Why I Am Not a Methodological Likelihoodist*

By Greg Gandenberger

March 14, 2014

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 methodological likelihoodism is false by arguing that an adequate self-contained methodology for science provides good norms of commitment vis-à-vis hypotheses, articulating minimal requirements for a norm of this kind, and proving that no purely likelihood-based norm satisfies those requirements.

Introduction

One of the guiding ideas in the philosophy of induction is that "saving the phenomena is a mark of truth" (Norton, 2005, 11). In other words, a hypothesis is confirmed to the extent that it correctly predicts what is observed; as Milne puts it, "prediction and confirmation are two sides of the same coin" (1996, 23).

Likelihoodism provides principles of evidential relevance and evidential favoring that accord with this idea. Its primary principle of evidential favoring is the *Likelihood Principle*, which says that the evidential meaning of a datum

*Thanks to Jake Chandler, Branden Fitelson, Clark Glymour, Satish Iyengar, Michael Lew, Edouard Machery, John Norton, Teddy Seidenfeld, and Jim Woodward for discussions that contributed to the development of this paper.

E for a set of hypotheses H depends only on how well each of those hypotheses predicts that datum—more precisely, on the *likelihood function* $\Pr(E|H)$[1] considered as a function of $H \in \mathbf{H}$, up to a constant of proportionality. Its primary principle of evidential favoring is the *Law of Likelihood*, according to which E favors H_1 over H_2 when and to the degree that the *log-likelihood ratio* $\mathcal{L} = \log[Pr(E|H_1)/Pr(E|H_2)]$ is greater than zero.[2]

Methodological likelihoodists such as Edwards (1972), Royall (1997), and Sober (2008) go beyond simply accepting the Likelihood Principle and the Law of Likelihood: they claim that it is possible to provide an adequate self-contained[3] methodology for science on the basis of likelihood functions alone. They aim to provide a methodology that combines the main advantages of Bayesian and frequentist methodologies without their respective disadvantages. Like Bayesian and unlike frequentist methods, likelihoodist methods conform to the Likelihood Principle, for which there are strong arguments (Gandenberger, forthcoming). Like frequentist and unlike Bayesian methods, likelihoodist methods avoid appeals to prior probabilities, which are often contentious.

[1] A more subtle account is needed when the sample space is continuous, so that $\Pr(E|H) = 0$ for a typical datum; see (Hacking, 1965, 57, 66–70), (Berger and Wolpert, 1988, 32–6), and (Pawitan, 2001, 23–4). In principle, this complication can be ignored in the context of any real experiment: real measurement techniques have finite precision, so real sample spaces are always discrete.

[2] The Law of Likelihood is often stated in terms of likelihood ratios rather than log-likelihood ratios. Nothing substantive hangs on this difference. Strictly speaking, the Law of Likelihood should be understood as the claim that evidential favoring increases monotonically with the likelihood ratio. Different monotonic functions of the likelihood ratio produce different permissible measurement scales. I use a logarithmic scale for ease of interpretation: zero indicates evidential neutrality; positive values indicate that the evidence favors H_1 over H_2 while negative values indicate the opposite; and the degree of evidential favoring provided by a pair of independent data is simply the sum of the degrees of evidential favoring provided by each datum individually. The base of the logarithms is immaterial.

[3] The claim that this methodology is self-contained is not meant to exclude methodological pluralism a la (Sober, 2008, 3, 356–8). Methodological likelihoodists need not believe that methods based on likelihood functions alone are appropriate for *all* scientific problems. However, they must believe that they are appropriate for *some* scientific problems, and not merely in the sense that they are appropriate when they would give the same answer as a reasonable Bayesian or frequentist method or in the sense that their outputs are useful as inputs for some other method, such as Bayesian updating. A common pluralist view that qualifies as a form of methodological likelihoodism is that Bayesian methods are appropriate when prior probabilities are "available" in some sense, while likelihoodist methods are appropriate when they are not, and that they are often unavailable in science (see e.g. Sober 2008, 32).

The purpose of this paper is to argue that methodological likelihoodism should nevertheless be rejected.

My argument against methodological likelihoodism rests on the following premises.

Premises

(1) An adequate self-contained methodology for science provides good norms of commitment vis-à-vis hypotheses.

(2) If there are good norms of commitment based on likelihood functions alone, then some rule of the following form is among them, where T is your total relevant evidence:[4]

> **P**roportion **R**elative **A**cceptance to (a **F**unction of) the **E**vidence (PRAFE): Accept H_1 over H_2 to the degree $f(\mathcal{L}) = f(\log[\Pr(T|H_1)/\Pr(T|H_2)])$, where f is some nondecreasing function such that $f(0) = 0$ and $f(a) > 0$ for some a.

(3) A good norm of commitment vis-à-vis hypotheses is compatible with the following rules:

(3A) Do not prefer H_1 to H_2 and H_3 to H_4 if H_1 is logically equivalent to H_4 and H_2 is logically equivalent to H_3 (where preferring one hypothesis to another is equivalent to accepting the former over the latter to a positive degree).

(3B) 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 following conditions are met:

i. H_1 and H_2 are mutually exclusive,

[4] The notion of relevant evidence can be formalized in terms of sufficient statistics (see Halmos and Savage 1949).

ii. the degree to which you accept H_1 over H_3 is well-defined, and

iii. the degree to which you accept H_2 over H_3 is well-defined and is not $-\infty$.

(4) A good norm of commitment vis-à-vis hypotheses is compatible with the following rule:

(4A) Do not prefer H_1 to $\sim H_1$ and $\sim H_2$ to H_2 if H_1 and H_2 are logically equivalent given your total evidence.

(3) and (4) each entails that no rule of the form given by (PRAFE) is a good norm of commitment vis-à-vis hypotheses. Thus, the conjunction of (1), (2), and either (3) or (4) entails that methodological likelihoodism is false.

