The Probative Value of Old Evidence

James Joyce (University of Michigan)

Most Bayesian epistemologists maintain that the overall evidential import of an item of data *x* on a hypothesis *h* is *entirely* a matter of the disparity between *h*'s prior probability (or odds) and its probability (odds) conditional on *x*. This "incrementalism" finds its fullest expression in the *Likelihood Principle* (LP), which says that the evidential import of *x* for some set of hypotheses *H* is entirely determined by the profile of likelihoods *Prob*(*x*|*h*) for *h* in *H*. LP marks a fault line between frequentist statisticians, whose practices invoke likelihoods for hypothetical data, and Bayesians and likelihoodists, who hold that only actual data is relevant to statistical inference. Glymour's *old evidence problem* undermines incrementalism and LP by identifying substantive evidential relationships that cannot be understood as disparities between priors and posteriors. But, Bayesians need not despair. The Bayesian toolkit has the resources to formulate a coherent non-incremental concept of evidence, the *probative value* of *x* for *h*, which reflects the degree to which the total evidence for *h* depends on *x*. I will argue that: (a) Probative value captures evidential relationships that incrementalists ignore, including those at issue in old evidence cases; (b) It does not compete with incremental value since both are needed for a full account of confirmation; (c) While incremental relationships ultimately compare likelihoods *Prob*(*x*|*h*) and *Prob*(*x*|*h**) for distinct hypotheses *h* and *h**, probative relationships compare *Prob*(*h*|*x*) and *Prob*(*h*|*x**) for distinct data points *x* and *x**; (d) While LP is appropriately invoked when the incremental impact of data is in question, it is out of place when probative relationships are under discussion; (f) This is because a component of LP, the Sufficiency Principle, has two readings, one plausible in incremental contexts, the other plausible in probative contexts.

The Probative Value of Old Evidence

Jim Joyce
Department of Philosophy
The University of Michigan
Ann Arbor, MI 48109 USA
email: jjoyce@umich.edu

MIT
February 13, 2009

A FAMOUS EPISODE IN THE PHILOSOPHY OF SCIENCE

Hempel (1945, 31-33) initially thought that the concept of confirmation should satisfy:

- CONSEQUENCE. Data that confirms a hypothesis *h* confirms any of *h*'s logical consequences.
- CONVERSE CONSEQUENCE. Data that confirms *h* also confirms any hypothesis that entails *h*.

"Tacking" Paradox. These imply that anything that confirms anything confirms everything!

For if data *x* confirms hypothesis *h* then it confirms *h* & *g* for any *g* (by CC) But, if *x* confirms *h* & *g* then it also confirms (*h* & *g*) v (~*h* & *g*) = *g*. And, this is true no matter how irrelevant *x* might be to *g*!

- Hempel endorsed CONSEQUENCE, and restricted CONVERSE to cases where *h* is an instance of *g*.

Carnap (1962, 468-81): The paradox is due to an equivocation on the term "confirmation".

- *Confirmation as Total Evidence*: With the addition of data *x*, the *total* amount of evidence in favor of *h* exceeds some (contextually defined) threshold.
- *Confirmation as Positive Relevance*: *x* increases the total evidence in favor of *h*.
  - CONSEQUENCE holds for total evidence, not positive relevance. CONVERSE fails for total evidence, and only holds for positive relevance when *g* is independent of *h* and of *x*.

Carnap was right, but did not go far enough! We need more notions of evidence!

TWO CONCEPTS OF POSITIVE RELEVANCE

What does it mean to say that data *x* is positively relevant to the hypothesis *h*?

> *Incremental Concept*. Adding *x* to the existing data *increases* the total evidence in favor of *h relative to its current level*. This captures the intuition that acquiring evidence for *h* makes *h more* worthy of belief than it was before the evidence was acquired.
>
> *Probative Concept*. The total evidence for *h* depends heavily on whether or not *x* obtains. This captures the intuition that *x* serves as an important part of the *basis* for our confidence in *h*.

My View: While Bayesians tend to focus on incremental evidence, the probative notion is also entirely legitimate and useful. Each captures a central aspect of our concept of evidence, and a complete theory of confirmation should embrace both.

My Goals: To explain what probative evidence is and to convince you to take it seriously.

A BRIEF OVERVIEW OF PROBABILISTIC CONFIRMATION THEORY

We will adopt the useful fiction of a rational subject' beliefs at any time *t* can be described by a *credal state* that represents her degrees of confidence in various propositions at *t*, and we think of these credal states as being determined exclusively by the subject's total evidence at that *t*.

> *Assumption*. Each total body of evidence X can be associated with a unique *credal state* that represents the system of beliefs that best reflects the evidence in X.

Note: I intend this assumption is a way that does not distinguish between objectivist and subjectivist views of the relationship between evidence and belief.

- Objectivist hold that there are objective inductive rules (like the principle of indifference or the prescription to maximize entropy) that single out the same credal state for everyone with the same total evidence.
- Subjectivists allow each individual to have his or her own "personalistic" inductive commitments, so that two people with the same total evidence might legitimately have different credal states.

> *Assumption*. Each total body of evidence is sufficiently definite and comprehensive to fix a unique *credence function* that assigns each proposition of interest a real-valued *degree of belief* between zero and one.

Note: This assumption is rarely satisfied, and we make it here only to keep things simple. When it fails credal states are represented by sets (or sets of sets) of credence functions. See Joyce (2005) for details.

WHAT IS A BODY OF EVIDENCE?

I am going to dodge this question in the hope that the details for the answer will not matter much to what follows. Here are two possible answers, either of which I believe would be consistent with what I'll have to say.

Propositional Model. A body of evidence is a set of propositions accepted as true (Levi) or known to be true (Williamson).

Probabilistic Model (Jeffrey, van Fraassen). A body of evidence is a set of (externally imposed) "constraints" on degrees of belief, like these:

- The credence for h should be in some subset S of [0, 1].
- The credence for h conditional on datum x should be in some subset S of [0, 1].
- h is independent of x, i.e., the credence for h conditional on e should coincide with h's unconditional credence.

Note: The propositional model can emulate many aspects of the probabilistic model by using "expert probabilities" of the form $q(h) = r$. One can achieve the effect, say, of having the credence for h fall in S by (i) endorsing the inductive principle that (given the other evidence) for every r such that $q(h) = r$ has positive credence, the credence for h conditional on $q(h) = r$ should be r, and then (ii) including $q(h) \in S$ as part of the total evidence.

Note: The probabilistic model can emulate many aspects of the propositional model by using constraints of the form "the credence for h should be 1".

STATIC COHERENCE

I am going to assume the standard Bayesian story about (synchronic) rationality.

> *Assumption*. Rational credences obey the *laws of probability*, i.e., for each body of evidence X there is a *probability* function P_X that captures the system of beliefs that best reflects X.

Conditional probabilities are typically "defined" as ratios of unconditional probabilities via the rule: $P_X(h \mid x) = P_X(h \,\&\, x) / P_X(x)$, provided $P_X(x)$ is defined. But, for purposes of confirmation theory it is better to interpret them in terms of their relationships to bodies of evidence.

Conditioning as Evidence Aggregation. Suppose that *x*'s truth is probabilistically compatible with evidence X in the sense that $P_X(x) > 0$, and let X+*x* be the body of evidence obtained by adding the datum that *x* is true to the total evidence. *Define* $P_X(h \mid x) = P_{X+x}(h)$.

