example analogous to Example 2 in which that rule would lead one to prefer the analogue of $H_{2\clubsuit}$ for that deck size to the analogue of H_s.

Of course, there is nothing special about the two of clubs. What makes Example 2 worrisome is that it illustrates the point that, for any PRAFE–rule, we can produce an example like it that will lead us to prefer *some* anomalous-deck hypothesis over H_s *no matter what card appears*. More generally, the Law of Likelihood will always say that the evidence favors the "maximally likely" hypothesis H_{ML} that the data had to turn out just as it did (so that $\Pr(E|H_{ML}) = 1$) over any hypothesis H_{CH} that makes the outcome a matter of chance (so that $\Pr(E|H_{CH}) < 1$). For any PRAFE–rule, if the probability that the chance hypothesis ascribes to the data is sufficiently low (because there are very many equally probable outcomes on that hypothesis, for instance), then that rule will direct one prefer the maximally likely hypothesis to the chance hypothesis no matter what the data may be.

A good norm of commitment would at least permit one to prefer chance hypotheses to maximally likely hypotheses in some cases. We have just seen that no PRAFE–rule does so. At best, such a rule allows one to be neutral between chance hypotheses and maximally likely hypotheses in some cases but forces one to prefer maximally likely hypotheses to chance hypotheses in others. Therefore, no PRAFE–rule is a good norm of commitment.

This Problem of Maximally Likelihood Hypotheses is particularly problematic when genuine indeterminism is a live option, as in quantum mechanics. The standard interpretation of quantum mechanics says that many events, such as radioactive decay, are "genuinely chancy." According to that interpretation, the laws of physics simply do not determine whether or not a radioactive atom will decay in a given time interval; they only give that event a probability. The Law of Likelihood would say that our entire body of data on radioactive decay events favors the hypothesis that those events had to occur exactly as they did over the hypothesis that they were genuinely chance to an enormous degree. Yet scientists do not accept the former over the latter. Thus, PRAFE–rules are seriously at odds with a highly successful and seemingly sensible scientific practice.

Some commenters argue that the Problem of Maximally Likely Hypotheses is a fatal objection to the Law of Likelihood itself (Mayo, 1996, 200–3). Advocates of the Law of Likelihood defend it by pointing out that (1) the Law of Likelihood faithfully reports changes in odds under Bayesian conditioning, and (2) the Law of Likelihood does *not* say that the evidence favors the hypothesis that the data-generating mechanism is deterministic (as opposed to some *particular* deterministic hypothesis such as $H_{2\clubsuit}$) over the hypothesis that its outputs are genuinely chancy. I contend that the debate on this point is moot for the purposes of this paper. If the critics of the Law of Likelihood are right, then so much the worse for methodological likelihoodism. If its defenders are right, then methodological likelihoodism still has a problem: responses to the Problem of Maximally Likely Hypotheses as an objection to the Law of Likelihood do not work as responses to it as an objection to rules of the form given by PRAFE.

Let me explain. It will be helpful to see how a Bayesian would handle Example 2. He or she has to put a prior probability distribution over H_s and the set of fifty-two anomalous-deck hypotheses corresponding to the fifty-two card types in a standard deck. Suppose for simplicity that he or she puts equal probability on each of the anomalous-deck hypotheses, so that his or her prior distribution puts some probability p on H_s and $(1-p)/52$ on each anomalous-deck hypothesis. In general, the odds for a pair of hypotheses after Bayesian conditioning is equal to the prior odds times the log-likelihood ratio:

$$\frac{\Pr(H_1|E)}{\Pr(H_2|E)} = \frac{\Pr(H_1)}{\Pr(H_2)}\frac{\Pr(E|H_1)}{\Pr(E|H_2)}$$

In this case, we have

$$\begin{aligned}\frac{\Pr(H_{2\clubsuit}|E)}{\Pr(H_s|E)} &= \frac{\Pr(H_{2\clubsuit})}{\Pr(H_s)}\frac{\Pr(E|H_{2\clubsuit})}{\Pr(E|H_s)} \\ &= \frac{(1-p)/52}{p} + (52)\end{aligned}$$

What happens for a Bayesian in this case is that the data rule out fifty-one of the fifty-two anomalous-deck hypotheses, and the prior probability on those hypotheses gets shifted to the one non-refuted anomalous-deck hypothesis. The probability of H_s does not change. Thus, the probability of $H_{2\clubsuit}$ relative to H_s increases substantially. However, the probability of the hypothesis H_a that some anomalous-deck hypothesis or other is true does not change at all.

Viewing the Law of Likelihood against this Bayesian backdrop helps explain what it is doing and why. It reports the factor by which Bayesian conditioning on the relevant evidence would increase the odds of one hypothesis against another if the prior probabilities were available. It does so even in cases like Example 2 that exhibit the Problem of Maximally Likely Hypotheses, which is a plausible reason to regard it as an appropriate measure of evidential favoring even in those cases.

For the purposes of this paper it does not matter whether or not one finds this response persuasive. What matters is that it does not work as a defense of PRAFE even if it does work as a defense of the Law of Likelihood.