I argue for (1)–(4) in Sections 1–4, respectively. I prove that (3) and (4) entail that no rule of the form given by (PRAFE) is a good norm of commitment vis-à-vis hypotheses in Appendices A and B, respectively. In Section 5 I respond to attempts to defend likelihoodist methods with reliabilist arguments.

1 Premise (1): An adequate self-contained methodology for science provides good norms of commitment vis-à-vis hypotheses

Science should help guide our commitments vis-à-vis hypotheses. It is not enough to say something about how the data are related to the hypotheses; we need to be able to "detach" the evidence and say something about the hypotheses themselves in light of the data. It would be odd for the author of a scientific paper to say that he or she does not care about evaluating the hypotheses he or she considers, but only wishes to assess how the data bear on them as evidence.[5]

Some methodological likelihoodists, such as Edwards,[6] at least suggest that purely likelihood-based methods can be used to guide our commitments vis-à-vis hypotheses. Others, such as Royal and Sober, are careful not to claim more for purely likelihood-based methods than that they correctly characterize data as evidence (e.g. Royall 1997, 3; Sober 2008, 32). But what are we to do with a characterization of data as evidence if not to use it to guide our commitments? And how are we to use it to guide our commitments without appealing to information not given by the likelihood function?

Royall and Sober provide no explicit answers to these questions. Royall takes the value of a correct characterization of data as evidence for granted.[7] Sober does not take it for granted, but he does not provide a clear argument for it either. He says that it is not enough to show that the Law of Likelihood "conforms to, and renders precise and systematic, our use of the informal concept" of evidential favoring: "what matters," he writes, "is whether [the Law of Likelihood] isolates an epistemologically important concept" (Sober, 2008, 35). I agree that it is not enough to vindicate methodological likelihoodism to show that the Law of Likelihood captures our informal concept of evidential favoring. But, depending on one's views about epistemological importance, it may also not be enough to show that the Law of Likelihood isolates an epistemologically important concept. The Law of Likelihood could be epistemologically important because, for instance, it is useful for explaining the so-called conjunction

[5]Of course, not every scientific paper should include an evaluation of the hypotheses considered therein. There may be relevant data from other sources, and the author of the paper may wish to leave the evaluation to his or her readers. These points are compatible with my claim that hypothesis evaluation is ultimately indispensable.

[6]Edwards states the Law of Likelihood in terms of support (1972, 30), but he describes it as providing the basis for a system of inference (e.g. 1972, 7) and relative degrees of belief (e.g. 1972, 28) without explanation or argument.

[7]Royall begins his (1997) with the bare assertion that the "most important task" of statistics "is to provide objective quantitative alternatives to personal judgement for interpreting the evidence produced by experiments and observational studies" (xi). I have found no argument for this claim in his writings.

fallacy.[8] It would not follow that the Law of Likelihood provides useful guidance for our commitments vis-à-vis hypotheses, as methodological likelihoodism requires.

Sober seems to take himself to show that the Law of Likelihood does isolate an epistemologically important concept. However, he does not explain how he takes himself to do so. It is a plausible guess that he takes himself to do so in his use of the Law of Likelihood to address seemingly well-formed and important question such as the following (Sober, 2008, 107–8):

- Are the imperfect adaptations that organisms exhibit evidence that they were not produced by an intelligent designer?

- Is the fact that bears in cold climates have longer fur than bears in warm climates evidence that fur length evolved by natural selection as an adaptive response to ambient temperature?

- Are the similarities that species exhibit evidence that they stem from a common ancestor?

The Law of Likelihood can indeed be used to provide defensible answers to these questions (see Gandenberger forthcoming, Gandenberger unpublished). However, the following questions remain: what are we to do with answers to these questions if not to use them to guide our commitments vis-à-vis the relevant hypotheses? And how are we to use them to provide such guidance without appealing to information not given by the likelihood function?

[8] As an example of the conjunction fallacy, most people give a higher probability to the statement that a character named Linda is a feminist bank teller then to the proposition that Linda is a bank teller. Because the population of feminist bank tellers is necessarily a subset of the population of bank tellers, these judgments are probabilistically incoherent. One possible explanation for this phenomenon is that people are responding to the fact that the vignette told about Linda favors the statement that she is a feminist bank teller over its negation (or confirms it) more than it favors the statement that she is a bank teller over its negation. See Tentori et al. (2013) for empirical evidence that seems to support this explanation.

Methodological likelihoodists have two options: (1) claim that science need not provide guidance for our commitments vis-à-vis hypotheses, or (2) provide good norms of commitment vis-à-vis hypotheses that are based on likelihood functions alone. Option (1) flies in the face of the commonsense idea that we do science in order to learn about the world (at least in some attenuated sense)[9] or to improve our ability to predict and control some part of it. It would be difficult to justify allocating time and tax dollars to science if all it could do were to generate data and hypotheses and tell us how that data is related to those hypotheses as evidence, without thereby giving us any guidance about what to believe or do. Traditionally, philosophers of science have sought a theory of evidence or confirmation so that they could use that theory to evaluate hypotheses in a principled way. The idea that characterizations of data as evidence are valuable in themselves is an unfortunate byproduct of this pursuit.

I take up option (2) in the next section.

2 Premise (2): If there are good purely likelihood-based norms of commitment, then (PRAFE) is among them

I argued in the previous section that making good on methodological likelihoodism requires providing a good purely likelihood-based norm of commitment.