So, relative to a total body of evidence, *h*'s probability conditional on *x* is its unconditional probability relative to the total evidence augmented by *x*.

> *Assumption*. Rational credences obey the *law of conditional probability*, so that $P_{X+x}(h) = P_X(h \,\&\, x)$ or, equivalently, $P_X(h \mid x) = P_X(h \,\&\, x) / P_X(x)$ when $P_X(x) > 0$.

Note: I make no assumptions about diachronic rationality here. Indeed, I don't think Bayesians need assumptions about diachronic rationality (but that's another story)!

EVIDENTIAL RELATIONSHIPS AS COMPARISONS OF STATES OF TOTAL EVIDENCE

I claim that much of our talk about evidential relationships involves making comparisons among states of *total evidence*. When we discuss the evidential bearing that x has on h,

i. We typically have in mind a potential "experiment" that is capable of producing (or has produced) x as a datum.

   - Formally, such an experiment can be represented by a partition $\langle x_1, x_2, \ldots, x_N \rangle$ of data statements that has x as a member and about which we have, or might come to have, information.

ii. We implicitly compare the total evidence for h, across a range of states of evidence that (a) agree about the total evidence for h in light of each potential data item, but (b) might differ about the probabilities they assign to the elements of the data partition.

   - Clause ii can be explained in terms of "Jeffrey shifts." P and Q are Jeffrey shifts of one another on $\langle x_1, x_2, \ldots, x_N \rangle$ with respect to h iff $P(h \mid x_n) = Q(h \mid x_n)$ for all x_n. Then,

$$Q(h) = \textstyle\sum_n q_n \cdot P(h \mid x_n) \text{ for some } q_n = Q(x_n) \geq 0 \text{ with } \textstyle\sum_n q_n = 1.$$

Note: If you dislike Jeffrey conditioning (but what's not to like?), you need not think of Q as arising from P directly via a Jeffrey shift. It could be, instead, that Q arises from P via the learning of some datum y that influences the evidence for h entirely by altering the evidence for the various x_n.

Note: I don't claim that this is the most general possible framework for evidence talk, but I do think is pretty general and that it suffices for capturing the essential points I want to make.

THE *PRAGMATICS* OF EVIDENCE TALK

- People typically make claims about evidential relationships as a means of conveying contextually salient information about the comparative balances of total evidence for or against hypotheses relative to various bodies of actual or potential data.

  A Key Point: This means that everything we say about evidence must be evaluated in terms of the information conveyed about differences in *total* evidence, e.g., talk about incremental evidence is not an end in itself; it is a means of conveying information about total evidence.

- In conversational contexts where such claims are usefully asserted there will typically be (a) at least some common background of “prior” evidence shared among conversational participants, coupled with (b) some differences in prior evidence, (c) a rough consensus about the legitimate differences in prior evidence and opinion, and (d) agreement about questions of interest.

- A claim about evidential relevance, made in such a context, will succeed just in case it conveys information that helps some conversational participants answer some questions of interest, but does so in a way that does not directly assert or presuppose any facts about total evidence that are legitimate points of disagreement in the context.

This constrains the ways in which information about disparities in total evidence can be conveyed.

Toy Example: Joe is being tested for heroin use. *h* is the hypothesis that he is a user. *x* says that he tests positive. We ask two experts, Jack and Jill, to tell us how a positive test should alter our opinions about *h*.

- Unlike us, the experts know the test's true positive and false positive rates: $P(x \mid h) = 0.8$ and $P(x \mid \sim h) = 0.3$.
- They also know, via an earlier test, that Joe's blood contains *enzyme-Z*.
- There is a legitimate disagreement in the scientific community about *enzyme-Z*. Jack's group thinks *enzyme-Z* is a strong indicator of recent heroin use, and that 90% of those people who have it are users. Jill's group sees it as mild contra-indicator of heroin use, and that only 1% of those who have it are users.
- By Jack' s estimate $P(h \mid x) = 0.96$. By Jill's it's $P(h \mid x) = 0.026$.

How should the experts answer our question?

- Jack would be wrong to have us set $P(h \mid x)$ at 0.96 because this value incorporates his views about *h*'s probability, which depend on his controversial views about enzyme-*Z*. Likewise for Jill.
- The best thing Jack and Jill can do, without introducing their own views, is to tell us that the value of the odds/likelihood ratio $P(x \mid h) / P(x \mid \sim h) = 8/3$ and advise us that, in the event that Joe tests positive, we should set our odds of Joe being a user at 8/3 our current odds, so that $P(h \mid x) = 8/3 \cdot P(h)$.
- This is the logically weakest thing they can say that will both answer our questions and will not introduce information that is judged controversial in the context (e.g., it is weaker than specifying the values of both $P(x \mid h)$ and $P(x \mid \sim h)$.

One Advantage

Thinking of evidence talk as providing information about disparities in states of total evidence helps to put seeming conflicts among measures of incremental evidence into perspective.

Example. Ellen is random member of a population of 100 students who took freshman calculus and who might be majoring in chemistry or math or both or neither. The distribution of students looks like this:

	Math, Chem	*Math, not Chem*	*Not Math, Chem*	*Not Math, not Chem*
Passed Calc	9	9	20	1
Failed Calc	1	1	10	49

Does learning that Ellen passed calculus provide more incremental evidence for the conclusion that she is a math major or for the conclusion that she is a chemistry major?

- The *difference measure*, $P(h \mid x) - P(h)$, says that Chem is better confirmed.
- The *likelihood ratio*, $P(x \mid h) / P(x \mid h^*)$, says that Math is better confirmed.

If we did not have the table in front of us and if we saw both functions as measuring some unitary quantity of "incremental evidence," this would seem problematic. But, the puzzle evaporates once we appreciate that these incremental measures are used to convey comparative information about states of total evidence (i.e., the whole table versus it's top row) in cases where we have incomplete information about these states.

Compare: Joe: "I did better in the stock market than Moe. I made $10K and he only made $5K."
Moe: "I did better. The worth of my portfolio increased at twice the rate that Joe's did."

ANOTHER ADVANTAGE: SOLVING THE "OLD EVIDENCE" PROBLEM?

Glymour's example of General Relativity being spectacularly confirmed by the "old" evidence of the anomalous advance of Mercury's perihelion (*T&E*, 85-93) poses at least *three* distinct problems for Bayesian confirmation theory:

Problem of New Hypotheses. How should an agent assign an *initial* subjective probability to a newly formulated hypothesis h in light of the evidence she already possesses? And, how can the mere introduction of h undermine the support of that existing hypotheses have in light of available evidence.

Problem of Logical Learning. How should a less than ideally rational agent revise her subjective probability for a hypothesis h when she discovers that h entails some item of data she already possesses?

Problem of Probable Evidence. How can a rational agent, who has a well-defined subjective probability for hypothesis h and who knows what h entails about her data, make sense of evidential relationships between h and evidence to which she assigns a subjective probability of 1? More generally, how can data that has a high initial probability whether or not h is true provide strong evidence in favor of h.

I am only interested in the third problem. (I regard the first two as *much* deeper and *much* harder to solve.)

The Problem: Suppose confirmation relations are cashed out in terms of a probability function P that assigns x a high probability.