Return to the point (1) that the Law of Likelihood faithfully reports changes in log-odds under Bayesian conditioning. This point does not help in defending PRAFE because PRAFE concerns one's *commitments themselves,* rather than *changes* in one's commitments. One could try to address this issue by saying that the state of commitment that PRAFE requires is the state at which one would arrive by starting in a neutral state and changing one's commitments in accordance with (a function of) the Law of Likelihood. PRAFE does effectively work in this way, but the fact that it thereby prevents one from ever preferring chance hypotheses to maximally likely alternatives in the cases to which it applies is a good reason to question the idea that this way of operating is a good one.

Next return to the point (2) that the Law of Likelihood does *not* say that the evidence favors the composite hypothesis that the data-generating mechanism is deterministic over the hypothesis that its outputs are genuinely chancy. In fact, it does not say anything definite about that pair of hypotheses because the hypotheses that the deck is anomalous does not entail probabilities for the possible observations. It is what statisticians call a *composite* hypotheses: a *disjunction* of simple statistical hypotheses each of which entails a different probability distribution for the possible observations. Composite hypotheses give rise to conditional probabilities for the possible observations only given a prior probability distribution over their disjuncts.[7]

For the Bayesian who assigns equal probabilities to the anomalous-deck hypotheses, the Law of Likelihood says that any outome of a single draw is evidentially *neutral* between H_a and H_s. In the same way, the evidence about radioactive decay does *not* favor the hypothesis that it is a deterministic hypothesis over the hypothesis that it is genuinely chancy, even though it favors the much more specific hypothesis that the data had to turn out exactly as it

[7] The likelihood $\Pr(E|H_1 \text{ or } H_2) = \Pr(E|H_1)\Pr(H_1|H_1 \text{ or } H_2) + \Pr(E|H_2)\Pr(H_2|H_1 \text{ or } H_2) = \Pr(E|H_1)\Pr(H_1)/\Pr(H_2) + \Pr(E|H_2)\Pr(H_2)/\Pr(H_1)$, which obviously depends on the prior probabilities $\Pr(H_1)$ and $\Pr(H_2)$.

did over the latter.

Again, one may or may not find this point persuasive as a defense of the Law of Likelihood. What matters for the purposes of this paper is that it does not work as a defense of PRAFE. Using it as such requires giving up the following closure principle: if you prefer H_1 to H_2, then you should prefer H_3 to H_2 for any logical consequence H_3 of H_1. PRAFE is at odds with this principle because it can lead one to accept a maximally likely hypothesis over a chance hypothesis when one should not accept a consequence of the maximally likely hypothesis over that chance hypothesis. For instance, it can lead one to accept $H_{2\clubsuit}$ over H_s when one should not accept H_a over H_s.

It is hard to see how one could get by without this closure principle. Giving it up means giving up the ability to do even very simple single-premise deductive inference with one's commitments. What kind of notion of acceptance does PRAFE yield if that notion does not support such inferences? There are of course problems with *multi-premise* closure principles, as illustrated by the lottery and preface paradoxes (Fumerton, 2009, 16–8), but one can use the probability calculus to explain why those problems arise: the probability of a conjunction is no greater than that of each conjunct and typically decreases as new conjuncts are added. By contrast, the probability calculus supports single-premise closure principles: if H_3 is a logical consequence of H_1, than $\Pr(H_3) \geq \Pr(H_1)$.

The Law of Likelihood can avoid this problem because an analogous closure principle for evidential favoring is not obviously compelling: such a principle would say that if E favors H_1 over H_3, then it favors H_2 over H_3 for any logical consequence H_2 of H_1. This principles is false if favoring one hypothesis over another entails increasing the odds of the former to the latter: we have seen that in Example 2, for instance, the data can increase the odds of $H_{2\clubsuit}$ against

H_s without increasing the odds of H_a (which follows from $H_{2\clubsuit}$) against H_s.

Because defenses of the Law of Likelihood against objections arising from maximally likely hypotheses do not work as defenses of PRAFE against analogous objections, it appeals that methodological likelihoodists have no response to the objection to PRAFE-rules presented in this section. Given that an adequate self-contained methodology for science would provide a good norm of commitment and that a good purely likelihood-based norm would have to be a PRAFE-rule, it follows that mtehodological likelihoodism is false.

6.2 Objections and Replies to Claim 3

Objection 1: A Pluralist Would Use a Bayesian Approach in These Cases

One might object that the Problem of Maximally Likely Hypotheses is only pressing because one often wants to say that the maximally likely hypothesis is antecedently implausible relative to the chance hypothesis. Methodological likelihoodists are typically pluralists. They would recognize cases in which the maximally likely hypothesis is antecedently implausible relative to the chance hypothesis as cases in which a likelihoodist approach is likely to yield a bad result, and would for that reason use a Bayesian or some other kind of approach

Reply to Objection 1

Methodological likelihoodists typically say that their approach is a fallback option to be used when empirically well-grounded prior probabilities are not available. In the kinds of cases under discussion, empirically well-grounded prior probabilities are not available (recall that in Example 2, for instance it was stipulated that the proportions of the relevant deck types are unknown), so this claim implies that a methodological likelihoodist approach is to be used.