A reasonable starting point for an attempt to provide such a norm is Hume's dictum that a wise person proportions his or her belief to his or her (total) evidence (1825, 111, paraphrased). We can increase the plausibility of this

[9]It is compatible with my claim in the section, for instance, that we do science only to learn approximate truths about observable phenomena, as some scientific anti-realists claim (e.g. Van Fraassen 1980).

already plausible dictum by generalizing it in two ways. First we can replace "belief" with "acceptance." The word "acceptance" here could be understood in a purely doxastic way, as indicating "what one holds in one's head," so to speak. Alternatively, it could be understood in a "behavioristic" way, as indicating something about what one would do in certain circumstances. I aim to show that my generalization of Hume's dictum has disastrous consequences on either interpretation, thereby casting doubt on the possibility of a good purely likelihood-based norm of commitment of either the doxastic or the behavioristic kind. I assume only that degrees of acceptance correspond to a qualitative preference ordering in the following way: accepting H_1 over H_2 to a positive degree indicates that one prefers H_1 to H_2, doing so to degree zero indicates that one has no preference between H_1 and H_2, and doing so to a negative degree indicates that one prefers H_2 to H_1.

We can generalize Hume's dictum in a second way by saying merely that a wise person proportions his or her acceptance to some function of the evidence, where that function satisfies the following mild constraints:

- it is **nondecreasing**, so that an increase in absolute value for evidential favoring without a change in sign[10] never leads to a decrease in degree of acceptance;
- it is **calibrated** in the sense that neutral evidence ($\log[\Pr(T|H_1)/\Pr(T|H_2) = 0$, i.e. $\Pr(T|H_1) = \Pr(T|H_2)$) leads to neutrality of acceptance (neither preferring H_1 to H_2 nor vice versa);
- it is **nontrivial** in the sense that it would lead one to accept H_1 over H_2 given sufficiently strong evidence favoring the former over the latter.

[10]The phrase "without a change in sign" is necessary because I do not assume that the function f is symmetric about zero. One might wish to add this assumption, but I do not need it.

Generalizing Hume's dictum in these two ways leads to the following class of purely likelihood-based norms of commitment vis-à-vis hypotheses, where T is one's total relevant evidence.

> **P**roportion **R**elative **A**cceptance to (a **F**unction of) the **E**vidence (PRAFE): Accept H_1 over H_2 to the degree $f(\mathcal{L}) = f(\log[\Pr(T|H_1)/\Pr(T|H_2)])$, where f is some nondecreasing function such that $f(0) = 0$ and $f(a) > 0$ for some a.

I cannot think of a more plausible yet nontrivial way to map the degrees of evidential favoring that the Law of Likelihood provides onto real-valued degrees of relative acceptance. There are no rival proposals in the literature to consider because methodological likelihoodists either deny that their methods provide guidance for belief or action (e.g. Royall and Sober) or suggest that they do provide such guidance but fail to provide a definite account (e.g. Edwards). If methodological likelihoodists wish to claim that there are good purely likelihood-based norms of commitment of a different form, then they need to say explicitly what those norms are and how they are good. In the meantime, (PRAFE)'s generality and intuitive plausibility warrant the claim that if there are good purely likelihood-based norms of commitment, then they include some norm of the form it provides.

Methodological likelihoodists may not be able to escape the difficulties for (PRAFE) that I present below even if there are purely likelihood-based norms of commitment more plausible than (PRAFE). I have argued for a particular kind of norm only because some constraints are necessary for proving definite results. But the problematic results for (PRAFE) that I present do not seem to depend on any quirk of (PRAFE) that could easily be removed, but rather from the fact that likelihood functions (unlike probability distributions) respect neither entailment relations among hypotheses nor logical equivalence among hypotheses given one's evidence. For that reason, it seems likely that any purely likelihood-based norm of commitment would suffer from similar problems.

3 Premise (3): A good norm of commitment is compatible with (3A) and (3B)

I have argued that an adequate self-contained methodology for science provides good norms of commitment vis-à-vis hypotheses and that if there are any purely likelihood-based norms of this kind then a norm of the form given by (PRAFE) is among them. It follows that methodological likelihoodism is false if no norm of the form given by (PRAFE) is a good one. In this section I argue that no norm of the form given by (PRAFE) is a good one because any norm of this kind can force one to violate either (3A) or (3B):

(3A) Do not prefer H_1 to H_2 and H_3 to H_4 if H_1 is logically equivalent to H_4 and H_2 is logically equivalent to H_3 (where preferring one hypothesis to another means accepting the former over the latter to a positive degree).

(3B) 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 following conditions are met:

i. H_1 and H_2 are mutually exclusive,

ii. the degree to which you accept H_1 over H_3 is well-defined, and

iii. the degree to which you accept H_2 over H_3 is well-defined and is not $-\infty$.

These rules are compelling. Take (3A). This rule seems innocuous in applications. For instance, it says not to prefer "all ravens are black" to "some ravens are white" while at the same time preferring "some white things are ravens" to "all non-black things are non-ravens." (Note that "some white things are ravens" is logically equivalent to "some ravens are white" and "all non-black things are non-ravens" is logically equivalent to "all ravens are black.") After all, one's preference between a pair of propositions should not depend on the form in which those propositions are stated.

Moreover, (3A) is completely trivial under various possible formalization of the notion of relative acceptance. For instance, if we interpret the degree to which one accepts A over B as one's log-odds $\log[\Pr(A)/\Pr(B)]$, then (3A) follows from the fact that probabilities do not change under substitution of logical equivalents.[11] In fact, (3A) follows from any formalization that allows substitution of logical equivalents.