How can x provide strong evidence for or against h (with $P(h) > 0$) given that $P(h|x)$ will only differ from $P(h)$ by a factor of $P(x \mid h) / P(x)$, which is sure to be small when $P(x) \approx 1$.

INCREMENTALISM

> On a Bayesian account, the effect of evidence *x* in confirming (or disconfirming) a hypothesis is solely a function of the increase (decrease) in probability that accrues to *x* when it is first determined to be true. (William Talbott, "Bayesian Epistemology," SEP)

The Incremental Concept of Evidence. Datum *x* provides incremental evidence for *h* only to the extent that adding *x* to the existing data *increases* the total evidence in favor of *h relative to its current level*. In other words, evidential relationships *exclusively* involve comparisons between the *current* total evidence and the current total evidence augmented by data, i.e., of the disparity between *h*'s prior probability $P(h)$ (or odds) and its probability (odds) conditional on *x*, $P(h \mid x)$.

- This captures the intuition that acquiring evidence for *h* makes *h more* worthy of belief than it was before the evidence was acquired.
- The incremental evidence that *x* provides for *h* is a matter of the disparity between *h*'s prior probability, which reflects all antecedently available evidence, and *h*'s posterior, which reflects the antecedent evidence augmented by *x*.
- Qualitatively, $P(h|x) > P(h)$. Quantitatively, $P(h|x)/P(h)$, $P(h|x) - P(h)$, $\text{Odds}(h|x)/Odds(h)$ might all be used to measure the incremental impact that leaning *x* has on *h*'s probability.
  - Aside: My view is that all these measures, and their comparative counterparts (e.g., the odds ratio $[\text{Odds}(h \mid x)/Odds(h)]/[\text{Odds}(g \mid x)/Odds(g)]$), are different but legitimate ways of describing *x*'s incremental impact. Different measures are used to answer different sorts of questions one might have about incremental evidence.

THE OLD EVIDENCE PROBLEM ASCENDANT

Facts: (A) There are cases in which x seems highly evidentially relevant to h even though $P(h) \approx P(h|x)$. (B) Many have the intuition that, in at least some contexts, the prior probability of some item of data is irrelevant to its value as evidence.

The murder was committed by The Joker, Moriarty or Lex Luthor. We are about to learn whether the victim died of a gunshot, poisoning, or a knife wound. We have a credible witness who claims to have heard a loud noise at the time of the murder. On this basis, we are 90% confident that the coroner will report that the cause of death was a gun. Given our background evidence about the way that these three villains operate, we assign the following probabilities:

	Joker (75.5%)	Moriarty (5.4%)	Luthor (17.3%)
x = gunshot (95%)	75%	3%	17%
y = poison (2.5%)	0.4%	2%	0.01%
z = knife (2.5%)	0.1%	0.4%	2%

- Learning that the death was caused by a gunshot does little to alter the probability of the Joker being the murderer since $P(J \mid x) = 0.79$, not far off the current value of 0.755. So, just as the incrementalist says, a learning experience that raises x's probability from 95% to100% provides little *new* evidence for thinking that the Joker did it.

- Yet, if you asked me to cite evidence against The Joker (before I get the coroner's report), I will *start* by saying that gunfire was the probable cause of death. Why? Because any case against the Joker must hang on x's truth-value since $P(J \mid x) >> P(J \mid y) = 0.16$, $P(J \mid z) = 0.04$.

My View: If you stick exclusively to an incrementalist picture, there is no way around the problem of old evidence. But, ***the problem lies with incrementalism, not probabilism!***

Some Incrementalist Fixes I Don't Like

Counterfactual Deletion: Compute evidential relations by deleting x from your body of evidence, thereby obtaining a new probability P^* for which $P^*(x)$ is not close to 1, and then assess the incremental impact of x relative to P^*.

> When you ask yourself how much support x gives h, you are plausibly asking how much knowledge of x would increase the credibility of h, which is the same thing as asking how much x boosts the credibility of h relative to what **else** you currently know. The 'what else' is just $K - \{x\}$. (Howson 1991, emphasis added)

Relativize to Background: Explain old evidence intuitions by appealing to changes in background data B. Say that x is evidence for h, relative to B, iff $P(h|x \,\&\, B) > P(h|B)$. Old evidence intuitions are views about the import of data in non-actual or unlikely background conditions.

> "One can ask whether x actually confirmed h, and this question is implicitly relative to the actual history of the relevant subjective probability assignment, in particular to the actual order in which relevant pieces of evidence came in. On the other hand, one can ask whether evidence x is (still and permanently), part of the evidence for h relative to some background knowledge B, and of course the answer will vary as we vary B. The first question is historical and depends on historical accident pertaining to the order in which evidence actually came in… [the second] is more like a logical relation, being insensitive to historical accident, and permanent. We think these distinct ideas are captured by the two ideas we have defined and have called "confirmation" and "evidence" (Eells and Fitelson, 2002)

While both suggestions are worth pursuing for other reasons, neither addresses the old-evidence problem. The problem really concerns how to make sense of the idea that x is evidence for h given the background knowledge we *actually* have, including any knowledge that is relevant to x.

THE RIGHT ANSWER (Joyce 1999, Christensen 1999)

- While Bayesians often focus on the incremental effects of evidence, the richness of probability theory allows us to characterize many different sorts of evidential relationships which are suited to answering many different sorts of questions about evidence.

- The intuitions about evidence that arise in the problem of probable evidence are best captured by a *non-incremental* notion of evidence that I (now) call *probative value*.

Def: An item of data x has high probative value relative to hypothesis h to the extent that the total evidence for h depends heavily on whether or not x obtains.

- o The probative value of x relative to h is a matter of the disparities between $P(h | x)$ and $P(h | y)$ where y ranges over items of data that are incompatible with x.

- In contrast with the incremental concept, probative value is *not* meant to capture the intuition that acquiring evidence for h makes h *more* worthy of belief. Rather, it captures the intuition that the evidence serves as an important part of the *basis* for our confidence in h.

My View: While Bayesians tend to focus on incremental evidence, the probative notion is also entirely legitimate and useful. Each captures a central aspect of our concept of evidence, and a complete Bayesian theory of evidence must embrace both.

IMPORTANT CAVEAT

The incremental and probative concepts do ***not*** compete with one another. They do different jobs!

- There is no conflict between saying that x provides little incremental evidence for h but has high probative value. (Example above.) Both claims can be right since each addresses a different question!

- There is no conflict between saying that x provides significant incremental evidence for h but has a low probative value. (But there are limits, see below.) For example, consider a case in which our initial estimates of causes of death are close to uniform are:

Example: Low Probative, High Incremental

	Joker (84.8%)	Moriarty (7.1%)	Luthor (8.1%)
x = gunshot (33.3%)	33.1%	0.1%	0.1%
y = poison (33.3%)	25.3%	2%	6%
z = knife (33.4%)	26.4%	5%	2%

- Learning that the death was caused by a gunshot will raise the probability of the Joker being the murderer to $P(J \mid x) = 0.994$ from its current value of 0.848, nearly 15%.
- Yet, if you asked me to cite evidence against The Joker, I will stress that he is likely to be the culprit *no matter what the coroner reports* since $P(J \mid y) = 0.76$, $P(J \mid z) = 0.79$. The case against him thus does not depend strongly on x's truth-value.

