QUOL

Qualitative, Objective Likelihoodism

1 The Law of Likelihood

The **law of likelihood** (LL) asserts that a piece of evidence E "favors" hypothesis H_1 over H_2 if $P(E|H_1) > P(E|H_2)$.[1] Sober argues that some of the greatest scientific achievements – from Darwin's arguments for common ancestry to Eddington's argument that the bending of light during an eclipse favors Einstein's theory of relativity – are applications of LL.[2] Moreover, proponents of LL argue that it provides an "objective" standard for evidence that Bayesian measures of confirmation, which depend upon an experimenter's subjective (prior) credences, cannot.[3] But likelihood functions are sometimes not intersubjectively shared, and conversely, some Bayesian "priors" are shared because they encode historical data. So in what sense, if any, is LL an "objective" standard of evidence?

In this paper, I provide a precise framework for characterizing the objectivity of **statistical principles**, i.e., theses like LL and the related **likelihood principle** (LP) that characterize either (i) when a single piece of evidence favors one hypothesis over another (and to what degree) or (ii) when two pieces of evidence are "equivalent."[4] Roughly, I argue that one statistical principle is more objective than another if the set of (potentially hypothetical) agents who endorse the former contains the set of agents who

[1] Hacking [1965], Royall [1997], and [Edwards, 1984]

[2] Sober and Steel [2015], [Sober, 2015, p. 82]

[3] Royall [1997, p. 10-11], for example, writes, "Whereas [LL] measures the support for one hypothesis H_1 relative to a specific alternative H_2, without regard either to the prior probabilities of the two hypotheses or to what other hypotheses might also be considered, the law of changing probability [which measures support by the difference between $P(H|E)$ and $P(H)$] measures support for H relative to a specific prior distribution over H and its alternatives ... The law of changing probability is of limited usefulness in scientific discourse because of its dependence on the prior probability distribution, which is generally unknown and/or personal." I have changed Royall's notation slightly. Concerning the related likelihood principle, [Berger and Wolpert, 1988, p. 67] claim, "It should be observed, first of all, that the LP is entirely objective, stating only that the evidence about e is contained in the likelihood function."

[4] I borrow the phrase "statistical principles" from [Wasserman, 2012], who is critical of LP. For defenses of LP, see [Birnbaum, 1962] and [Berger and Wolpert, 1988].

endorse the latter. There are at least two common reasons one principle might be more objective than another in this sense: (1) the former might be acceptable to agents with more diverse beliefs and values, and (2) the latter may be acceptable only to agents meeting implausibly strict "coherence" or "rationality" constraints. Here, my analysis of objectivity draws on a long tradition in philosophy of science [Douglas, 2009, Lloyd, 1995, Longino, 1990, Potter, 2006].[5]

In section two, I use this framework to explain why probabilistic principles like LL and LP characterize evidential considerations that are acceptable (in a sense to be clarified) to the community consisting exclusively of Bayesian agents. My results make precise *one* sense in which LL and LP are more objective standards for evidence than measures of confirmation that rely on prior probabilities.

My results are preliminary. But the framework I introduce, I believe, is philosophically important because it provides a method for *finding new* plausible statistical principles. Specifically, my framework allows one (1) to investigate whether LP, which characterizes *sufficient* conditions for evidential equivalence, also provides necessary conditions or whether some alternative principle is needed, (2) to extend LL to *composite* hypotheses in a systematic fashion, and (3) to motivate and justify *purely qualitative* analogs to LL and LP.

In section three, I investigate statistical principles for agents who fail to adhere to Bayesian coherence norms. Specifically, I state qualitative analogs of LL and LP and, using the same framework described in section two, I prove the qualitative principles would be endorsed by a community of agents whose members might have non-additive, non-normalized, non-Archimedean credences. My results are important because they suggests that, to reject Bayesians' conclusions about stopping rules, for example, classical statisticians might need to reject extremely weak coherence axioms.

I should note that "qualitative" axioms for probability are well understood and predate the standard, quantitative probability theory due to Kolmogorov.[6] What is novel about section three is the use of qualitative axioms

[5] Gelman and Hennig [2017] argue that statisticians ought to abandon discussions of "objectivity" in favor of clearer concepts like "transparency", "consensus", and more. The concepts Gelman and Hennig identify, I believe, are similar to the various philosophical concepts of objectivity I discuss below.

[6] For discussions of qualitative probability relations, see [Keynes, 2004] [Koopman, 1940a,b], [De Finetti, 1937], and [Savage, 1972, Chapter 3]. For conditions under which those qualitative relations are representable as probability functions, see [Fishburn, 1986] and [Krantz et al., 2006].

and a definition of intersubjective agreement to derive purportedly objective likelihoodist principles that (a) constrain the set of acceptable statistical techniques but (b) do not entail Bayesianism.