The objection under discussion supposes instead that a methodological likelihoodist approach is to be used unless one of the hypotheses in question is antecedently implausible relative to the other. This approach sounds suspiciously like using likelihoodist methods if and only if they would yield essentially the same result as Bayesian methods. This approach is essentially Bayesianism in disguise. It might work fairly well, but it is not a distinctive alternative to Bayesianism.

Objection 2: Different Norms Are Appropriate for Different Hypotheses

Another possible response to Example 2 is to propose that different purely likelihood-based norms of commitment should be used for different pairs of hypotheses. For instance, one should require a higher degree of favoring for a given degree of acceptance when comparing a maximally likely hypothesis to a chance alternative than when comparing two chance alternatives because the maximally likely hypothesis is much more specific about what kinds of observations to expect.

Reply to Objection 2

This response is similar to the previous one, but it proposes modifying the methodological likelihoodist approach for different pairs of hypotheses rather than restricting it to certain kinds of pairs. It faces the same basic problem: to the extent that it solves the Problem of Maximally Likely Alternatives, it does so by bringing methodological likelihoodism into closer alignment with Bayesianism, thereby giving up the goal of providing a distinctive alternative to Bayesianism and frequentism.

Bayesianism is one way of using different purely likelihood-based norms of commitment for different pairs of hypotheses: it says to accept H_1 over H_2 to the degree $c_{12} \Pr(E|H_1)/ \Pr(E|H_2)$, where degrees of relative acceptance are understood as posterior odds and c_{12} is the prior odds of H_1 against H_2. While the approach is purely likelihood-based for each pair of hypotheses, it is of course not purely likelihood-based across pairs of hypotheses: it is based on a combination of likelihood functions and prior probabilities.

If the proposal under discussion is to use a methodological likelihoodist approach with $f(x) = x$ as long as the prior odds for the pair of hypotheses in question are not too different from one and to use a Bayesian approach otherwise, then it is just a kind of approximate Bayesianism in disguise. If it is anything else, then the methodological likelihoodist owes us an account of what it is and why it is preferable to that approach.

7 Claim 4: A good norm of commitment would allow preferences between logical opposites to be invariant under substitution of logical equivalents given one's evidence

7.1 A Problem Involving Logically Related Hypotheses

In this section I argue that no norm of commitment of the form given by PRAFE is a good one because any such norm can force one to violate the following rule:

> (4A) If H_1 and H_2 are logically equivalent given your evidence,[8] then do not prefer H_1 to $\neg H_1$ and $\neg H_2$ to H_2.

(4A) is intuitively compelling. After all, if H_1 and H_2 are logically equivalent given your total evidence, then your evidence entails that H_1 is true if and only

[8] H_1 and H_2 are logically equivalent given evidence E if and only if the conjunction of E_T and H_1 entails H_2 and the conjunction of E_T and H_2 entails H_1.

if H_2 is true. To take a particular case, (4A) prohibits someone who knows that no ravens are red from preferring "all ravens are either black or red" to its negation while dispreferring "all ravens are black" to its negation. For someone who knows that no ravens are red, those hypotheses have the same content and thus should be assessed alike.

Something like (4A) is often taken for granted in formal theories. It is a common trick in performing conditional probability calculations to replace A in an expression of the form $\Pr(A|C)$ with some B that is logically equivalent to A given C but not otherwise. For instance, one might replace $\Pr(\sum_{i=1}^{3} X_i = 3|X_1 = X_2 = 1)$ with $\Pr(X_3 = 1|X_1 = X_2 = 1)$, since $\sum_{i=1}^{3} X_i = 3$ is logically equivalent to $X_3 = 1$ given $X_1 = X_2 = 1$. Standard textbooks simply take for granted the permissibility of such maneuvers. Thus, if we were to identify the degree to which one accepts H_1 over H_2 with the odds $\Pr(H_1)/\Pr(H_2)$, then, we would have (4A) automatically. Likewise for any other formalization of relative acceptance that permits substitution of logical equivalents given one's evidence.

I prove that PRAFE can force one to violate (4A) in Appendix A. Here is roughly how the proof goes.[9] Let H_1 be the conjunction of some proposition A with E, and let H_2 be just the proposition A. Suppose that E is one's total relevant evidence with respect to H_1 and H_2. For instance, H_1 might be the proposition that Coins 1 and 2 both land heads, H_2 and A the proposition that Coin 2 lands heads, and E the proposition that Coin 1 lands heads. I show that in cases of this kind, for any constant a, one can construct a probability distribution over A and E such that the likelihood ratio of H_1 against $\neg H_1$ and $\neg H_2$ against H_2 both exceed a. Thus, given any norm of the form given by PRAFE there is a possible experimental outcome that would lead you to prefer H_1 over $\neg H_1$ and $\neg H_2$ over H_2. In this way, PRAFE can force you to violate (4A).

[9] The proof I give generalizes an argument given by Seidenfeld (1985).