- The choice as to whether to speak in terms of incremental or probative value in a given context has to do with the *pragmatics* of evidence questions.

  - Sometimes we want to describe the evidential import of data in ways that can be used equally well by people will differential initial information about the hypothesis under test. In particular, we sometimes seek quantities, like the likelihood ratio, that people with different priors for a hypothesis can utilize when computing their posterior probabilities. So, in this context we try to ask questions about evidential relationships whose answers do not presuppose any specific views about $P(h)$. Incremental notions are useful here.

  - Sometimes we aim to describe evidential relationships that can be used by people with different initial information about what data they are likely to receive. Here we try to ask questions about evidential relationships that do not presuppose any specific "predictive distribution" over the potential data. Probative notions are useful here.

Let's explore the distinction between probative and incremental evidence a bit more rigorously and in more detail.

Some Terminology

An ***experiment*** $\mathcal{E} = \langle \mathcal{H}, \mathcal{X}, P \rangle$ is composed of the following elements:

- $\mathcal{H} = \{h_1, h_2, \ldots, h_M\}$ is a partition of *simple* hypotheses. A *composite hypothesis* H is a subset of $\mathcal{H}$.
- $\mathcal{X} = \{x_1, x_2, \ldots, x_N\}$ is a partition of *simple experimental results*. Composite *experimental results* X are subsets of $\mathcal{X}$.
- For each fixed $x \in \mathcal{X}$, there is a *likelihood function* $L_x(\bullet) = P(x\,|\,\bullet)$ that maps each $h \in \mathcal{H}$ to the probability of x given h. Think of $L_x(h) = P(x\,|\,h)$ as the degree to which h *predicts* x.
- For each fixed $h \in \mathcal{H}$, there is a *conditioning function* $C_h(\bullet) = P(h\,|\,\bullet)$ that maps each $x \in \mathcal{X}$ to the probability of h given x. Think of $C_h(x) = P(h\,|\,x)$ as the total evidence for h given x.

As the notation suggests, Bayesians think of the likelihood and conditioning functions as related to a ***prior*** probability function P defined the algebra generated by $\mathcal{H} \cup \mathcal{X}$.

The specification of a prior requires either (A) an unconditional probability distribution over the hypotheses in $\mathcal{H}$ plus the likelihood function, or (B) an unconditional probability distribution over the data points in $\mathcal{X}$ plus the conditioning function.

Note: The likelihood function for x is determined only up to multiplication by a positive constant λ_x, which means that only likelihood *ratios* $P(x \mid h) / P(x \mid h^*)$ are meaningful.

COMPARING INCREMENTAL IMPACT AND PROBATIVE VALUE

(within the context of a specific experiment $\mathcal{E}$)

Both the incremental and probative notions *agree* that

★ Questions about the comparative evidential relationships that different data items in $\mathcal{X}$ bear to a hypothesis *h* involve asking about the conditional probability function C_h.

★ Question about the comparative evidential relationships a single that data item *x* bears to different hypotheses in $\mathcal{H}$ involve asking about the likelihood function L_x.

INC_1 *Irrelevance of Prior Views about Hypotheses*. *When likelihoods are fixed*, the incremental impact of *x* on any hypothesis is independent of the probability distribution over $\mathcal{H}$.

- The incremental impact of evidence is invariant under *Jeffrey shifts* on $\mathcal{H}$. When $Q(\bullet) = \sum_h q_h \cdot P(\bullet \mid h)$ likelihood ratios are preserved even though $Q(h)$ and $P(h)$ might differ.
- Any question about the relative incremental impact of an individual item of datum *x* on hypotheses in $\mathcal{H}$ can be answered solely by appeal to (ratios of) values of $L_x = \{P(x \mid h_1), P(x \mid h_2), \ldots, P(x \mid h_M)\}$.
- Talk of the incremental impact of evidence is useful when answering questions that (a) involve disputes about probabilities of hypotheses, but (b) do not involve disagreement about the predictive properties of these hypotheses relative to the data (which is what a likelihood function captures).

This is not true for probative value, which can vary with changes in probabilities for hypotheses even when likelihoods are fixed!

PR_1 *Irrelevance of Prior Views about Data*. *When conditional probabilities are fixed*, the probative value of x on any hypothesis is independent of the probability distribution over $\mathcal{X}$.

- The incremental impact of evidence is invariant under *Jeffrey shifts* on $\mathcal{X}$. When $Q(\bullet) = \sum_x q_x \cdot P(\bullet \mid x)$ ratios of conditional probabilities are preserved even though $Q(x)$ and $P(x)$ might differ.
- Any question about the relative probative values of impact of an individual item of data x on a given hypothesis $h \in \mathcal{H}$ can be answered solely by appeal to values of $C_h = \{P(h \mid x_1), P(h \mid x_2),\ldots, P(h \mid x_N)\}$.
- Talk of the probative value of evidence is useful when answering questions that (a) involve disputes about probability of receiving various items of potential data, but (b) do not involve disagreement about the total evidence that one would have for some hypothesis given this data (which is what P_h captures).

This is not true for incremental impact, which can vary with changes in probabilities for data statements even when conditional probabilities are fixed!

THE KEY CONTRAST BETWEEN INCREMENTAL IMPACT AND PROBATIVE VALUE

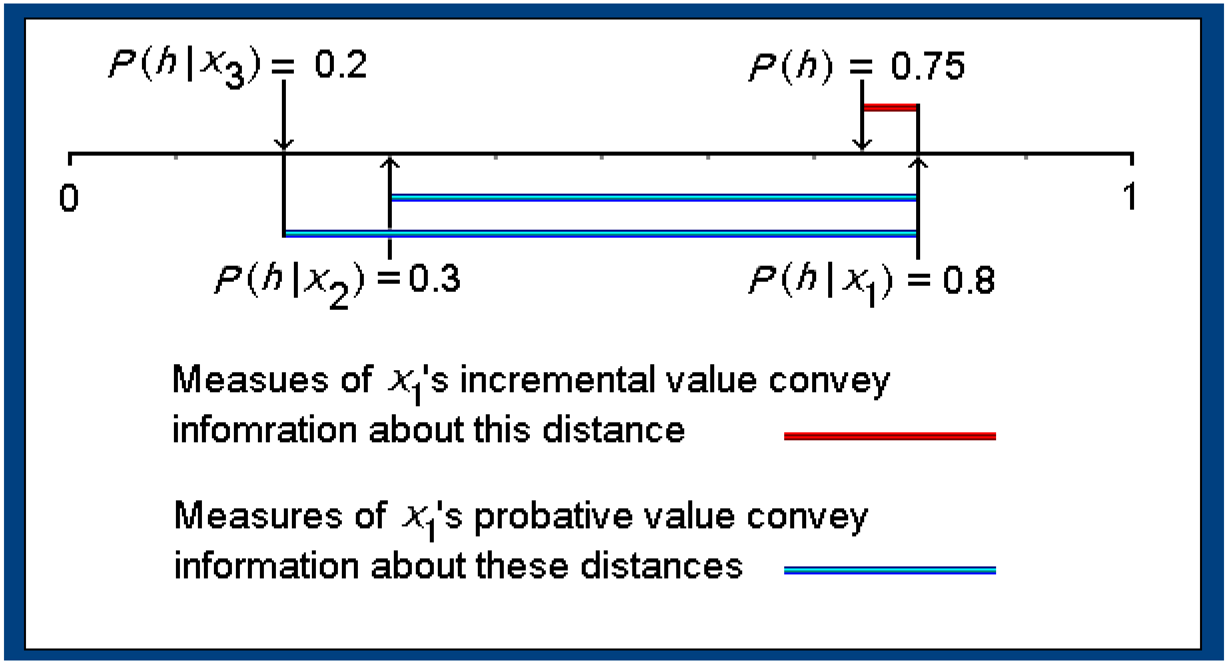
(within the context of a specific experiment $\mathcal{E}$)

INC_2 *Irrelevance of Alternative Evidence*. Given a fixed value of $P(h)$, the incremental impact of any data item x on h does depend on $P(x)$, but it does NOT depend on the probability distribution $P(y|{\sim}x)$ over $y \in \mathcal{X} \sim x$.