Now take (3B). Roughly speaking, this rule directs one to accept a disjunction over an alternative claim more than one accepts one of its disjuncts over that claim, provided that one is not willing to dismiss the other disjunct completely relative to that claim. The restriction to cases in which the degree to which one accepts H_2 over H_3 is not $-\infty$ rules out cases in which one completely rejects H_2 relative to H_3. Again, this rule seems innocuous in applications. For instance, it directs one to accept "either all ravens are black or some white and the rest of 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 degree to which one accepts "all ravens are black" over "some ravens are red" is well-defined and the degree to which one accepts "some ravens are white and the rest are black" over "some ravens are red" is well-defined and

[11] It is arguably permissible in some sense for subjective degrees of belief to vary under substitution of logical equivalents. For instance, one would hardly blame a person of average mathematical ability who was attempting to assess the size of a cubic box for assigning different probabilities 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} (Rescorla, unpublished, 18–9), even though those hypotheses are equivalent. But this case does involve a failure to be fully rational; it is just a failure of logical omniscience rather than a failure of probability assessment. Thus, though excusable, it is not rationally permissible in any strong sense.

is not $-\infty$. This application of (3B0 seems obligatory. After all, "either all ravens are black or some are white and the rest are black" encompasses more possibilities than "all ravens are black," so it makes sense to accept the former over some third claim to a greater degree than letter, provided that one gives any credence at all to the additional possibilities it encompasses.

Like (3A), (3B) 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 log- odds $\log[\Pr(A)/\Pr(B)]$, then (3B) 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 $\log[\Pr(H_1 \text{ or } H_2)/\Pr(H_3)] > \log[\Pr(H_1)/\Pr(H_3)]$ when $\Pr(H_2)/\Pr(H_3) > 0$ and H_1 and H_2 are mutually exclusive. An analogous argument would work under any analogous interpretation that uses an additive (or superadditive)[12] calculus.

It is possible to give a very simple argument that (PRAFE) forces one to violate (3B) without assuming (3A). However, this argument makes an objectionable assumption. Suppose you were to run a completely uninformative experiment: you have me flip a coin with unknown bias p for heads but to report "heads" regardless of how it lands. Then $\Pr(E|p = p^*) = 1$ for all $0 \leq p^* \leq 1$. The Law of Likelihood entails that that outcome is neutral between any pair of $H_1 : 0 \leq p < 1/3$, $H_2 : 1/3 \leq p < 2/3$, and $H_3 : 2/3 \leq p < 1$. But it also implies that it is neutral between (H_1 or H_2) and H_3. This combination of claims is not problematic as long as we are only talking about evidential favoring. But using (PRAFE) to translate talk about evidential favoring into talk of acceptance yields violations of (3B).

[12] A superadditive calculus f such as the Dempster-Shafer calculus (Dempster, 1968)) is one whose axioms guarantee only that $f(H_1 \text{ or } H_2) \geq f(H_1) + f(H_2)$ when H_1 and H_2 are mutually exclusive.

If you find this argument against (PRAFE) convincing, then so much the better for my thesis. I do not place much weight on it because it has a weak point. One could claim that the size of the interval for p that a hypothesis posits is relevant to its assessment, either as part of the total relevant evidence with respect to that hypothesis or as a factor apart from the evidence that should play some role in a rule such as (PRAFE) for translating degrees of favoring into degrees of acceptance. I suspect that this claim is unsustainable, but I would rather make it moot than argue against it. I do so by providing a more elaborate argument in which I apply (PRAFE) only to intervals of equal sizes in two different parameterizations of the hypothesis space and use (3B) to generate a violation of (3A).

That argument is given in Appendix A. Here is roughly how it goes. I construct a hypothetical experiment the outcome of which is evidentially neutral between hypotheses A, B, and C according to the Law of Likelihood. By stipulation, that outcome is one's only evidence about those hypotheses. (PRAFE) thus requires you to be neutral between hypotheses A, B, and C. (3B) thus requires you to prefer (A or B) over C. I then consider an alternative set of hypotheses A', B', and C' between which the outcome is also evidentially neutral according to the Law of Likelihood. By an analogous argument, (PRAFE) and (3B) require you to prefer (A' or B') over C'. It thereby requires you to violate (3A). For the hypotheses are constructed so that (A or B) is logically equivalent to C' and (A' or B') is logically equivalent to C. Thus, (PRAFE) forces you to violate either (3A) or (3B).

Because (PRAFE) can force[13] you to violate either (3A) or (3B), the fact that (3A) and (3B) are compelling warrants the claim that (PRAFE) is not a good norm of commitment vis-à-vis hypotheses. It follows from this claim together

[13] "(PRAFE) can force you to violate..." should be understood as shorthand for "any norm of the form given by (PRAFE) can force you to violate..."

with the claims I argued for in the previous two sections that methodological likelihood is false.

4 Premise (4): A good norm of commitment is compatible with (4A)

In this section I argue that (PRAFE) is not a good norm of commitment for a second reason: it can force one to violate (4A):

> (4A) Do not prefer H_1 to $\sim H_1$ and $\sim H_2$ to H_2 if H_1 and H_2 are logically equivalent given your total evidence.

(4A) is compelling. It is similar to (3A) in that it requires degrees of acceptance to respect a certain kind of logical equivalence. Is stronger than (3A) in that requires only logical equivalence given one's evidence. It is weaker than (3A) in that it has implications only for preferences between hypotheses and their negations.

Like (3A) and (3B), (4A) seems innocuous in applications. It 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, these hypotheses have the same content and thus should be assessed alike.

Like (3A), (4A) is completely trivial under any formalization of relative acceptance that allows substitution of logical equivalents.

I prove that (PRAFE) can force one to violate (4A) in Appendix B. Here is roughly how the proof goes. 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. I show that for any constant a, one can construct a probability distribution over A and E such that the log-likelihood ratios of H_1

against $\sim H_1$ and $\sim H_2$ 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 one to prefer H_1 over $\sim H_1$ and $\sim H_2$ over H_2. In this way, (PRAFE) can force you to violate (4A).