2 Intersubjectivity and Statistical Principles

2.1 Statistical Problems and Interpretations of Probability

Let Θ represent the **simple hypotheses** under investigation; a **composite hypothesis** $H \subseteq \Theta$ is a disjunction of simple hypotheses.[7] Let $\mathcal{E}$ represent the experiments that might be performed. Each experiment $\mathbb{E} \in \mathcal{E}$ has a set of possible outcomes $\Omega_{\mathbb{E}}$, which represents the data one might obtain. Statisticians assume that, for each experiment $\mathbb{E}$ and each hypothesis $\theta \in \Theta$, there is a probability measure $P_\theta^{\mathbb{E}}$ over $\Omega_{\mathbb{E}}$. The function $P_\theta^{\mathbb{E}}(\cdot)$ is called a **likelihood function**. Bayesians claim that, when an experiment $\mathbb{E}$ is conducted, a rational scientist's credences ought to be representable by a function Q that assigns probabilities to *both* hypotheses in Θ and experimental outcomes in $\Omega_{\mathbb{E}}$

For example, imagine we find a coin and want to know if it's biased towards heads. One experiment $\mathbb{E}_1$ consists in flipping the coin ten times; another $\mathbb{E}_2$ consists in flipping the coin until two heads are observed. In $\mathbb{E}_1$, the set of possible outcomes $\Omega_{\mathbb{E}_1}$ contains all binary sequences of length ten, representing the ten coin flips. In $\mathbb{E}_2$, the set $\Omega_{\mathbb{E}_2}$ contains all finite binary sequences containing exactly two ones (representing the two heads), where the last digit is one (because flipping stops after the second head).

Among statisticians, the set of hypotheses Θ would typically be represented by the unit interval $[0, 1]$, where $\theta \in \Theta$ represents the hypothesis that, on a given thrown, the coin's bias (i.e. chance of landing heads) is precisely θ. A likelihood function $P_\theta^{\mathbb{E}_1}$ in experiment $\mathbb{E}_1$ specifies how likely various sequences of ten tosses of the coin are, if the coin's bias equals θ.

Philosophers, however, should carefully distinguish **scientific hypotheses**, like "The coin is evenly weighted", from **statistical hypotheses**, like "The coin's bias is 1/2 and flips are iid." Statistical hypotheses, in my terminology, *determine* a probability distribution over experimental outcomes, whereas a scientific hypothesis may determine a distribution only in conjunction with additional background hypotheses. For example, the scientific hypothesis "The coin is evenly weighted" entails the "The coin's bias is 1/2

[7] To avoid unnecessary measure-theoretic complications, I assume that the set of hypotheses Θ is finite and that the set of experimental outcomes is countable. I also assume all probability measures are defined on the appropriate power set algebra.

and flips are iid" if one makes additional assumptions about who or what is throwing the coin (e.g., that it is not a machine designed to throw only heads).

Scientists are typically interested in testing scientific hypotheses, not just statistical ones. For instance, although a psychologist might officially test the statistical hypothesis that a purported psychic is no better than chance in predicting a coin flip, she likely wants to test the scientific hypothesis that precognition does not exist. Nonetheless, scientists often share theoretical commitments and beliefs about measurement techniques that allow one to treat an idealized statistical hypothesis as if it were entailed by a scientific hypothesis (and so a test of the former is a test of the latter).

The distinction between statistical and scientific hypotheses is important for interpreting a likelihood function $P_\theta^{\mathbb{E}}$. Some statisticians and philosophers claim that likelihood functions are more objective than an experimenter's credences over hypotheses.[8] But in what sense are likelihoods more "objective?" If θ represents a true statistical hypothesis, then $P_\theta^{\mathbb{E}}$ specifies the probability of obtaining particular data in virtue of the *meaning* of θ. In contrast, if θ represents a scientific hypothesis – and this is almost always the case – then $P_\theta^{\mathbb{E}}$ might be best interpreted as representing *widely shared credences* about how likely various data are under supposition that the hypothesis represented by θ is true. Thus, when scientific hypotheses are under investigation, claiming that prior probabilities are more "subjective" than likelihoods is sometimes misleading. The difference between the two probability functions is often a matter of degree, not kind: likelihoods are more "objective" because, in virtue of shared theoretical commitments, more scientists share likelihood functions than share priors.

If likelihood functions often represent (shared) subjective credences, one might conjecture that LL and LP are likewise "more objective" simply because they characterize when a diverse group of learners will agree about some evidential question. That is roughly the conclusion of the next two subssections.

[8] See quotation in footnote 3 from [Royall, 1997, p. 10-11]. See also [Fisher, 1956, p. 59] who writes, "Probabilities obtained by a fiducial argument are objectively verifiable in exactly the same sense as probabilities assigned in games of chance."

2.2 Law of Likelihood and Favoring

Although proponents (and critics) of the law of likelihood (LL) agree that "favoring" is not a pre-theoretic notion,[9] they nonetheless agree that certain intuitions about "favoring" can be used in support or criticism of LL. Observing 95 heads in 100 tosses of a coin, for instance, intuitively favors the hypothesis that the coin is weighted (vs. not). But many proponents of LL distinguish evidence that favors a hypothesis from evidence that justifies believing a hypothesis [Royall, 1997]. A bloody glove, for example, favors the hypothesis H that a suspect is guilty, but if the suspect has a trustworthy alibi, then a juror is not justified in believing H.

Further, an observation may favor one hypothesis over another, without providing good evidence for either. For example, imagine a white metal object is found near a tin mine and emits a bluish-white color when burned. Then the evidence favors the hypothesis that the object is cerium (which burns blue) over the hypothesis that it is cadmium (which burns red). Intuitively, however, the evidence best supports the hypothesis that the object is tin (which burns bluish-white) and neither cerium nor cadmium.

Here is one natural way of formalizing how some philosophers use the word "favoring." Consider the example again in which you find a silvery object near a tin mine that emits a bluish-white color when burned. Call that evidence E. Let H_1 represent the hypothesis that the object is cerium and let H_2 denote the hypothesis the object is cadmium. Normally, E would not increase in your confidence in either H_1 or H_2. However, if you knew all along that either H_1 or H_2 is true, then intuitively, E ought to increase your confidence in H_1. This is how I will explicate "favoring."