7.2 Objections and Replies

Objection 1: H_1 and H_2 Are Not Mutually Exclusive

One obvious response to this argument is to restrict PRAFE or even the Law of Likelihood itself to hypotheses that are not logically related in the way that H_1 and H_2 in my argument are logically related. In fact, I have elsewhere argued on other grounds for restricting the Law of Likelihood to mutually exclusive hypotheses (manuscript, Section 3). H_1 and H_2 are not mutually exclusive, so one might think that this restriction would prevent the problem discussed here from arising.

Reply to Objection 1

In fact, restricting the Law of Likelihood to mutually exclusive hypotheses does not prevent the problem I have discussed here from arising. H_1 and H_2 in my argument are not mutually exclusive, but the Law of Likelihood is not applied to the comparison between H_1 and H_2: it is applied to the comparison between H_1 and its negation and to the comparison between H_2 and its negation.

It is presumably permissible to apply PRAFE to a hypothesis and its negation when they both have well-defined likelihood functions, as they do in the example I use. To prevent the problem discussed in this section from arising, one would have to place restrictions not on individual applications of PRAFE, but on combinations of applications of PRAFE. It would be difficult at best to find independent motivation for such restrictions. It would be even more difficult to abide by such restrictions because doing so would require individual scientists and perhaps even communities of scientists to keep track of which hypotheses they had considered in order to avoid asking questions that are logically related in ways to can lead to the kind of trouble discussed here.

Objection 2: PRAFE does not apply to hypotheses and their negations

The Law of Likelihood very often does not apply in a straightforward way to a hypothesis and its negation because at most one of those hypotheses entails a definite probability distribution for the data. For instance, if one supposes that a particular datum arises from a normal distribution with variance one, then the hypothesis that the mean of the distribution is zero (or any other definite value) entails a particular probability distribution over possible observations drawn from that distribution, while the negation of that hypothesis does not.

Reply to Objection 2

It is true that the Law of Likelihood very often does not apply in a straightforward way to a hypothesis and its negation. However, it does do so in some cases, and the case used to generate the violation of (4A) in this section is among them. There is nothing particularly strange about cases of this kind that would motivate a restriction on PRAFE or the Law of Likelihood that would exclude them.

It may be true that the problem discussed in this section would not often arise in practice because the kinds of cases in which it can arise are rare. However, they are sufficiently simple and straightforward example that the fact that rules of the form given by PRAFE fail to handle them in a sensible way suggests that PRAFE is deeply misguided.

Objection 3: H_1 and H_2 are not simple statistical hypotheses

There is something that distinguishes cases in which the hypotheses in question are logically related in the way that H_1 and H_2 are related from the kinds of examples that are commonly considered in likelihoodist writings: H_1 and

H_2 are not "statistical hypotheses," that is, hypotheses about the stochastic properties of a data-generating mechanism. My construction requires that at least one of them say something about the evidence, which would presumably be a datum produced by such a mechanism. Some likelihoodists do restrict the Law of Likelihood to statistical hypotheses (e.g. Hacking 1965, 59 and Edwards 1972, 57).

Reply to Objection 3

While some likelihoodists restrict the Law of Likelihood to statistical hypotheses, they seem to do so not for any principled reason but simply because they have statistical applications in mind. It is not clear that the restriction has any principled basis. Moreover, it has the unfortunate consequence of restricting the scope of the Law of Likelihood substantially. For instance, it would not allow one to apply the Law of Likelihood to high-level, substantive scientific theories, as Sober does with the theory of evolution and the theory of intelligent design (Sober, 2008). In addition, it does not address the other problems for PRAFE presented in this paper. Thus, restricting the Law of Likelihood to statistical hypotheses might allow one to avoid violations of (4A), but it does so in a way that is not independently well-motivated, has a high cost, and leaves other problems untouched.

8 Claim 5: A good norm of commitment would allow degrees of acceptance to increase under non-trivial disjunctions without violating invariance of preferences under substitution of logical equivalents

8.1 A Problem Arising From Diffrent Ways of Carving Up the Same Hypothesis Space

In this section I argue that no norm of the form given by (PRAFE) is a good one because no such norm allows one to abide by (5B) without in some cases violating (5A):

(5A) If H_1 is logically equivalent to H_3 and H_2 is logically equivalent to H_4, do not prefer H_1 to H_2 and H_4 to H_3

(5B) Accept (H_1 or H_2) over H_3 to a degree greater than that to which you accept H_1 over H_3 when the latter is well-defined, H_1 and H_2 are mutually exclusive, and you accept H_2 over H_3 to some degree greater than 0.

Recall that, by convention, accepting H_2 over H_3 to degree zero means fully rejecting H_2 relative to H_3, rather than being neutral between H_2 and H_3.

(5A) is slightly different from (4A) but just as compelling. It is stronger than (4A) in that the hypotheses to which it applies need not be logical opposites. However, the fact that the pairs of hypotheses to which (4A) applies are logical opposites played no role in my argument motivating (4A). (5A) is weaker than (4A) in that the relevant pairs of hypotheses do need to be logically equivalent full stop, rather than merely logically equivalent given one's evidence.