- Likewise, given priors for h and g, the relative incremental impact of x on h and g does not depend on the probability distribution $P(y)$ over $y \in \mathcal{X} \sim x$.

Given a prior probability for *h*, the incremental impact of learning *x* on *h*'s probability does not depend on the probability with which one might have received evidence other than *x*, whereas *x*'s probative value depends crucially on the probability of such "unseen data".

INCREMENTAL VS. PROBATIVE FOR A FIXED HYPOTHESIS



For a fixed *h* and potential data points with $P(x_1) = 0.9$, $P(x_2) = 0.09$, and $P(x_1) = 0.01$,

- Assertions about the incremental evidence that x_1 provides for *h* describe the disparity between $P(h \mid x_1)$ and $P(h)$.
- Assertions about x_1's probative value for *h* describe the disparities between $P(h \mid x_1)$ and either $P(h \mid x_2)$ or $P(h \mid x_3)$ or both.

ONE MEASURE OF PROBATIVE VALUE

While I do not think there is any one correct measure of *x*'s probative value (or incremental value), one useful kind of probative measure coincides with the *most extreme* incremental effect that *x* can have on the total evidence for a hypothesis.

- More precisely, if $\boldsymbol{I}_P(h, x)$ is a measure of incremental value defined relative to a probability *P*, there are three measures of probative value

$$\boldsymbol{I}_P^{+}(h, x) = \max\{\, \boldsymbol{I}_Q(h, x):\ Q \text{ a Jeffrey shift of } P \text{ on } \langle x_1, x_2, \ldots, x_N \rangle \}$$

$$\boldsymbol{I}_P^{-}(h, x) = \min\{\, \boldsymbol{I}_Q(h, x):\ Q \text{ a Jeffrey shift of } P \text{ on } \langle x_1, x_2, \ldots, x_N \rangle \}$$

$$\boldsymbol{I}_P^{\#}(h, x) = \text{extreme}\{\, \boldsymbol{I}_Q(h, x):\ Q \text{ a Jeffrey shift of } P \text{ on } \langle x_1, x_2, \ldots, x_N \rangle \}$$

When $\mathcal{H} = \{h, {\sim}h\}$ and $\mathcal{X} = \{x, {\sim}x\}$ and where *Q* ranges over all Jeffrey shifts on *P* on $\mathcal{X}$,

Difference: $\boldsymbol{I}_P(h, x) = P(h \mid x) - P(h)$;

$$\boldsymbol{I}_P^{\#}(h, x) = P(h \mid x) - P(h \mid {\sim}x) \quad \text{(the Christensen measure)}$$

Odds Ratio: $\boldsymbol{I}_P(h, x) = P(x \mid h) / P(x \mid {\sim}h)$

$$\boldsymbol{I}_P^{\#}(h, x) = [P(h \mid x) / P(h \mid {\sim}x)] / [P({\sim}h \mid x) / P({\sim}h \mid {\sim}x)]$$

The Problem of Probable Evidence

- Highly probable data has little incremental value, but it can have high probative value for *h* (even when $P(x) = 1$)!

- In paradigmatic "old evidence" cases $P(h|x) >> P(h|y)$ for all $y \in \mathcal{X} \sim x$, but the disparity between $P(h|x)$ and $P(h)$ is small because $P(x) \approx 1$. In effect, the differences between $P(h|x)$ and the various $P(h|y)$ are "hidden" by *x*'s high probability.

- Our intuitions about the value of old evidence in such cases concern the probative value of the data, not its incremental evidential impact. (See Joyce 1999, 203-214; Christensen 1999)

- In other words, in these cases we are interested in an assessment of *x*'s evidential import that does not depend on the way probabilities are distributed over $\mathcal{X}$.

- This is plausible reading of our current situation with regard to the GR/Mercury case.

Aside: Isn't $P(h \mid x)$ undefined *when* $P(x) = 0$?

- No. We often have better epistemic access to conditional probabilities than to unconditional ones; we know their values directly, without thinking of them as ratios $P(A \mid C) = P(A \,\&\, C) / P(C)$. The tip-off is when we find the probability of a conjunction $P(A \,\&\, C)$ by calculating it as a constant multiple of C's probability, as $P(A \,\&\, C) = P(A \mid C) \cdot P(C)$.

  Example: What's the probability that I have a dog given that I often carry doggie treats. Do you have any credence at all in either "Jim has a dog and carries doggie treats" or "Jim carries doggie treats."

- In such cases, we can make sense of conditional probabilities conditional on events having probability zero because our estimate of $P(A \mid C)$ does not depend on knowing $P(A \,\&\, C)$ or $P(C)$.
- These are ***not*** counterfactual probabilities: $P(h \mid x)$ does not measure the probability h would have if x were true. (This would be given by the *image* of h on x or by the "causal probability" $P^x(h) = \sum_k P(k) \cdot P(h \mid x \,\&\, k)$ where the k are a partition of "dependency hypotheses").
- $P(h \mid x)$ is best understood, along the lines of Rényi (1955), as a regular "factual" probability. It tells you how probable h is if x *is* true, not how probable h would be if x were true.
- The hallmark of such a "factual" probability is that it obeys a conditional version of Bayes' theorem: $P(A \,\&\, B \mid C) = P(A \mid C \,\&\, B) \cdot P(B \mid C)$ even when $P(C) = 0$.
- The bases of our knowledge of these conditional probabilities are laws of nature, causal laws, objective chances or frequency observations.

Is Surprising Data Better Evidence?

Some have claimed that hypotheses are better confirmed by the surprising data they predict than by the unsurprising data they predict. Others have felt that the prior probability of an item of data should be irrelevant to its confirming potential (Hempel 1968, p. 38). Both views are right, as well as wrong.

- The first view reflects this truth: for fixed "likelihoods" $P(x \mid h) > P(x \mid \sim h)$, the degree to which x *incrementally* confirms h increases with decreases in x's probability.

- The second view reflects this truth: for fixed posterior probabilities $P(h \mid x) > P(h \mid \sim x)$, the degree to which *x is probative for h* is independent of x's probability.

RESPONSES TO SOME OBJECTIONS TO THE PROBATIVE NOTION

- (Fitelson 2001): Any legitimate notion of evidence must satisfy:

(**F**) *x* provides more evidence than *y* does for *h* if and only if $P(h|x) > P(h|y)$

All incremental measures satisfy (**F**); no probative measure does.

Reply. While (**F**) holds for incremental evidence, it ignores what is crucial to probative. *h*'s posterior probability in light of *x* might reflect the contributions of background data, including *y* (if *y* is highly probable), far more than it reflects the contributions of *x*.