Formally, suppose Q, which represents an agent's credences, is a probability measure that assigns probabilities to *both* hypotheses and data. Let $supp(Q) = \{\theta \in \Theta : Q(\theta) > 0\}$ be the set of simple hypotheses that the agent assigns positive probability; $supp(Q)$ is called the **support** of Q in Θ. Given a set of experiments $\mathcal{E}$, I say an agent's conditional degrees of belief **match likelihoods over** $\mathcal{E}$ if $Q(\cdot|\theta) = P_\theta^{\mathbb{E}}(\cdot)$ for all $\mathbb{E} \in \mathcal{E}$ and all $\theta \in supp(Q)$. As discussed in §2.1, statisticians typically assume that likelihood functions are widely shared, and so it's reasonable to restrict attention to agents whose credences match likelihoods.

Let E be an outcome of an experiment $\mathbb{E}$. Given two hypotheses $H_1, H_2 \subseteq \Theta$ such $H_1 \cap H_2 = \emptyset$, say E **Bayesian favors** H_1 to H_2 if $Q(H_1|H_1 \cup H_2) < Q(H_1|E \cap (H_1 \cup H_2))$ (or equivalently $Q(H_2|H_1 \cup H_2) > Q(H_2|E \cap (H_1 \cup H_2)))$

[9] See [Sober, 2015, p. 81] and [Fitelson, 2007]. See [Chandler, 2013] for a discussion of "intuitive" support for a collection of related principles about "favoring."

whenever Q matches likelihoods. In other words, evidence E Bayesian favors one hypothesis H_1 to another H_2 if *every* Bayesian (no matter her prior) would be more confident in H_1 upon learning E, assuming she knew either H_1 or H_2 were true.

Theorem 1 *Suppose Θ is finite. Then E Bayesian favors H_1 to H_2 if and only if* $\min_{\theta \in H_1} P^{\mathbb{E}}_{\theta}(E) > \max_{\theta \in H_2} P^{\mathbb{E}}_{\theta}(F)$. *When $H_1 = \{\theta_1\}$ and $H_2 = \{\theta_2\}$ are simple hypotheses, it follows that* LL *entails that E favors H_1 to H_2 (i.e., $P^{\mathbb{E}}_{\theta_1}(E) > P^{\mathbb{E}}_{\theta_2}(E)$) if and only if E Bayesian favors H_1 to H_2.*

I do not claim that theorem 1 is mathematically deep. Further, [Chandler, 2013][Observation 5] has discussed the right-to-left direction in the special case in which H_1 and H_2 are simple hypotheses. The left-to-right direction, however, has an important philosophical consequence. Namely, even if likelihood functions represent (shared) subjective credences, there is a precise sense in which LL is an "objective" evidential principle: it characterizes precisely when all Bayesians, *regardless of prior beliefs and values/utilities*, will agree about when one piece of evidence favors one hypothesis to another.

I should contrast Theorem 1 with existing work. Some philosophers and statisticians have advocated extending LL to composite hypotheses as follows:

† E favors H_1 to H_2 if and only if $\sup_{\theta \in H_1} P_{\theta}(E) > \sup_{\theta \in H_2} P_{\theta}(E)$.

The ratio $\sup_{\theta \in H_1} P_{\theta}(E) / \sup_{\theta \in H_2} P_{\theta}(E)$ is also used in likelihood ratio tests [Casella and Berger, 2002, §8.2.1, p. 375]. † has two virtues: (i) it reduces to LL when the hypotheses are simple and (ii) it is entailed by the conjunction of LL and two fairly intuitive axioms [Bickel, 2012]. Yet † conflicts with what some Bayesians will regard as evidence because † entails that E favors H_1 over H_2 in some cases in which there are priors Q for which $Q(H_1|H_1 \cup H_2) > Q(H_1|E \cap (H_1 \cup H_2))$, i.e., cases in which some rational agents' comparative confidence in H_1 would be *reduced* upon observing E. Theorem 1 provides an alternative generalization of LL that (i) also reduces to LL when the hypotheses are simple, (ii) is consistent with Bayesianism by design, and (iii) saves many of the same intuitions as † because it entails one direction of †. Perhaps most importantly, my method for deriving an alternative version of LL – by securing it in intersubjective agreement – provides a principled way of finding other evidential principles, including purely qualitative ones, which I will describe in upcoming sections.

2.3 Comparing Pieces of Evidence

Theorem 1 is a corollary of another result that will motivate my general framework for deriving qualitative statistical principles. Two pieces of evidence E and F might favor the same hypothesis, but E might be stronger evidence than F. For instance, 75 heads in a 100 tosses favors the hypothesis that a coin is biased towards heads, but 90 heads in a 100 tosses seems to favor that hypothesis even more. If the definition of "Bayesian favoring" correctly formalizes intuitions about evidence, then it is natural to generalize that definition as follows. Given two hypotheses $H_1, H_2 \subseteq \Theta$ such $H_1 \cap H_2 = \emptyset$, say E **Bayesian favors** H_1 to H_2 **at least as much as** F does if $Q(H_1|E \cap (H_1 \cup H_2)) \geq Q(H_1|F \cap (H_1 \cup H_2))$ for all priors Q that match likelihoods. In other words, observing E would raise one's confidence in H_1 at least as much as F if one knew H_1 or H_2 to be true. By definition, E Bayesian favors H_1 to H_2 (simpliciter) if and only if E Bayesian favors H_1 to H_2 at least as much as $\Omega_{\mathbb{E}}$. So theorem 1 follows from the next result.