Violations of (5A) again seem unacceptable. H_1 is true if and only if H_3 is true, and likewise for H_2 and H_4, so one's assessment of H_1 in relation to H_2 should be the same as one's assessment of H_3 relative to H_4. For instance, (5A) says not to prefer "all ravens are black" to "some ravens are white" while at the same time preferring "some white things are ravens" (equivalent to "some ravens are white") to "all non-black things are non-ravens" (equivalent to "all ravens are black"). One's preference between a pair of propositions should depend on the content of those propositions rather than on the form in which those propositions are expressed.

(5A), like (4A), would hold automatically in a variety of possible formalizations of the notion of relative acceptance. It is typically assumed in probability theory, for instance, that logically equivalent propositions have the same probability. Thus, if one used a person's odds $\Pr(H_1)/\Pr(H_2)$ to formalize the degree to which he or she accepts H_1 over H_2, then (4A) would hold automatically. Likewise for any formalization that allows substitutions of logical equivalents.

Violations of rules like (4A) and (5A) might of course be *excusable.* We would not be inclined to criticize harshly a person of average mathematical ability who was attempting to assess the size of a cubic box for assigning different probabilities of the top of his or her head to the proposition that each side of the box is 27 inches long and the proposition that the box has volume $19,683$ in.^{3} even though those hypotheses are equivalent.[10] But of course the person in question has made a mistake, even if we would not be inclined to chastise him or her for it. Similarly, we would tolerate some violations of (4A) and (5A) in practice, but that does not mean that we should accept norms that can produce such violations.

(5B) is also compelling. If H_1 and H_2 are mutually exclusive, then the set of possible states of affairs in which (H_1 or H_2) holds is a strict superset of the

[10] I owe this example to Rescorla unpublished, 18–9

set of possible states of affairs in which H_1 holds. If one accepts H_2 over H_3 to some degree greater than 0, then one does not entirely dismiss the possibility that one of those additional states of affairs obtains. Thus, it seems that one should accept (H_1 or H_2) over H_3 to a degree greater than that to which one accepts H_1 over H_3.

Consider a particular case: (5B) directs one to accept "either all ravens are black or some are white and the rest are black" over "some ravens are red" to a degree greater than that to which one accepts "all ravens are black" over "some ravens are red," provided that the latter is defined and the degree to which one accepts "some ravens are white and the rest are black" over "some ravens are red" is greater than zero. This directive seems compelling. After all, "either all ravens are black or some are white and the rest are black" holds in all possible states of affairs in which "all ravens are black" holds and some in which it does not, so it makes sense to accept the former over some third claim to a greater degree than latter, provided that one takes at all seriously the set of possible states of affairs in which the former holds and the latter does not.

Like (4A) and (5A), (5B) would hold under a variety of possible formalizations of the notion of relative acceptance. For instance, if we again interpret the degree to which one accepts A over B as one's odds $\Pr(A)/\Pr(B)$, then (5B) follows from the axioms of probability. Probabilities obey finite additivity, meaning that $\Pr(H_1 \text{ or } H_2) = \Pr(H_1) + \Pr(H_2)$ when H_1 and H_2 are mutually exclusive. It follows that $\Pr(H_1 \text{ or } H_2)/\Pr(H_3) > \Pr(H_1)/\Pr(H_3)$ when $\Pr(H_2)/\Pr(H_3) > 0$ and H_1 and H_2 are mutually exclusive. An analogous argument would work under any similar interpretation that uses an additive (or superadditive)[11] calculus.

[11] A superadditive calculus f such as the Dempster-Shafer calculus (Dempster, 1968)) is one the axioms of which guarantee only that $f(H_1 \text{ or } H_2)$ is *greater than or equal to* $f(H_1)+f(H_2)$ when H_1 and H_2 are mutually exclusive.

The following example shows there are cases in which no PRAFE–rule allows one to abibe by (5B) without violating (5A).[12] (For the sake of readability, some details are given in Appendix B.)

Example 3. Suppose that a mad genius has mixed water and wine in a bottle. You know that the ratio r of water to wine is in the interval $(1/2, 2]$. The mad genius knows the value of r but refuses to tell it to you. He does agree to run three trials of an experiment each possible outcome of which yields relevant, non-neutral evidence concerning the hypotheses $H_1^r : r \in (1/2, 1]$, $H_2^r : r \in (1, 3/2]$ $H_3^r : r \in (3/2, 2]$.[13] As it turns out, the data from the three trials is collectively neutral with respect to those hypotheses. (A description of the experiment and its result are given in Appendix B.) PRAFE thus requires accepting H_1^r over H_3^r to degree one (i.e., being neutral between them). Consistency with (5B) thus requires consistency with accepting (H_1^r or H_2^r) over H_3^r to a degree greater than one.