For example, it was surely true as of the third week of October that

P(Obama wins election | Obama wins Alabama) > *P*(Obama wins election | Obama wins California)

But, this was only because the probability on the left already reflected the evidence that Obama was very likely to win California, while the one on the right already reflected the evidence that Obama was almost certain to lose Alabama. This does not alter the fact that, at the time, our high confidence an Obama victory depended far more heavily on his winning California than on his winning Alabama.

- (Fitelson & Eells 2002): suggest that any legitimate notion of evidence must satisfy:

(**E**) If *x* entails *h*, then *x* provides the maximal amount of evidence for *h*.

Some incremental measures satisfy (**E**), but no probative measure does: *y* can have a higher probative value for *h* than *x* does even when *x* entails *h* and *y* does not.

Example: Sue is randomly chosen from this population:

	Elderly	Middle Age	Young
Back Pain	792	50	1
No Back Pain	8	50	99

h = Sue not young; *x* = Sue is middle age; *y* = Sue has back pain.

x, which entails *h*, has *less* probative value than *y* does for *h*.

Reply. This only seems wrong when we view evidence in its incremental guise. Though learning that Sue is middle age maximizes total evidence for *h*, we still have strong evidence for *h* even if we learn that she is not middle age (since most are elderly). So, *h*'s probability does not depend heavily on *x*. In contrast, while learning *y* will not markedly increase the evidence for *h* (since most people have back pain and few of them is young), learning ~*y* would leave us with much less total evidence for *h*. Facts about Sue's back pain are thus more central to the balance of total evidence for *h* than facts about her middle-aged status.

A DECISIVE (?) OBJECTION: PROBATIVE EVIDENCE VIOLATES LP! (Steele 2007)

Edna's rash suggests that she has a particular virus (*h*), which can be accompanied by a fever (*x*), a headache (*y*), or can be asymptomatic (*z*). The proportions of people in a North American population with Edna's rash who exhibit these symptoms are:

North America (probability *P*)

	X	*y*	*z*
h	6.66%	1.11%	2.22%
~*h*	9%	45%	36%

However, among people who have recently spent time in Borneo the proportions are different.

Borneo (probability *Q*)

	X	*y*	*z*
h	60%	10%	20%
~*h*	1%	5%	4%

- The likelihoods are identical. In particular, $P(x|h) = Q(x|h) = 0.666$ and $P(x|{\sim}h) = Q(x|{\sim}h) = 0.1$. However, $P(h|x)/P(h|{\sim}x) = 18.74$ and $Q(h|x)/Q(h|{\sim}x) = 1.27$.
- So, this measure of probative force (and all others) violates the likelihood principle.

But, is **LP** a reasonable constraint on evidence? Let's discuss.

THE LIKELIHOOD PRINCIPLE

If $\mathcal{E}$ and $\mathcal{E}^*$ are any two experiments with the same parameter space [but perhaps different sample spaces],…and if *x* and *y* are any respective outcomes determining the same likelihood function, then the evidential meaning of observing *x* in experiment $\mathcal{E}$ is identical to the evidential meaning of observing *y* in $\mathcal{E}^*$… Reports of experimental results in scientific journals should in principle be descriptions of likelihood functions. (Birnbaum 1962, 271-272)

All the information which the data provide concerning the relative merits of two hypotheses is contained in the likelihood ratio of the hypotheses on the data. (Edwards 1972, 30)

LP: Let $\mathcal{E} = \langle \mathcal{H}, \mathcal{X}, L_x, C_h \rangle$ and $\mathcal{E}^* = \langle \mathcal{H}, \mathcal{Y}, L_y, C_h \rangle$ be experiments with the same hypotheses. For fixed $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, if there is a constant $\lambda > 0$ such that $L_x(x \mid \bullet) = \lambda \cdot L_x(y \mid \bullet)$, then the evidence *x* provides for any *h* in $\mathcal{E}$ is the same as the evidence that *y* provides for *h* in $\mathcal{E}^*$.

Consequences:

- Only the data actually observed should be relevant to conclusions drawn from an experiment. *x*'s "evidential meaning" should not depend on probabilities of form $P(x^* \mid h)$ where $x^* \neq x$.
- Information about the experimental design is irrelevant once the likelihood function for the data is known. E.g., if we are not told whether $\mathcal{E}$ or $\mathcal{E}^*$ was run but only know that the likelihood function for the data is proportional to $P(x \mid \bullet)$ we should draw the same inferences as we would if we knew which experiment was run. See "A Pill of a Problem" (at the end) for an example.
- All evidential relationships are incremental!

STRICT LIKELIHOODISM

Strict Likelihoodism. Any assertion about "what data x shows" must hold solely in virtue of facts about x's likelihood ratios for *simple* hypotheses. All legitimate evidential relationships are invariant across experiments that agree about likelihood ratios involving x.

Legitimate:

- Incremental changes in *relative* odds: $[P(h \mid x)/P(g \mid x)]/[P(h)/P(g)] = P(x \mid h)/P(x \mid g)$
- Maximum Likelihood Estimation: Estimate $F(h)$ as $F(\hat{h})$ where $P(x \mid \hat{h})/P(x \mid h) \geq 1$ for all h.

NOT Legitimate:

- Incremental changes in *absolute* odds: $[P(h \mid x)/P(\sim h \mid x)]/[P(h)/P(\sim h)] = P(x \mid h)/P(x \mid \sim h)$. When $\sim h$ is composite, $P(x \mid \sim h)$ will depend on the prior probability distribution over hypotheses and not merely on the likelihoods for simple hypotheses.
- *p*-values give the probabilities of (counterfactually) observing data at least as improbable as that actually observed given a "null hypotheses". These depend on $P(y \mid h_{null})$ for $y \neq x$.
- Probative evidential relationships, like having a high value of $P(h \mid x)/P(h \mid \sim x)$, since these depend on the values of $P(y \mid h)$ for $y \neq x$ as well as the prior distribution over hypotheses.

NOTE: This is a *philosophical* thesis. It offers a kind of partial analysis of the concept of evidence, according to which two items of data that are predicatively equivalent with respect to the hypotheses under consideration are "evidentially equivalent" relative to those hypotheses.

STANDARD ARGUMENT FOR THE LIKELIHOOD PRINCIPLE

Birnbaum (1962): The Likelihood Principle is equivalent to the combination of two intuitively compelling theses about evidence:

Sufficiency Principle. If a certain statistic is known to captures all the relevant information contained in any given body of data that might be obtained in the experiment, then two bodies of data that have the same value for that statistic should coincide in evidential import.

Conditionality Principle (roughly): Suppose that you decide to perform either experiment $\mathcal{E}$ or experiment $\mathcal{E}^*$ depending upon whether a fair coin comes up heads or tails. If the coin comes up heads, then the conclusions you draw from $\mathcal{E}$ should not depend on what data $\mathcal{E}^*$ might have produced had the coin landed tails.

- I'll not discuss Conditionality. I will show that Sufficiency, and so LP, is a reasonable constraint on incremental evidence, but unreasonable for probative evidence.

The Sufficiency Principle

A statistic $T: \mathcal{X} \rightarrow \Re$ is *sufficient at value t* when $P(x \mid T(x) = t \,\&\, h) = P(x \mid T = t)$ for any $x \in \mathcal{X}$ and any $h \in \mathcal{H}$. *T* is a *sufficient statistic* (per se) when it is sufficient at all its values.