Theorem 2 *Suppose Θ is finite, and let E and F be outcomes of experiments $\mathbb{E}_1$ and $\mathbb{E}_2$ respectively. Then E Bayesian favors H_1 to H_2 at least as much as F does if and only if*

$$\min_{\theta \in H_1} P_\theta^{\mathbb{E}_1}(E) / \max_{\theta \in H_2} P_\theta^{\mathbb{E}_1}(E) > \max_{\theta \in H_1} P_\theta^{\mathbb{E}_2}(F) / \min_{\theta \in H_2} P_\theta^{\mathbb{E}_2}(F).$$

I should contrast theorem 2's implications with existing proposals for measuring evidential strength. Some philosophers and statisticians argue that the ratio $\sup_{\theta \in H_1} P_\theta^{\mathbb{E}_1}(E) / \sup_{\theta \in H_2} P_\theta^{\mathbb{E}_1}(E)$ is a numerical measure of *how strongly* evidence E favors H_1 over H_2. That thesis is strictly stronger than LL, which does not entail the degrees of "favoring" can be numerically quantified. Proponents of the stronger thesis are, therefore, committed to the claim that E is stronger evidence for H_1 over H_2 if $(*)$ $\sup_{\theta \in H_1} P_\theta^{\mathbb{E}_1}(E) / \sup_{\theta \in H_2} P_\theta^{\mathbb{E}_1}(E) > \sup_{\theta \in H_1} P_\theta^{\mathbb{E}_2}(F) / \sup_{\theta \in H_2} P_\theta^{\mathbb{E}_2}(F)$. Theorem 2 entails that the inequality $*$ might hold, and yet, one's confidence in H_1 might be *lower* upon learning E than upon learning F. Intuitively, stronger evidence for a hypothesis ought to increase one's confidence more than weaker evidence, and so theorem 2 suggests the ratio $\sup_{\theta \in H_1} P_\theta^{\mathbb{E}_1}(E) / \sup_{\theta \in H_2} P_\theta^{\mathbb{E}_1}(E)$ is not an appropriate measure of evidence. For the same reason, theorem 2 speaks against virtually every numerical measure of confirmation other than the so-called "difference measure" $Q(H|E) - Q(H)$, which depends upon the prior probability of the hypothesis H.[10]

[10] See [Evans, 2015] for a list of proposed measures of confirmation.

Although I will not defend so here, my suspicion is that debates about measures of confirmation are misguided because they presume that strength of evidence (i.e., confirmation) is numerical. Yet no argument, to my knowledge, establishes that one can always compare the strengths of two different pieces of evidence, let alone that those strengths can be quantified on some meaningful scale (e.g., interval or log). My definition of Bayesian favoring is compatible with the claim that certain pieces of evidence are incomparable in strength. Why? It's possible that E raises one Bayesian agent's confidence more than F does, but F raises a second's confidence more than E. In such a case, the two agents disagree about which piece of evidence is stronger. To claim that evidence is numerically quantifiable seems to entail that, when such disagreements arise, either (i) one agent is wrong in virtue of having irrational degrees of belief or (ii) both agents are correct because strength of evidence is merely personal/subjective in the sense that it is always a function of a single agent's degrees of belief. Both options are too radical, I think. The first is incompatible with the intuition that reasonable disagreement is possible, and the second is in tension with the fact that often there is widespread agreement about which experimental outcomes would favor particular hypotheses, even when there is substantial reasonable disagreement about how likely the hypotheses are.

One might conjecture that it's too demanding to require *all* Bayesians to agree that some piece of evidence is stronger than another and that such a requirement would entail that, more often than not, two pieces of evidence will be incomparable in strength by that standard. As it turns out, however, theorem 2 entails that the definition of Bayesian favoring agrees with fairly robust intuitions about evidence, like the intuition that, all other things being equal, a larger sample provides more evidence than a smaller one.

Example 1: Let $\mathbb{E}_1$ be an experiment in which a coin is tossed 10 times and $\mathbb{E}_2$ be an experiment in which a coin is tossed 100 times. Let H_1 be the hypothesis "The coin is fair" and let H_2 be any hypothesis of the form "The coin has bias θ" where $\theta \neq 1/2$. Then observing fifty heads in $\mathbb{E}_2$ Bayesian favors H_1 over H_2 more than observing five heads in $\mathbb{E}_1$.[11]

The example is important because recent work by [Vieland, 2017] proposes to supplement LL by additional constraints on numerical measures of confirmation, and chief among those constraints is the requirement that larger samples are typically better. I stress that this fact falls naturally out

[11]Contact the author for a proof.

of the definition of Bayesian favoring, and so no *ad hoc* additional evidential principles are required.

As the example shows, theorem 2 allows one to compare pieces of evidence E and F that might be outcomes of *distinct* experiments $\mathbb{E}_1$ and $\mathbb{E}_2$. It is, therefore, natural to ask when pieces of evidence from different experiments are equally informative. This is what the likelihood principle was proposed to do.