Because (PRAFE) can force you to violate (4A), the fact that (4A) is compelling warrants the claim that (PRAFE) is not a good norm of commitment vis-à-vis hypotheses. It follows from this claim together with the claims I argued for in Sections 1 and 2 that methodological likelihoodism is false.

It is worth noting that one could avoid violating (4A) by adopting a restricted version of (PRAFE) that applies only to statistical hypotheses—that is, hypotheses that are simply about the stochastic properties of the data-generating mechanism. H_1 in my proof—the conjunction of some proposition A with E itself—is not a statistical hypothesis because it makes a direct statement about the outcome produced by the data-generating mechanism. Some likelihoodists do restrict the Law of Likelihood in this way (e.g. Hacking 1965, 59 and Edwards 1972, 57). However, 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 fact that (PRAFE) can force one to violate either (3A) or (3B). Thus, restricting the Law of Likelihood to statistical hypotheses has a high cost and is insufficient to avoid the major difficulties for methodological likelihoodism presented in this paper.

5 Against a reliabilist response

The fact that (PRAFE) can force one to violate either (3A) or (3B) and (4A) disqualifies it from consideration as a general principle of rationality. One might attempt to rescue methodological likelihoodism by lowering one's standards. Perhaps no purely likelihood-based norms of commitment are among the canons of rationality, but such norms are nevertheless useful in practice when deployed judiciously. This move may not appeal to most philosophers, but similar moves are common among statisticians (e.g. Chatfield 2002, Kass 2011, and Gelman 2011).

The idea that (PRAFE) is useful when deployed judiciously is plausible only if it has some redeeming quality that at least partially compensates for the fact that it is inconsistent with the conjunction of (3A) and (3B) and with (4A). What could that redeeming quality be? Here are four candidates from Royall (2000, 760):

(I) intuitive plausibility,

(II) consistency with other axioms and principles,

(III) objectivity, and

(IV) desirable operational implications.

I am willing to grant (I)-(III) for the sake of argument. Those virtues are not sufficient, even jointly, to vindicate methodological likelihoodism. It still needs to be shown that methods based on likelihood functions alone can provide useful guidance for our commitments vis-à-vis hypotheses.

(IV) is prima facie more promising. It refers to the purported fact that purely likelihood-based methods are guaranteed to perform well in certain senses in the indefinite long run if used over and over again with varying data. Appeals to guarantees about long-run performance are the hallmark of frequentism, but Bayesians cite such results as well, perhaps most often in the form of convergence theorems (e.g. Doob 1949). The exact significance of various facts about long-run operating characteristics is a matter of dispute, but there is no disputing the basic idea that we want techniques that we can reasonably expect to yield good results.

Unfortunately for this line of response, likelihoodist appeals to (IV) generate many problems. The most damning of these problems is that *the operating characteristics that likelihoodists appeal to are not operating characteristics of purely likelihood-based methods.* Instead, they are operating characteristics of methods that use likelihood functions in a frequentist way.

Let me explain. By definition, purely likelihood-based methods are not sensitive to differences between experimental outcomes that are not reflected in the likelihood function. One such fact concerns the distinction between *fixed* and *random* hypotheses. Fixed hypotheses are specified without reference to the data, while random hypotheses are specified in terms of the data. For instance, the hypothesis that the mean of the distribution that produced the data is *zero* is a fixed hypothesis, while the hypothesis that it is the *sample average* (the sum of the data values divided by the number of data values) is a random hypothesis, because the value of the sample average depends on the data while the value of the number zero does not.

By contrast, frequentist methods violate the likelihood principle by being sensitive to the distinction between fixed and random hypotheses. A frequentist may draw different conclusions about the hypothesis that the mean of a distribution is zero depending on whether he or she set out to test the hypothesis that the mean is zero or set out to test the hypothesis that the mean is the sample average, which turned out to be zero.

Whether sensitivity to the distinction during fixed and random hypotheses is a good feature for a method to have or not is a topic for another occasion. The key points for present purposes are (1) such sensitivity cannot be present inpurely likelihood-based methods, and (2) it is necessary for the long-run operating characteristics that likelihoodists erroneously cite in support of their methods. I will illustrate these claims for the *universal bound*, which is the fact that the probability of a likelihood ratio of at least k for any given fixed, false hypothesis against the true hypothesis is at most $1/k$ (i.e., $\Pr_{H_0}(\Pr(E|H_1)/\Pr(E|H_0) \geq k) < 1/k$) (Royall, 2000, 762-3). The same point holds for other results can concerning the performance characteristics of methods based on likelihood functions, including both the tighter bounds that Royall derives for specific distributions (2000) and likelihood ratio convergence theorems (Hawthorne, 2012).

An example[14] due to (Armitage, 1961) is a counterexample to a generalized version of the universal bound that applies to fixed as well as random hypotheses. I will simply describe the main features of the Armitage example here; see (Cox and Hinkley, 1974, 50–1) for details. The example involves taking observations until the sample average $\bar{x}$ is at least a specified distance away from zero. That distance decreases as the number of observations increases. It does so at a rate that is fast enough that the experiment is guaranteed to end in finite time,[15] but slow enough to ensure that according to the Law of Likelihood its final outcome strongly favors the hypothesis that the true mean equals x-bar over the hypothesis that it equals zero. For any k, there is an experiment of this kind such that $\Pr_{H_0}(\Pr(E|H_1)/\Pr(E|H_0) \geq k)$ is 1—a maximally severe violation of the universal bound. I have argued elsewhere (Gandenberger, unpublished) that this example should not be regarded as a counterexample to the Law of Likelihood

[14]Strictly speaking, Armitage provides a class of examples rather than a single example. I am using the word "example" as a convenient shorthand.

[15]Technically, the experiment ends "almost surely" (i.e., with probability one) in finite time.

itself. However, it is a counterexample to attempts to use the universal bound to support the use of purely likelihood-based methods.