- *T*'s values capture all relevant information about $\mathcal{H}$ that is contained in the data. Given the value of $T(x)$, further facts about *x* carry no additional relevant information about $\mathcal{H}$.
- The conditional probability of observing *x* does not depend on which hypothesis is true once the value of a sufficient statistic is known.
- If *T* is a sufficient statistic, then $T(x) = T(y)$ only if the *posterior* probability of each hypothesis is the same conditional on *x* and on *y*, so that $P(h \mid x) = P(h \mid y)$ for all $h \in \mathcal{H}$.

Examples:

- In a Bernoulli process the number of successes is a sufficient statistic (order is irrelevant).
- In sampling from a normal distribution with unknown mean and variance the sample mean is a sufficient statistic for the distribution mean (the extreme sample values do not matter).

Sufficiency. Let statistic *T* be sufficient in both $\mathcal{E}$ and $\mathcal{E}^*$. If $T(x) = T(y)$ for fixed $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, then the evidence *x* provides for any *h* in $\mathcal{E}$ is identical to the evidence *y* provides for *h* in $\mathcal{E}^*$.

SUFFICIENCY AS A PRINCIPLE OF INCREMENTAL EVIDENCE

Probative value violates Sufficiency. So, if Sufficiency really is undeniable, then probative value is an illegitimate way to think about evidence.

KEY POINT: Since incremental evidential relations are a function solely of disparities between prior and posterior probabilities, it follows that *Sufficiency is a principle of incremental evidence*.

It says two things:

S_1 Facts about the data that go beyond what is needed to specify values of sufficient statistics are irrelevant to questions about the evidential import of data on hypotheses.

Given an experiment $\mathcal{E} = \langle \mathcal{H}, \mathcal{X}, L_x, C_h \rangle$ and a sufficient statistic T for $\mathcal{E}$, consider the new experiment $\mathcal{E}_T = \langle \mathcal{H}, \mathcal{X}_T, L_{T(x)}, C_h \rangle$ where $\mathcal{X}_T = \{T(x_1), T(x_2), \ldots, T(x_N)\}$. S_1 says that the evidential import of $x \in \mathcal{X}$ for on hypotheses in $\mathcal{H}$ should be the same as the evidential import of $T(x) \in \mathcal{X}_T$ for hypotheses in $\mathcal{H}$.

S_2 Facts relevant to the assessment of x's evidential import are entirely a function of the likelihoods $P(T(x)|h)$ for $h \in \mathcal{H}$: the values of $P(T(y)|h)$ for $T(y) \neq T(x)$ do not affect x's evidential import.

*P*ROBATIVE *V*ALUE *O*BEYS S_1

It would be problematic if probative value violated S_1, which does seem undeniable. Fortunately, the probative value of x in $\mathcal{E}$ is identical to that of $T(x)$ in $\mathcal{E}_T$.

To see why, recall that

★ Questions about the relative probative values of data items in $\mathcal{X}$ for an individual hypothesis h can be answered by appeal the conditioning function for h: $C_h = \{P(h|x_1), P(h|x_2),\ldots, P(h|x_N)\}$.

But, $C_h = \{P(h|T = T(x_1)), P(h|T = T(x_2)),\ldots, P(h|T = T(x_N))\}$.

Likewise

★ Questions about the relative probative value of x for hypotheses in $\mathcal{H}$ can be answered by appeal to the likelihood function for x: $L_x = \{\lambda_x P(x|h_1), \lambda_x P(x|h_2),\ldots, \lambda_x P(x|h_M)\}$.

But, since $P(x|h_m) = P(x|T(x) = t)\cdot P(T= T(x)|h_m)$, and since $\mu_x = P(x|T(x) = t) > 0$ does not depend on h, $L_x = \{\mu_x P(T = T(x_1)|h_1), \mu_x P(T= T(x)|h_2),\ldots, \mu_x P(T= T(x)|h_M)\}$

Moral: Any question about the probative value of x we ask in $\mathcal{E}$ is identical to a question about the probative value of $T(x)$ in $\mathcal{E}_T$. So, S_1 is satisfied by probative evidence.

PROBATIVE VALUE VIOLATES S_2. SO WHAT?

Probative value violates S_2 because the probative value of $T = T(x)$ is a function of both likelihoods of the form ${}_xP(T(x)|h)$ and those of the form $P(T(y)|h)$ for $T(y) \neq T(x)$.

But, this is only a problem if we *assume* that all evidential relationships are incremental. S_2 only makes sense if we think of evidential questions as being exclusively concerned with the ways in which new data *changes* initial states of total evidence. All that matters to such changes is where we are now, evidencewise, and where we will end up when we learn x or $T(x)$.

But, is there any good reason to focus exclusively on incremental evidential relationships?

Standard Answer: As Bayes' theorem teaches, posterior odds, $P(h|x) / P(g|x)$, are just prior odds, $P(h) / P(g)$, multiplied by likelihood ratios, $P(x|h) / P(x|g)$. So, posterior odds reflect both the "subjective" contribution of the prior, and the "objective" contributions of the data, as reflected in the likelihood ratios. Insofar as we want data "to speak for itself" we should focus on likelihood ratios since these capture the pure import of the data, uncontaminated by priors.

Reply: First, priors are often based on objective information and likelihood ratios often reflect subjective judgments. For purposes of evaluating the import of data, the distinction between what is objective and what is not is largely a matter of which probabilistic facts are agreed upon in a context and which are not. Second, whereas incremental relationships are invariant across certain kinds of disagreements (those about probabilities of hypotheses), probative relationships are invariant across other kinds of disagreements (about the probability of data). We need both!

MORALS

(a) It is quite true that, insofar as *x*'s *incremental* impact is concerned, it matters not a whit what the data might have looked like had *x* not been observed.

(b) It is also quite true than **LP** and **SP** are entirely legitimate when understood as principles of incremental evidence.

(c) But, neither LP nor SP justifies a prohibition on probative value.

- Even if the probability of some unobserved *y* is irrelevant when one wants to know how much observing *x* will ***change*** one's evidential position with respect to *h*, *y*'s probability can be highly relevant when one is interested in knowing how central *x* is to one's total evidence for *h*.

- This question explicitly depends on what evidence one will have had for *h* if one does not observe *x*! **TO ANSWER IT WE NEED PROBATIVE VALUE!**

References

Birnbaum, Allan (1962) "On the foundations of statistical inference," *Journal of the American Statistical Association* **57**: 269-326.

Christensen, David (1999) “Measuring Confirmation,” *Journal of Philosophy* **96**: 437-461.

Edwards, Anthony W. (1992) *Likelihood*, 2nd edition, Baltimore: Johns Hopkins University Press

Fitelson, Branden (2001) *Studies in Bayesian Confirmation Theory*, Ph.D. Dissertation, University of Wisconsin. Available at http://fitelson.org/thesis.pdf.

Fitelson, Branden and Eells, Ellery (2002) “Symmetries and Asymmetries in Evidential Support,” *Philosophical Studies* **107**: 129–142.

Glymour, Clark (1980) *Theory and evidence*. Princeton: Princeton University Press.