2.4 LP and Evidential Equivalence

Let $\{P_\theta^{\mathbb{E}_i}\}_{\theta \in \Theta}$ $(i = 1, 2)$ be the likelihood functions associated with the experiments $\mathbb{E}_1$ and $\mathbb{E}_2$ respectively; the experiments may be distinct or the same. Here, I assume that the same set of hypotheses Θ is being tested in both experiments. The **likelihood principle** (LP) asserts that, given an outcome E of $\mathbb{E}_1$ and an outcome F of $\mathbb{E}_2$, if there is a constant $c > 0$ such that $P_\theta^{\mathbb{E}_1}(E) = c \cdot P_\theta^{\mathbb{E}_2}(F)$, then E and F are **evidentially equivalent**. The notion of evidential equivalence is undefined, but it is typically assumed that, if E and F are evidentially equivalent, then one ought to draw the same inferences about the hypotheses in Θ upon learning E and as one would upon learning F.

For example, suppose there are two experiments designed to test among three hypotheses θ_1, θ_2, and θ_3; the likelihoods of observing various outcomes are described in the table below. LP entails that observing A in Experiment 1 is evidentially equivalent to observing C in Experiment 2 because the probability of observing C is, no matter what hypothesis is true, always twice that of observing A.

Experiment 1

	Observe A	Observe B
θ_1	10%	90%
θ_2	20%	80%
θ_3	5%	95%

Experiment 2

	Observe C	...
θ_1	20%	...
θ_2	40%	...
θ_3	10%	...

Here is a controversial application of LP. Consider the experiment $\mathbb{E}_1$ in which a coin is flipped exactly ten times versus an experiment $\mathbb{E}_2$ in which the coin is flipped until two heads are observed. Let E be the event that two heads are observed in the first experiment and F be the event that ten tosses are made in the second. If θ is the bias of the coin, then:

$$\begin{aligned} P_\theta^{\mathbb{E}_1}(E) &= \binom{10}{2}\theta^2(1-\theta)^8 \text{ and} \\ P_\theta^{\mathbb{E}_2}(F) &= \binom{9}{1}\theta^2(1-\theta)^8. \end{aligned}$$

Letting $c = \binom{10}{2}/\binom{9}{1} = 5$, it follows that $P_\theta^{\mathbb{E}_1}(E) = c \cdot P_\theta^{\mathbb{E}_2}(F)$. Thus, LP entails that E and F are evidentially equivalent.

LP is controversial, in part, because it is intimately connected to Bayesianism. How so? The following calculation (which is well-known) shows that if an agent's credences matches likelihoods and there is some $c > 0$ such that $P_\theta^{\mathbb{E}_1}(E) = c \cdot P_\theta^{\mathbb{E}_2}(F)$ for all θ, then $Q(\cdot|E) = Q(\cdot|F)$, i.e., the agent's degrees of belief upon learning E would be same as they would be upon learning F. Let $H \subseteq \Theta$ be any (simple or composite) hypothesis. Then:

$$\begin{aligned}
Q(H|E) &= \frac{Q(E \cap H)}{Q(E)} && \text{definition of conditional probability} \\
&= \frac{\sum_{\theta \in H \cap supp(Q)} Q(E \cap \{\theta\})}{\sum_{\theta \in supp(Q)} Q(E \cap \{\theta\})} && \text{law of total probability} \\
&= \frac{\sum_{\theta \in H \cap supp(Q)} P_\theta^{\mathbb{E}_1}(E) \cdot Q(\theta)}{\sum_{\theta \in supp(Q)} P_\theta^{\mathbb{E}_1}(E) \cdot Q(\theta)} && \text{because } Q(\cdot|\theta) = P_\theta(\cdot) \text{ for all } \theta \in supp(Q) \\
&= \frac{\sum_{\theta \in H \cap supp(Q)} c \cdot P_\theta^{\mathbb{E}_2}(F) \cdot Q(\theta)}{\sum_{\theta \in supp(Q)} c \cdot P_\theta^{\mathbb{E}_2}(F) \cdot Q(\theta)} && \text{as } P_\theta^{\mathbb{E}_1}(E) = c \cdot P_\theta^{\mathbb{E}_2}(F) \\
&= \frac{\sum_{\theta \in H \cap supp(Q)} P_\theta^{\mathbb{E}_2}(F) \cdot Q(\theta)}{\sum_{\theta \in supp(Q)} P_\theta^{\mathbb{E}_2}(F) \cdot Q(\theta)} && \text{canceling } c \\
&= Q(H|F) && \text{reversing the above steps}
\end{aligned}$$

This is why Bayesianism is often said to entail LP.

To motivate my framework, it will help to introduce some terminology. Suppose E and F are outcomes of experiments $\mathbb{E}_1$ and $\mathbb{E}_2$ respectively; I assume neither that $\mathbb{E}_1 \neq \mathbb{E}_2$ nor that $E \neq F$. Say E and F are **Bayesian posterior equivalent** relative to Θ if $Q(H|E) = Q(H|F)$ for all hypotheses $H \subseteq \Theta$ and all *all* priors Q that match likelihoods, i.e., *all* Bayesian agents' would update on E in the same they would be upon learning F. As shown above, if LP entails that two events E and F are evidentially equivalent (i.e., if $P_\theta^{\mathbb{E}_1}(E) = c \cdot P_\theta^{\mathbb{E}_2}(F)$ for some $c > 0$ and all θ), then E and F are Bayesian posterior equivalent. The converse is also true.