Likelihood functions do not distinguish between fixed and random hypotheses, so purely likelihood-based methods cannot distinguish between them either. Thus, results such as the universal bound that hold only for fixed hypotheses do not support the use of purely likelihood-based methods. Methodological likelihoodists who wish to claim that purely likelihood-based methods are useful when deployed judiciously need to find some other support for that view.

6 Conclusion

Methodological likelihoodism is true only if (PRAFE) is a good purely likelihood-based norm of commitment. (PRAFE) is not a good purely likelihood-based norm of commitment because it can force one to violate both the combination of (3A) and (3B) and (4A). Therefore, methodological likelihoodism is false. The results concerning long-run operating characteristics that methodological likelihoodists sometimes cite in support of their methods do not help their cause because those results concern frequentist methods that use likelihood functions rather than purely likelihood-based methods.

References

Armitage, Peter. 1961. "Contribution to 'Consistency in Statistical Inference and Decision'." *Journal of the Royal Statistical Society. Series B (Methodological)* 23:30–1. ISSN 00359246.

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

Chatfield, Chris. 2002. "Confessions of a pragmatic statistician." *Journal of the Royal Statistical Society: Series D (The Statistician)* 51:1–20. ISSN 1467-9884. doi:10.1111/1467-9884.00294.

Cox, David and Hinkley, David. 1974. *Theoretical Statistics*. London: Chapman and Hall.

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

Doob, Joseph L. 1949. "Application of the theory of martingales." *Le calcul des probabilites et ses applications* 23–27.

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.

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

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

Gelman, Andrew. 2011. "Bayesian Statistical Pragmatism." *Statistical Science* 26:10–11. doi:10.1214/11-STS337C.

Hacking, I. 1965. *Logic of Statistical Inference*. Cambridge University Press. ISBN 9780521290593.

Halmos, Paul R. and Savage, L. J. 1949. "Application of the Radon-Nikodym Theorem to the Theory of Sufficient Statistics." *The Annals of Mathematical Statistics* 20:225–241. doi:10.1214/aoms/1177730032.

Hawthorne, James. 2012. "Inductive Logic." In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Winter 2012 edition.

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

Kass, Robert E. 2011. "Statistical inference: The big picture." *Statistical science: a review journal of the Institute of Mathematical Statistics* 26:1.

Lewis, David. 1981. "A subjectivists guide to objective chance." In *Ifs*, 267–297. Springer.

Milne, Peter. 1996. "log[P(h/eb)/P(h/b)] Is the One True Measure of Confirmation." *Philosophy of Science* 63:pp. 21–26. ISSN 00318248.

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. doi: 10.1086/587822.

Pawitan, Y. 2001. *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford science publications. OUP Oxford. ISBN 9780198507659.

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. ISSN 01621459.

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

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

Tentori, K, Crupi, V, and Russo, S. 2013. "On the determinants of the conjunction fallacy: Confirmation versus probability." .

Van Fraassen, B.C. 1980. *The Scientific Image*. Clarendon Library of Logic and Philosophy. Clarendon Press. ISBN 9780198244271.

A Proof that (PRAFE) can force one to violate either (3A) or (3B)

Proof

Suppose a mad genius has mixed water and wine in a bottle. You know only 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 perform three independent rolls of a three-sided die with weights that depend on r as shown in the following table (ignore the final column for the moment).

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

For instance, if r is in the interval $(1/2, 1]$, then the mad genius will report the results of three rolls of a three-sided die such that the probability of 1 is 1/2, the probability of 2 is 1/3, and the probability of 3 is 1/6.

Suppose the mad genius reports the outcomes 1, 2, and 3. This outcome has the same probability $(1/2)(1/3)(1/6) = 1/36$ under each of the hypotheses $H_1 : 1/2 < r \leq 1$, $H_2 : 1 < r \leq 3/2$, and $H_3 : 3/2 < r \leq 2$. It is your total relevant evidence, so (PRAFE) says to accept each of those hypotheses over each of the others to degree zero (i.e., to be neutral among them). Thus, (3B) says to prefer (H_2 or H_3) : $1 < r \leq 2$ to $H_1 : 1/2 < r \leq 1$ to a degree greater than zero.

Now consider the ratio r' of wine to water instead of the ratio r of water to wine. Before the die roll, you have no information about r' except that it is in the interval $[1/2, 2)$. The table above gives the probability distributions for the die roll outcomes under each possible value of r'. The outcome $\{1, 2, 3\}$ has the same probability $(1/2)(1/3)(1/6) = 1/36$ under each of the hypotheses $H'_1 : 1/2 \leq r' < 2/3$, $H'_2 : 2/3 \leq r' < 1$, and $H'_3 : 1 \leq r' < 2$. Moreover, it has the same probability under each of the hypotheses $H^*_1 : 1/2 \leq r' < 1$, $H^*_2 : 1 \leq r' < 3/2$, and $H^*_3 : 3/2 \leq r' < 2$. For if $H^*_1 : 1/2 \leq r' < 1$ is true, then either $1/2 \leq r' < 2/3$ or $2/3 \leq r' < 1$. Either way, the probably of the outcome is 1/36. Thus, if H^*_1 is true, then the probably the outcome is 1/36. If $H^*_2 : 1 \leq r' < 3/2$ is true, then $1 \leq r' < 2$, so the probably the outcome is 1/36. Likewise for H^*_3. The die roll is your total relevant evidence, so (PRAFE) says to accept each of the hypotheses H^*_1, H^*_2, and H^*_3 over each of the others to degree zero (i.e., to be neutral among them). Thus, (3B) says to prefer (H^*_2 or H^*_3) to H^*_1 to a degree greater than zero.