Howson, Colin (1991) “The 'Old Evidence' Problem,” *British Journal for the Philosophy of Science* **42**: 547- 555

Joyce, James M. (1999) *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press.

---- (2005) “How Probabilities Reflect Evidence,” *Philosophical Perspectives* **19**, 2005: 153-179.

Rényi, A. (1955). “On a New Axiomatic Theory of Probability,” *Acta Mathematica Academiae Scientiarum Hungaricae*, 6:286–335.

Steele, Daniel. (2007) “Bayesian Confirmation Theory and the Likelihood Principle”, *Synthese* **156**: 55-77.

A PILL OF A PROBLEM

You have a disease that kills 50% of its victims inside of a week. Fortunately, BigPharmaCo makes two pills that can cure you, but there are side-effects. Pill-*h*, which is expensive to make, cures 60% of patients but kills the other 40%. Pill-~*h* can be made cheaply, but it cures 40% and kills 60%. To help BP increase profits, the FDA is allowing it to market both pills under the same name at a high price. The pills are indistinguishable except that each is marked with a lot number, and all pills from a given lot are of the same type.

While BP keeps lot numbers secret, researchers may perform statistical studies to determine which lots are *h*-pills and which are ~*h*-pills. Usually, the FDA requires that these studies be conducted according to

> **Protocol-A**: Give pills with the same lot number to twenty randomly chosen patients. See how many die as a side effect. Report out $\langle t = 20, s_{20} \rangle$ where $t = 20$ is the number of trials (fixed in advance) and s_{20} is the number of patients who died in 20 trials.

However, under pressure from BigPharma, the FDA is also allowing (optional stopping)

> **Protocol-B**: Give pills with the same lot number to twenty randomly chosen patients. See how many die from side effects. If the number of deaths s_{20} is eight or fewer, report out $\langle t = 20, s_{20} \rangle$. If not, keep administering the drug to randomly selected patients until the ratio of deaths to trials s_t / t is less than 0.40. Then, stop and report out $\langle t, s_t \rangle$. When $t = 1000$, stop the study, publish nothing, and deny that any study was even attempted.

To fund their studies, researchers are allowed to charge individuals a fee of $100 to learn which protocol was used to conducting a study. So, a standard report will look like this

> "*t* subjects took pills from lot #xx, s_t died from side effects. To learn which protocol was used send a certified check for $100 to…"

Of course if $t > 20$, it is crazy to spend the money since you know Protocol-B was used.

You have acquired a pill marked **Lot #551** (and have no chance of getting another). After scouring the literature, you find only one report which reads:

> "20 patients took pills from lot #**551**, 8 died from side effects. To learn which protocol was used send a certified check for $100 to…"

Some Questions:

- Should you pay the money?
- Should you take the pill if you don't pay?
- Should you take the pill if you learn that Protocol-A was used?
- Should you take the pill if you learn that Protocol-B was used?

You decide consult three statistician friends to see what you should do.

Mr. F: “Pay the $100. If the data was gathered using Protocol-A, take the pill. Data from Protocol-B is worthless! Protocol-B is set up so that evidence contrary to *h* can never emerge. No matter what results the researchers might have gotten, they weren’t going to report anything unless it supported the hypothesis that pills in Lot #551 are of type *h*. If you learn Protocol-B was used, you should not let the results of the study influence your views about the safety and efficacy of your pill.”

Rev. B: “Do not pay, just take the pill. It doesn’t matter which protocol was used! What matters is what *actually* happened: in a population of randomly selected patients, 20 took the drug and 8 died. Counterfactual questions about what would have happened if other results had been observed are irrelevant to what you should believe. What should matter is that the data *actually* obtained makes *h* highly probable! I am convinced that BP is not so immoral as to market more ~*h*-pills than *h*-pills. I’ve done the math and, on that basis, the probability that your pill is of type *h* is at least 0.835!”

Ms. L: “Rev. B is right about the money. It doesn’t matter which protocol was used: the evidential import of the data is the same either way. But, I’m not sure where he gets off with the ‘BP is not so immoral…’ junk. Who cares about how confident he is about *h*? Fact is, we have no objective information about that. If BP is marketing 99% ~*h* pills then the probability of *h* is less than 0.05 given the results of either study. If they are marketing 99% type-*h* pills then the probability is 0.998. Anything in between is possible. What the data say, and all they say, is that the reasonable odds to assign *h* are about five times higher than they were before the data came in!”

Who’s right?

***LP says: Rev B and Ms. L are Correct*!**

Likelihoods in Protocol-A

	$P(\langle t, s\rangle \mid h \,\&\, A) =$	$P(\langle t, s\rangle \mid {\sim}h \,\&\, A) =$
$t = 20$	$\binom{20}{s}\cdot(2/5)^{s}\cdot(3/5)^{20-s}$	$\binom{20}{s}\cdot(3/5)^{s}\cdot(2/5)^{20-s}$
$t \neq 20$	0	0

Likelihoods in Protocol-B

	$P(\langle t, s\rangle \mid h \,\&\, B) =$	$P(\langle t, s\rangle \mid {\sim}h \,\&\, B) =$
$t < 20$	0	0
$t = 20,\ s_{20} \leq 8$	$\binom{t}{s}\cdot(2/5)^{s}\cdot(3/5)^{t-s}$	$\binom{t}{s}\cdot(3/5)^{s}\cdot(3/5)^{t-s}$
$t > 20,\ s_{20} \leq 8$	0	0
$s_t/0.4 \geq t > 20,\ s_{20} > 8$	$\binom{t}{s}\cdot(2/5)^{s}\cdot(3/5)^{t-s}/N_h$	$\binom{t}{s}\cdot(2/5)^{s}\cdot(3/5)^{t-s}/N_{\sim h}$
ELSE	0	0

N_h and $N_{\sim h}$ are (different) normalizing constants

At $t = 20$ and $s_{20} \leq 8$ the likelihoods are the same in either study given both h and $\sim h$. Thus, the data $\langle t = 20, s_{20}\rangle$ has the same "evidential meaning" in either study, according to **LP**.

➢ **REDUNDANCY**: Redundant evidence may seem to pose problems for both the probative and incremental concepts of evidence.

Example (Christensen 1999): *h* = there are deer in the woods; *x* = there are deer tracks in the woods; *y* = there are deer droppings in the woods. Neither *x* nor *y* has much probative (or incremental) value when the other is known since $P(h|x)$ is close to $P(h|{\sim}x)$ when $P(y) \approx 1$, and $P(h|y)$ is close to $P(h|{\sim}y)$ when $P(x) \approx 1$. But, we have the intuition that both *x* and *y* provide strong evidence for *h* even when the other is known.

Reply: It is entirely *right* that neither *x* nor *y* is good evidence for *h* when the other is known, either in the incremental or in the probative sense. This is just the result we want!

The sense that *x* and *y* each provide strong evidence for *h* even when the other is known is explained by two things:

- Conditional on the other being false, each provides strong incremental and (perhaps) probative evidence for *h*. (Note: I think we can often make sense of probabilities of the form $P(h \mid x \,\&\, {\sim}y)$ or $P(x \mid h \,\&\, {\sim}y)$ even when we know *y* is true.)

- The disjunction $x \vee y$ might have high *probative* value for *h* even when both its disjuncts are highly probable. Your confidence in *h* depends neither on *x* nor *y* alone, but it hinges strongly on the fact that one or the other is true.