Theorem 3 *The following three statements are equivalent:*

1. LP *entails that E and F are evidentially equivalent.*

2. *E and F are Bayesian posterior equivalent.*

3. *For any two disjoint hypotheses H_1 and H_2, E Bayesian favors H_1 to H_2 at least as much as F and vice versa (i.e., $Q(H_1|E \cap (H_1 \cup H_2)) = Q(H_1|F \cap (H_1 \cup H_2))$ for all priors Q that match likelihoods).*

See the appendix for a proof. Although the theorem is mathematically simple, it's philosophically significant. The theorem says that LP characterizes precisely when all Bayesians will agree that two pieces of evidence are equivalent.[12] In other words, if the objectivity of LP is grounded in its concordance among Bayesians, then *there are no other principles characterizing evidential equivalence*. Furthermore, the theorem shows that LP, like LL, is in a sense derivable from a more general principle: that one piece of evidence for a hypothesis is stronger than another when it raises all Bayesian agent's degrees of belief to a greater extent (i.e., when one piece of evidence Bayesian favors the hypothesis more than another).

Like the definition of Bayesian favoring, the definition of Bayesian evidential equivalence captures strong intuitions about the value of larger samples.

Example 2: Imagine you find an urn with balls with k many different colors, but you do not know how many balls of each color are in the urn. So you're interested in testing various hypotheses of the form "Half the balls are yellow; one quarter are red, and one quarter are blue." Let $\mathbb{E}_1$ be the experiment in which n many balls will be drawn from the urn, and let $\mathbb{E}_2$ be the experiment in which $m > n$ many balls will be drawn. Then no sample from experiment $\mathbb{E}_1$ is Bayesian posterior equivalent to a sample from $\mathbb{E}_2$.

2.5 Likelihoodism, Bayesianism, and Objectivity

My argument for the objectivity of LL and LP differs in important ways from existing defenses of the objectivity of frequentist and Bayesian infer-

[12] A caveat: LP concerns experiments with the same parameter space Θ. One might wish to draw evidential comparisons between outcomes of experiments $\mathbb{E}_1$ and $\mathbb{E}_2$ with distinct parameter spaces $\Theta \times \Gamma_1$ and $\Theta \times \Gamma_2$ if (i) if Γ_1 and Γ_2 are sets of nuisance parameters that affect the likelihood functions in $\mathbb{E}_1$ and $\mathbb{E}_2$, but (ii) the experimenter is only interested in which member of Θ is actual. The non-informative nuisance parameter principle is the only evidential principle I know linking experiments with different parameter spaces [Berger and Wolpert, 1988]; more research is needed here.

ence. Drawing on [Lloyd, 1995], Douglas [2009] distinguishes eight senses of "objectivity" that are sought in science. In Douglas' terminology, I have argued LL and LP could be used in highly concordantly objective processes. A process is **concordantly objective** if it is guaranteed to produce some type of *agreement* among scientists. Theorem 1 entails that, if a committee of Bayesian scientists agree to take a particular action if the evidence gathered in an experiment (Bayesian) favors H_1 over H_2, then it suffices for the lead experimenter to announce what LL says – and forgo describing all of the data – in order for the scientists to reach an agreement.

Statisticians and philosophers frequently defend the concordant objectivity of Bayesian reasoning in a different way, namely, by appealing to "merging of opinions" theorems [Blackwell and Dubins, 1962, Huttegger, 2015, Kalai and Lehrer, 1994, Schervish and Seidenfeld, 1990, Stewart and Nielsen, 2018]. Roughly, these theorems entail that if two Bayesian agents assign positive probability to the same events, then the two agents are certain that, as data accumulates, their posterior probabilities will "merge" to assigning probability one to the same set of events in the limit. Savage [1967, p. 316] summarizes the theorems as follows:

> *[Subjective Bayesianism] does thus require holders of extremely diverse sets of opinions to agree closely with one another when presented with suitable common evidence. This alone seems to me an adequate model of the objectivity of scientific knowledge.*

Importantly, such theorems concern *asymptotic* agreement, and although one can sometimes use the proofs to extract constructive bounds on sample size by which certain levels of agreement are reached, no philosophers, to my knowledge, have shown that merging of opinions is likely given the Bayesian models used in practice. In contrast, although my argument promises a far weaker type of agreement (as the agents need not agree upon how likely various events are), Theorems 3 and 1 make no assumptions about the indefinite collection of data and therefore apply to experiments with even small samples.

Unfortunately, my arguments thusfar provide little reason to believe likelihoodist methods are objective in two other senses in which other statistical methods are thought to be.

First, merging-of-opinions theorems ought to be distinguished from **Bayesian consistency theorems**, which assert that, in a wide variety of circumstances, if data is collected indefinitely, a *single* Bayesian agent's posterior will concentrate on the *true* hypothesis, as long as she is sufficiently openminded.[13] Of course, consistency theorems can be used to derive merging-of-opinions theorems: two agents who eventually both believe the truth will also eventually agree. But the converse need not hold. Consistency theorems, therefore, undergird an argument that Bayesian reasoning has **manipulable objectivity** in Douglas' terminology: they show that, with enough data, Bayesians' credences will match the world in a particular way. Classical/frequentist methods are typically defended in the same way, i.e., by showing an estimator is unbiased, consistent, etc.

Although my arguments thusfar do not support the claim that likelihoodist methods possess manipulable objectivity in this sense, the the maximum likelihood estimator (MLE) is consistent in many settings [Casella and Berger, 2002, Theorem 10.1.12] and the MLE is a standard likelihoodist method. There is also growing research characterizing in what sense the likelihood ratio is a "truth-tracking" measure of evidence [Lele, 2004]. My results do not, I conjecture, suggest that one is obliged to use any particular type of estimator, however. More research is necessary here.