Problems arise when we apply the same kind of reasoning to the ratio r' of wine to water. r' must be in the interval $[1/2, 2)$. The pieces of evidence from the mad genius's experiment are individually non-neutral between $H_1^{r'} : r' \in (1/2, 1]$ and $H_3^{r'} : r' \in (3/2, 2]$, but collectively neutral with respect to those hypotheses and $H_2^{r'} : r' \in (1, 3/2]$. PRAFE thus requires accepting $H_1^{r'}$ over $H_3^{r'}$ to degree one (i.e., being neutral between them). Consistency with (5B) thus requires consistency with accepting ($H_1^{r'}$ or $H_2^{r'}$) over $H_3^{r'}$ to a degree greater than one. However, H_1^r is logically equivalent to ($H_2^{r'}$ or $H_3^{r'}$) (r is between 1/2 and 1 if and only if $r' = 1/r$ is between 1 and 2), and $H_1^{r'}$ is logically equivalent to

[12]This example is a variant on a well-known example due to Von Mises (1981).

[13]PRAFE does not require that the intervals have the same size, but using intervals of different sizes would give rise to needless objections. For instance, one might think that one could rescue PRAFE by restricting to hypotheses that have "equal empirical content" as a kind of objective surrogate for prior probability. In order to do so, one would have to give an account of "equal empirical content" such that H_1^r, H_2^r, and H_3^r do not have equal empirical content despite being corresponding to intervals of the same size, or likewise for $H_1^{r'}$, $H_2^{r'}$, and $H_3^{r'}$, without ruling out so much that PRAFE becomes essentially useless.

(H_2^r or H_3^r) (r' is between 1/2 and 1 if and only if $r = 1/r'$ is between 1 and 2). Thus, (5A) rules out consistency with (5B) by ruling out preferring H_1^r to (H_2^r or H_3^r) and preferring $H_1^{r'}$ to ($H_2^{r'}$ or $H_3^{r'}$).

8.2 Objection and Reply

Objection: The Law of Likelihood Does Not Apply to Disjunctions

One might object that (5B) is inappropriate in this context because it refers to the degree to which one accepts a disjunction over another hypothesis, but PRAFE generally does not apply to disjunctions. If H_1 and H_2 entail different probabilities for E, then $\Pr(E|H_1 \text{ or } H_2)$ depends on the prior probabilities of H_1 and H_2. (See footnote 7.) Thus, likelihood ratios involving (H_1 or H_2) are not well defined when prior probabilities are not available, which is exactly when a typical pluralistic methodological likelihoodist would use the Law of Likelihood.

Reply

This objection is based on a misunderstanding of my argument. I do not claim that a good norm of commitment should *yield* a degree of acceptance for (H_1 or H_2) over H_3 that is greater than the degree of acceptance that it yields for H_1 over H_3 in the relevant circumstances. It could yield no degree of acceptance for the pair of hypotheses at all, as PRAFE typically does in the kinds of cases in which a pluralistic methodological likelihoodist would use the Law of Likelihood. I claim only that a good norm of commitment would *allow* a degree of acceptance for (H_1 or H_2) over H_3 that is greater than the degree of acceptance that it yields for H_1 over H_3, without generating violations of (5A). That claim seems highly plausible, and my argument shows that it implies that PRAFE–rules are not good norms of commitment, without applying

PRAFE–rules to disjunctions.

9 Conclusion

Methodological likelihoodists claim that it is possible to provide an adequate self-contained methodology for science on the basis of likelihood functions alone. However, the aims of science require a methodology that provides a good norm of commitment, and I have argued that any purely likelihood-based norm violates three minimal requirements for such a norm.

If methodological likelihoodism is false and the Likelihood Principle is true, as I argue elsewhere (Gandenberger 2014, Gandenberger manuscript), then it is impossible to provide an adequate, closed methodology for science based on the evidential meaning of the data alone. Something else is needed, such as prior probabilities or a non-evidential approach justfied by its long-run operating characteristics.

References

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

Dempster, AP. 1968. "A generalization of Bayesian inference (with discussion)." *Journal of the Royal Statistical Society Series B. v30 i2* 205–247.

Edwards, A.W.F. 1972. *Likelihood. An Account of the Statistical Concept of Likelihood and Its Application to Scientific Inference. [Mit Fig. U. Tab.]*. Cambridge University Press.

Fumerton, Richard. 2009. *Epistemology*. John Wiley & Sons.

Gandenberger, Greg. 2014. "A New Proof of the Likelihood Principle." *The British Journal for the Philosophy of Science* .

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

Good, I.J. 1985. "Weight of Evidence: A Brief Survey." In *Bayesian statistics 2: proceedings of the Second Valencia International Meeting, September 6/10, 1983*, 249–64. North-Holland.

Hacking, I. 1965. *Logic of Statistical Inference.* Cambridge University Press.

Hume, David. 1825. *Essays and Treatises on Several Subjects*, volume 2. Edinburgh: Bell and Bradfute.

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

Neyman, J. 1957. ""Inductive Behavior" as a Basic Concept of Philosophy of Science." *Revue de l'Institut International de Statistique / Review of the International Statistical Institute* 25:7–22.

Norton, John. 2005. "A Little Survey of Induction." In Peter Achinstein (ed.), *Scientific Evidence: Philosophical Theories and Applications*, 9–34. Johns Hopkins University Press.

—. 2008. "Ignorance and Indifference." *Philosophy of Science* 75:45–68.