But now we have violated (3A). (H_2 or H_3) : $1 < r \leq 2$ is equivalent to $H^*_1 : 1/2 \leq r' < 1$, and $H_1 : 1/2 < r \leq 1$ is equivalent to (H^*_2 or H^*_3) : $1 \leq r' < 2$. Yet we have accepted (H_2 or H_3) over H_1 and (H^*_2 or H^*_3) over H^*_1. Therefore, (PRAFE) can force you to violate either (3A) or (3B).

Discussion

Two objections to this proof are worth discussing. First, statisticians distinguish between "simple" and "complex" statistical hypotheses. A simple statistical hypothesis says that the data- generating mechanism follows a particular probability distribution. A complex statistical hypothesis is a disjunction of simple statistical hypotheses. The Law of Likelihood applies in the first instance only to simple statistical hypotheses. It might seem that the hypotheses we consider here are complex statistical hypotheses. After all, they are disjunctions of more specific hypotheses about the value of r. One could claim that for this reason (PRAFE), properly understood, does not apply to those hypotheses.

That response to the example will not work. The hypotheses we have considered are disjunctions of hypotheses that posit particular values for r. But H_1, H_2, H_3, H_1', H_2', H_3', H_2^*, and H_3^* are not disjunctions of hypotheses that posit different probability distributions for the outcome of the die roll. They are instead disjunctions of hypotheses that all imply the same probability distribution for the outcome of the die roll. A likelihoodist who denied that we can say $\Pr(A|H) = a$ because $\Pr(A|H_i) = a$ for all H_i in some partition of H would be in deep trouble. It is this assumption that allows us to ignore irrelevant partitions of our hypotheses, which can always (or at least virtually always) be found. (For instance, we routinely ignore the fact that the hypothesis H that a given coin is fair can be partitioned into the hypothesis H_1 that the coin is fair and the moon is made of green cheese and the hypothesis H_2 that the coin is fair and the moon is not made of green cheese. In the same way, we can ignore the fact that the hypothesis $1/2 < r \leq 1$, for instance, can be partitioned into hypotheses of the form $r = 1/2 + \epsilon$ for $0 < \epsilon \leq 1/2$.)

Now, H_1^* is a disjunction of hypotheses not all which posit the same probability distribution for the outcome of the die roll. But we can arrive at a likelihood for H_1^* on $\{1,2,3\}$ in several ways. One way, used in the proof, is to interpret likelihoods as probabilities entailed by hypotheses and to use disjunction elimination to derive that H_3^* entails $\Pr(\{1,2,3\}) = 1/36$. But Bayesians and some likelihoodists want to interpret likelihoods as conditional probabilities.

There are at least two ways to get a likelihood for H_1^* under this interpretation. One way is to invoke a version of the law of total probability: if $B = (B_1 \text{ or } B_2)$, then $\Pr(A|B) = \Pr(A|B_1)\Pr(B_1|B) + \Pr(A|B_2)\Pr(B_2|B)$. Thus,

$$\begin{aligned}
\Pr(\{1,2,3\}|H_1^*) &= \Pr(\{1,2,3\}|H_1')\Pr(H_1'|H_1^*) + \Pr(\{1,2,3\}|H_2')\Pr(H_2'|H_1^*) \\
&= 1/36\Pr(H_1'|H_1^*) + 1/36\Pr(H_2'|H_1^*) \\
&= 1/36[\Pr(H_1'|H_1^*) + \Pr(H_2'|H_1^*)] \\
&= 1/36\Pr(H_1' \text{ or } H_2'|H_1^*) \\
&= 1/36\Pr(H_1^*|H_1^*) \\
&= 1/36
\end{aligned}$$

Now, some likelihoodists would reject this argument. The result does not depend on the values of $\Pr(H_1'|H_1^*)$ and $\Pr(H_2'|H_1^*)$, but the argument does mention those values and thus assumes that they exist. Some likelihoodists would reject that assumption.

For a likelihoodist who interprets likelihoods as conditional probabilities and rejects the existence of probabilities that are not objectively well-defined, there is still another way to get a likelihood for H_1^*: invoke the Principal Principle. The Principal Principle (Lewis, 1981) says that one's credence for A given a proposition which entails that the chance of A is x and no inadmissible information[16] should be x: $\mathrm{Cr}(A|H) = x$ where H entails $\mathrm{Ch}(A) = x$ and contains no inadmissible information. If we interpret $\Pr(\{1,2,3\}|H_1^*)$ as a credence, then it follows that $\Pr(\{1,2,3\}|H_1^*) = \Pr(\{1,2,3\}|H_1' \text{ or } H_2') = 1/36$, because ($H_1'$ or H_2') entails $\mathrm{Ch}(\{1,2,3\}) = 1/36$ by a disjunction elimination. If we interpret $\Pr(\{1,2,3\}|H_1^*)$ as a chance rather than a credence, then we do not need the Principal Principle but only the transparently obvious chance-chance principle which says that $\mathrm{Ch}(A|H) = x$ where H entails $\mathrm{Ch}(A) = x$ and contains no inadmissible information.

Second, one could claim that the outcomes of the die rolls are not part of one's total relevant evidence with respect to the hypotheses under consideration. One's total relevant evidence with respect to those hypotheses is the empty set. After all, one's assessment of those hypotheses makes no difference to one's assessment of the probability of that outcome. This claim is somewhat reasonable, but it does not help. We could simply ask what the probability is of one's total relevant evidence with respect to the hypotheses under consideration in the empty set is under each of those hypotheses. For each hypothesis, that probability is simply the probability that the outcome of the die roll is 1, putting us back where we started.