Second, Sprenger and Hartmann [2019, §11.5] argue that Bayesian reasoning possesses **interactive objectivity**, which is similar to concordant objectivity but requires agents to reach consensus via dialogue, debate, and negotiation. Employing an interesting case study of precognition research, Sprenger and Hartmann argue that the Bayesian practice of making one's prior probability function public allows scientists to more quickly identify methodological assumptions, to pinpoint errors, to find sources of disagreement, to propose alternative methods for analyzing data, and to correct any problems in others' reasoning.

I do not see a way of defending LP and LL as contributing to an interactively objective process. Likelihood functions are often shared, and that suggests that the acceptance of a likelihood function is the *result* of an interactively objective process in which scientists ascertain which statistical hypotheses follow from their joint theoretical commitments. A close case study of liklihoodist methods in practice – like Sprenger and Hartmann's case study of Bayesianism – is necessary here.

2.6 Objectivity and Coherence

Philosophers who are skeptical of Bayesian coherence norms, however, might think all of the arguments thusfar concern only a very thin type of objec-

[13] See [Savage, 1972, §3.6, pp. 47-49] and [Diaconis and Freedman, 1986a]. For conditions under which Bayesian updating is not consistent, see [Freedman and Diaconis, 1982], [Freedman and Diaconis, 1983],[Diaconis and Freedman, 1986b], and [Diaconis et al., 1998].

tivity. Because Bayesian rationality is overly demanding, the community of agents that are required to endorse LP and LL is rather small. In what ways is Bayesianism too demanding?

Bayesianism famously requires logical omniscience. How so? The requirement that one's credences are representable by a *function* entails that one's credence in an event ought not depend upon how the event is described; so rational agents must recognize logically equivalent descriptions of the same event, no matter how complex the equivalence.[14] Of course, Bayesians recognize that logical omniscience is an idealization. Fallible agents, like you and me, hopefully assign increasingly higher credence to more tautologies over time and become more rational by Bayesian standards. But few philosophers would hold the average person epistemically blameworthy for failing to recognize that some complex mathematical theorem is a logical consequence of the axioms of set theory.

Recognizing that certain Bayesian norms are idealizations, however, raises a problem for my defense of the objectivity of LP and LL. If few people meet Bayesian constraints, then showing that all Bayesians would endorse LL and LP is not convincing evidence that those statistical principles are highly intersubjective. In the next section, therefore, I argue that there are qualitative analogs of LL and LP that can be endorsed by agents meeting weak coherence norms that dot not require neither logical omniscience (as degrees of belief need not be normalized); the analog of LP does not even require that one's credences be "additive" in the way probabilities are.

3 Qualitative Likelihoodism

To develop qualitative versions of LL and LP, I simply replace two types of probability functions – likelihoods and priors — with qualitative orders satisfying particular properties.

Let $\mathcal{F}$ be an algebra over $\Theta \times \Omega_{\mathbb{E}}$; the set $\mathcal{F}$ represents the events over which agents have credences. Suppose $\mathcal{F}_{\Omega_{\mathbb{E}}}$ is an algebra on $\Omega_{\mathbb{E}}$ such that $\{\Theta \times A : A \in \mathcal{F}_{\Omega_{\mathbb{E}}}\} \subseteq \mathcal{F}$. Here, $\mathcal{F}_{\Omega_{\mathbb{E}}}$ represents the set of observable events over which the agent has credences, rather than the events concerning the hypotheses Θ. I assume there is a weak order (i.e., a reflexive, total, and transitive binary relation) $\sqsubseteq$ on $\mathcal{F}_{\Omega_{\mathbb{E}}} \times \Theta$. The idea is that $A|\theta \sqsubseteq B|\eta$

[14] Kolmogorov's axioms concern functions on σ-algebras, but many philosophers state probability axioms for functions of sentences in some formal language (e.g., [Talbott, 2013]). When axiomatized in the latter way, logical omniscience is entailed by the additivity axiom and the requirement that all tautologies are assigned probability one; together, those axioms require that logically equivalent sentences be assigned the same probabilities.

represents the claim that "experimental outcome B is at least as likely under supposition η as outcome A is under supposition θ." I write $A|\theta \equiv B|\theta$ if $A|\theta \sqsubseteq B|\theta$ and vice versa. The relation $\sqsubseteq$ is the qualitative analog of the set of likelihood functions in an experiment.

Bayesians assume that an agent's conditional credences are representable conditional probabilities. I weaken that assumption and assume that agents' credences are representable by a weak ordering $\leq$. More precisely, let $\Delta = \Theta \times \Omega_{\mathbb{E}}$. I assume there is some set of events $\mathcal{N} \subseteq \mathcal{F}$ that the agent thinks are maximally unlikely; $\mathcal{N}$ stands for "null." I will assume beliefs are represented by a binary relation $\leq$ on $\mathcal{F} \times (\mathcal{F} \setminus \mathcal{N})$ satisfying a soon-to-be specified subset of the axioms below. The idea is that a relation $\leq$ represents an agent's *suppositional beliefs*, in the sense that $A|B \leq C|D$ if the agent regards C as at least as likely under supposition D as A would be under supposition B.

Here are some plausible coherence axioms that for an agent's suppositional beliefs. For all $A, B, C \in \mathcal{F}$ (or $\mathcal{F} \setminus \mathcal{N}$ if the symbol appears to the right of the $|$ sign)

C0. $A|\theta \leq B|\eta$ if and only if $A|\theta \sqsubseteq B|\eta$, i.e., agents agree upon likelihoods,

C1. $\leq$ is a weak ordering,

C2. $\Delta \in \mathcal{F} \setminus \mathcal{N}$; and $A \in \mathcal{N}$ if and only if $A|\Delta \sim \emptyset|\Delta$

- Here, $A|B \sim C|D$ if and only if $A|B \leq C|D$ and vice versa.