Rescorla, Michael. unpublished. "Some Epistemological Ramifications of the Borel-Kolmogorov Paradox." .

Royall, Richard. 2000. "On the Probability of Observing Misleading Statistical Evidence." *Journal of the American Statistical Association* 95:pp. 760–768.

Royall, R.M. 1997. *Statistical Evidence: A Likelihood Paradigm.* Monographs on Statistics and Applied Probability. London: Chapman & Hall.

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. North-Holland.

Sober, Elliott. 2008. *Evidence and Evolution: The Logic Behind the Science.* Cambridge University Press.

Von Mises, R. 1981. *Probability, Statistics, and Truth.* Dover Books on Mathematics Series. Dover Publications.

A Proof that PRAFE can force one to violate (4A)

> (4A) If H_1 and H_2 are logically equivalent given your total evidence, then do not prefer H_1 to $\neg H_1$ and $\neg H_2$ to H_2.

Proof

Let X_1 and X_2 record the outcomes of independent coin flips. If the first coin lands heads, then $X_1 = 1$. Otherwise $X_1 = 0$. Likewise for the second coin and X_2. Let E be the evidence $X_1 = 1$, H_1 the hypothesis $X_1 = X_2 = 1$, and H_2 the hypothesis $X_2 = 1$. Suppose that E is the only information one has about X_1 and X_2.

Fix the function f such that (PRAFE) says to accept H_1 over H_2 to degree $f(\mathcal{L}) = f(\Pr(T|H_1)/\Pr(T|H_2))$. PRAFE requires that there be some constant a such that $f(a) > 1$ (and thus $f(x) > 1$ for all $x > a$, since f is nondecreasing). I need to show that there is a joint distribution for X_1 and X_2 such that $\Pr(E|H_1)/\Pr(E|\neg H_1) > a$ and $\Pr(E|\neg H_2)/\Pr(E|H_2) > a$.

I need to show how to construct for any a a joint distribution over X_1 and X_2 such that such that $\Pr(E|H_1)/\Pr(E| \sim H_1) > a$ and $\Pr(E| \sim H_2)/\Pr(E|H_2) > a$. From the fact that one can do so, it follows that PRAFE can force one to prefer H_1 to $\neg H_1$ and $\neg H_2$ to H_2. But given E, H_1 and H_2 are equivalent. Therefore, (PRAFE) forces one to violate (4A).

Here is the needed construction. Let a be some value greater than 1/2 of x such that $f(a) > 1$ for all $x > a$. Let $b = 2a/(2a+1)$. Then assign probabilities to outcomes according to the following table.

X_1 \ X_2	0	1	
0	$\frac{1-b}{4}$	b	$\frac{1+3b}{4}$
1	$\frac{1-b}{4}$	$\frac{1-b}{2}$	$\frac{3-3b}{4}$
	$\frac{1-b}{2}$	$\frac{1+b}{2}$	1

Table 1: Each of the four cells in the interior of the table give the probability that X_1 has the value for the given row and X_2 has the value for the given column. Thus, for instance, the bottom-right cell in the interior of the table says that the probability that $X_1 = X_2 = 1$ is $(1-b)/2$. The values all the way to the right give the probability that X_1 has the value associated with that row. The values all the way at the bottom give the probability that X_2 has the value associated with that column. The 1 in the bottom-right cell indicates that the row and column totals are both 1, as they should be.

It is easy to show that this table does indeed specify a probability distribution. Because $a > 1/2$, $b = 2a(2a+1)$ is strictly between 1/2 and 1. It follows that all of the probabilities in the table are between zero and one. In addition, the marginal probabilities (those shown in the leftmost column and the bottom row) are the sums of the relevant joint probabilities (those in the interior of the table), and the marginal probabilities sum to one. Thus, the distribution is additive, and the axioms of probability are satisfied.

I will now show that $\Pr(E|H_1)/\Pr(E|\neg H_1)$ and $\Pr(E|\neg H_2)/\Pr(E|H_2)$ are both greater than a, completing the proof.

$$
\begin{aligned}
\frac{\Pr(E|H_1)}{\Pr(E|\neg H_1)} &= \frac{\Pr(E \mathbin{\&} H_1)}{\Pr(H_1)} \frac{\Pr(\neg H_1)}{\Pr(E \mathbin{\&} \neg H_1)} \\
&= \frac{\Pr(X_1 = X_2 = 1)}{\Pr(X_1 = X_2 = 1)} \frac{\Pr(\neg(X_1 = X_2 = 1))}{\Pr(X_1 = 1 \mathbin{\&} X_2 = 0)} \\
&= \frac{1 - \Pr(X_1 = X_2 = 1)}{\Pr(X_1 = 1 \mathbin{\&} X_2 = 0)} \\
&= \frac{1 - (1-b)/2}{(1-b)/4} \\
&= \frac{4}{1-b} - 2 \\
&= \frac{4}{1 - (2a)/(2a+1)} - 2 \\
&= \frac{8a+4}{2a+1-2a} - 2 \\
&= 8a + 4 - 2 \\
&= 8a + 2 \\
&> a
\end{aligned}
$$