Now, a natural response to this maneuver is to claim that (PRAFE) applies only to non-empty bodies of relevant evidence (or to non-neutral bodies of evidence if evidence can be both relevant and neutral). One could, for instance, adapt the approach to ignorance that Norton (2008) develops by assigning the same non-numerical degree of relative acceptance I to all pairs of contingent hypotheses in cases of neutral evidence. However, this response faces at least two difficulties. First, it creates unnatural discontinuities. On this proposal, we are not required to formulate any commitments about the hypotheses H_1,

[16]See (Lewis, 1981) for a discussion of admissibility. Roughly speaking, information is in admissible if it speaks to the *outcome* of a chance process rather than to its stochastic properties.

H_2, and H_3 in light of the evidence in the example as it stands—not even a commitment of neutrality. But suppose we modified the example slightly by adding some tiny quantity ϵ to the probability of 1 under each possible value of r and subtracting it from the probability of 3 under each possible value of r. Then we would be required to formulate commitments, namely minute preferences for H_2 over H_1 and over H_3. The idea that neutral evidence requires no commitments while arbitrarily slightly non-neutral evidence requires completely definite commitments is hard to accept.

Second, this response sets up a game of cat-and-mouse that seems unlikely to end well for the methodological likelihoodist. An example involving exactly neutral evidence is needed to illustrate a conflict among (PRAFE), (3A), and (3B) only because I made those principles very weak so that they would command nearly universal assent. Similar but slightly stronger principles would conflict in cases of non-neutral evidence. For instance, (3B) says to accept (H_1 or H_2) over H_3 to a degree greater than that to which you accept H_1 over H_3 under the relevant conditions. Presumably, in each case in which (3B) applies, there is some definite amount by which the former should exceed the latter. In the case at hand, for instance, there is some positive number t such that it is at least permissible to accept (H_2 or H_3) over H_1 to degree t. One could use this margin t to generate the same kind of argument in a similar case with sufficiently slightly non-neutral evidence.

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

(4A) Do not prefer H_1 to $\sim H_1$ and $\sim H_2$ to H_2 if H_1 and H_2 are logically equivalent given your total evidence.

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, if the first coin lands heads, then $X_2 = 1$. Otherwise $X_2 = 0$. 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 $f(\mathcal{L}) = f(\Pr(T|H_1)/\Pr(T|H_2))$. By the formulation of (PRAFE), there's some constant a such that $f(a) > 0$ (and thus $f(x) > 0$ for all $x > a$, since f is nondecreasing). I will show in a moment the following:

> (*) For any a, there is a joint distribution for X_1 and X_2 such that $\Pr(E|H_1)/\Pr(E|\sim H_1) > a$ and $\Pr(E|\sim H_2)/\Pr(E|H_2) > a$.

Thus, (PRAFE) forces one to prefer H_1 to $\sim H_1$ and $\sim H_2$ to H_2. But given E, H_1 and H_2 are equivalent. Therefore, (PRAFE) forces one to violate (4A).

I will now prove (*) by showing 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 $\Pr(E|\sim H_2)/\Pr(E|H_2) > a$. Let a be some value greater than 1/2 of x such that $f(a) > 0$ for all $x > a$. Choose a $b > (2a-1)/(2a+1)$. Then assign probabilities to outcomes according to the following table.

$X_1 \backslash X_2$	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

Here is a derivation of the relevant likelihood ratios.

$$
\begin{aligned}
\frac{\Pr(E|H_1)}{\Pr(E|\sim H_1)} &= \frac{\Pr(E \;\&\; H_1)}{Pr(H_1)} \frac{\Pr(\sim H_1)}{\Pr(E \;\&\; \sim H_1)} \\
&= \frac{\Pr(X_1 = X_2 = 1)}{\Pr(X_1 = X_2 = 1)} \frac{\Pr(\sim (X_1 = X_2 = 1))}{\Pr(X_1 = 1 \;\&\; X_2 = 0)} \\
&= \frac{1 - \Pr(X_1 = X_2 = 1)}{\Pr(X_1 = 1 \;\&\; X_2 = 0)} \\
&= \frac{1 - (1-b)/2}{(1-b)/4} \\
&= \frac{4}{1-b} - 2 \\
&= \frac{4 - 2 + 2b}{1-b} \\
&= 2\frac{1+b}{1-b}
\end{aligned}
$$

This quantity is monotonically increasing in b. Thus, it follows that if $b > (2a-1)/(2a+1)$,[17] then

$$
\begin{aligned}
\Pr(E|H_1) \Pr(E|\sim H_1) &> 2\frac{1 + (2a-1)/(2a+1)}{1 - (2a-1)/(2a+1)} \\
&= 2\frac{2a + 1 + 2a - 1}{2a + 1 - 2a + 1} \\
&= 2\frac{4a}{2} \\
&= 4a \\
&> a
\end{aligned}
$$

And now the other likelihood ratio.

[17] $b > (2a-1)/(2a+1)$ is always permissible, because $a > 1/2$ implies $0 < (2a-1)/(2a+1) < 1$, and $0 < b < 1$ implies that the distribution in the table above is consistent with the axioms of probability.

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

This quantity is monotonically increasing in b. Thus, it follows that if $b > (2a-1)/(2a+1)$, then

$$
\begin{aligned}
\frac{\Pr(E|\sim H_2)}{\Pr(E|H_2)} &> 1/2\frac{1+(2a-1)/(2a+1)}{1-(2a-1)/(2a+1)} \\
&= 1/2\frac{2a+1+2a-1}{2a+1-2a+1} \\
&= 1/2\frac{4a}{2} \\
&= a
\end{aligned}
$$

Discussion

Note that restricting (PRAFE) so that it applies only to mutually exclusive hypotheses does not block this proof. (PRAFE) is applied only to the comparison between H_1 and $\sim H_1$ and the comparison between H_2 and $\sim H_2$. Those pairs of hypotheses are of course mutually exclusive. H_1 and H_2 are not mutually exclusive, but (PRAFE) is not applied to the comparison between H_1 and H_2 directly.