$$
\begin{aligned}
\frac{\Pr(E|\neg H_2)}{\Pr(E|H_2)} &= \frac{\Pr(E \mathbin{\&} \neg H_2)}{\Pr(\neg H_2)} \frac{\Pr(H_2)}{\Pr(E \mathbin{\&} H_2)} \\
&= \frac{\Pr(X_1 = 1 \mathbin{\&} X_2 = 0)}{\Pr(X_2 = 0)} \frac{\Pr(X_2 = 1)}{\Pr(X_1 = X_2 = 1)} \\
&= \frac{(1-b)/4}{(1-b)/2} \frac{(1+b)/2}{(1-b)/2} \\
&= \frac{1}{2} \frac{1+b}{1-b} \\
&= \frac{1}{2} \frac{1+2a/(2a+1)}{1-2a/(2a+1)} \\
&= \frac{1}{2} \frac{2a+1+2a)}{2a+1-2a} \\
&= \frac{1}{2}(4a+1) \\
&= 2a + \frac{1}{2} \\
&> a
\end{aligned}
$$

B Desciption of the Experiment Showing the Inconsistency of PRAFE with the Conjunction of (5A) and (5B)

The outline of a proof that PRAFE is inconsistent with the conjunction of (5A) and (5B) is given in the main text. What is missing is a description of an experiment that yields three pieces of evidence that are individually non-neutral but collectively neutral with respect to H_1^r, H_2^r, and H_3^r, individually non-neutral between $H_1^{r'}$ and $H_3^{r'}$, and collectively neutral with respect to $H_1^{r'}$ and $H_3^{r'}$. One experiment of this kind consists of rolling three times a three-sided die with weights that depend on r as shown in table 2.

Each possible outcome has a probability that varies with r, but the sequence

If r is in the interval...	... then Pr(1) =	... then Pr(2) =	... then Pr(3) =
$(1/2, 1]$	1/2	1/3	1/6
$(1, 3/2]$	1/6	1/2	1/3
$(3/2, 2]$	1/3	1/6	1/2

Table 2: The probabilities of the possible die roll outcomes as a function of the ratio of wine to water r

of outcomes $\{1, 2, 3\}$, for instance, has probability $1/2 \times 1/3 \times 1/6 = 1/36$ regardless of r. Thus, that sequence contains three pieces of evidence that are individually non-neutral but collectively neutral with respect to H_1^r, H_2^r, and H_3^r.

Because there is a one-to-one correspondence between values of r and values of r', we can also say that the die has weights that depend on r' in the following way.

If r' is in the interval...	... then Pr(1) =	... then Pr(2) =	... then Pr(3) =
$[1, 2)$	1/2	1/3	1/6
$[2/3, 1)$	1/6	1/2	1/3
$[1/2, 2/3)$	1/3	1/6	1/2

Table 3: The probabilities of the possible die roll outcomes as a function of the ratio of water to wine r'

If $H_1^{r'}$ is true (meaning that r' is in the interval $[1/2, 1)$), then r' is in either $[1/2, 2/3)$ or $[2/3, 1)$. Now, a likelihoodist cannot compute the likelihood $\Pr(E|H_1^{r'})$ on $E = 1$, $E = 2$, or $E = 3$ without assigning prior probabilities to $[1/2, 2/3)$ and $[2/3, 1)$ given $H_1^{r'}$. However, $\Pr(1|H_1^{r'})$ must be some weighted average of $\Pr(1|r' \in [1/2, 2/3)) = 1/3$ and $\Pr(1|r' \in [2/3, 1)) = 1/6$, and thus must be less than $\Pr(1|H_3^{r'}) = \Pr(1|r' \in [3/2, 2)) = 1/2$. Thus, a likelihoodist can say that $E = 1$ favors $H_3^{r'}$ over $H_1^{r'}$. By an analogous argument, a likelihoodist can say that $E = 3$ favors $H_1^{r'}$ over $H_3^{r'}$. It is indeterminate for a likelihoodist whether $E = 2$ favors $H_1^{r'}$ over $H_3^{r'}$ or not, but all I need is that the evidence be collectively neutral without all of the pieces of evidence being individually neutral. (If they were all neutral, then I have conceded that it may be acceptable for one to withhold judgment rather than to be neutral among $H_1^{r'}$, $H_2^{r'}$, and $H_3^{r'}$.) $\Pr(\{1,2,3\}|H_1^{r'})$ must be some weighted average of $\Pr(\{1,2,3\}|r' \in [1/2, 2/3)) = 1/36$ and $\Pr(\{1,2,3\}|r' \in [2/3, 1)) = 1/36$, so it must be $1/36$.

Now, if either $H_2^{r'}$ or $H_3^{r'}$ is true, then r' is in $[1, 2)$. Thus, all possible outcomes of the experiment are neutral between $H_2^{r'}$ and $H_3^{r'}$. In addition, $\Pr(\{1,2,3\}|H_3^{r'}) = 1/36$, so the evidence is collectively neutral between $H_1^{r'}$ and $H_3^{r'}$. This point completes the proof given in the main text.