C3. A. $A|A \sim B|B$, and

B. $\Delta|\Delta \geq A|B$.

C4. $A \cap B|B \sim A|B$,

C5. Suppose $A \cap B = A' \cap B' = \emptyset$. If $A|C \geq A'|C'$ and $B|C \geq B'|C'$, then $A \cup B|C \geq A' \cup B'|C'$; moreover, if each hypothesis is $>$, then the conclusion is $>$.

C6. Suppose $C \subseteq B \subseteq A$ and $C' \subseteq B' \subseteq A'$. If $B|A \leq C'|B'$ and $C|B \leq B'|A'$, then $C|A \leq C'|A'$. Further, the inequality in the consequent of that conditional is strict if either inequality in the antecedent is strict.

C7. Every standard sequence is finite, where a sequence $A_1, A_2, \ldots$ is **standard** if (i) $A_n \subset A_{n+1}$ for all n, and (ii) $A_1|A_2 \sim A_n|A_{n+1} < \Delta|\Delta$ for all n.

As noted above, statisticians typically assume that likelihood functions are widely shared, which is why axiom C0 requires that agents share the qualitative likelihood ordering $\sqsubseteq$. The remaining axioms C1-C7 are [Krantz et al., 2006]'s axioms for qualitative conditional probability, and it is easy to see they are necessary for the ordering $\leq$ to be representable as a conditional probability function. A well-known result in [Kraft et al., 1959] shows the axioms are not sufficient for the ordering to have a probabilistic representation.

In the last section, I explained why many philosophers (including me) believe that Bayesian coherence requirements are too strict. Some of the above axioms are closely analogous to the axioms for probability theory, and so my goal is to derive qualitative versions of LL and LP using as few of the axioms as possible. Which are most suspect?

Axiom C3B is perhaps most questionable; it is the analog of the thesis that rational agents must assign probability one to all tautologies, which would entail that rational agents must never be unsure about the truth of a mathematical theorem, no matter how complex. Axiom C5 is likewise suspect, as it's analogous to the probability axiom that $P(\varphi \vee \psi) = P(\varphi) + P(\psi)$ when φ and ψ are inconsistent. One should expect C5 is most problematic, therefore, in conjunction with Axiom C3B. Why? If one assumes (i) $P(\varphi) = 1$ for all tautologies φ and (ii) $P(\varphi \vee \psi) = P(\varphi) + P(\psi)$ when φ and ψ are inconsistent, it follows that $P(\varphi) = P(\psi)$ for any two logically equivalent formulae φ and ψ, even when φ and ψ are *contingent*. That is the requirement that rational agents must be logically omniscient. So one might expect the conjunction of C5 and C3B to have similarly implausible implications.

Axioms C1 and C7 are also questionable, though less-so than C3B and C5. Axiom C1 (together with C0) entails that qualitative likelihoods are totally ordered, i.e., that $A|\theta \leq B|\eta$ or $B|\eta \leq A|\theta$ for any two simple hypotheses θ and η . But as noted above, it's not clear that agents will always agree that two events are comparable in probability, let alone which event is more likely. Axiom C7, which is called an Archimedean principle, entails that it's irrational for an agent to assign "infinitely" greater probability to some events rather than others. While it's not entirely obvious, Archimedean principles like C7 often require rejecting widely-accepted weak-dominance principles in decision theory [Pedersen, 2014].

In the following two subsections, I show that one can derive qualitative versions of LL and LP that forgo the use of these controversial coherence constraints.

3.1 The Qualitative Law of Likelihood

LL asserts that, when H_1 and H_2 are simple hypotheses, E favors H_1 to H_2 if $P_{H_1}(E) > P_{H_2}(E)$. The obvious qualitative analog, therefore, is the following:

Qualitative Law of Likelihood (QLL): If H_1 and H_2 are simple hypotheses, then E favors H_1 to H_2 if $E|H_1 > E|H_2$.

In the quantitative case, I showed LL was equivalent to Bayesian favoring. Recall, I motivated the definition of Bayesian favoring by appealing to the intuition that evidence E favors one hypothesis H over another if and only if E raises all rational agents' confidence in H. To try to capture that intuition, I defined E to Bayesian favor H_1 to H_2 if $Q(H_1|E \cap (H_1 \cup H_2)) \geq QH_1|H_1 \cup H_2)$ for all priors Q that match likelihoods. That definition of "favoring", however, is plausible only if one assumes that all rational agents satisfy Bayesian coherence norms. But an analogous definition might work. Let $\mathcal{C}$ be a set containing C0 and some subset of the axioms C1-C7. Say $E \in \mathcal{F}_{\Omega_{\mathbb{E}}}$ **$\mathcal{C}$-favors** H_1 to H_2 if $H_1|E \cap (H_1 \cup H_2) \geq H_1|H_1 \cup H_2$ for all orderings $\leq$ satisfying the axioms in $\mathcal{C}$. The goal is characterize $\mathcal{C}$-favoring (for various subsets of the axioms C1-C6) using only features of the shared likelihood relation $\sqsubseteq$.

Theorem 4 *Suppose Θ is finite and that $H_1, H_2 \subseteq \Theta$ are disjoint hypotheses. Let $\mathcal{C}$ contain the axioms C0, C1, C3A and C4-C6. If $E|\theta_2 \sqsubset E|\theta_1$ for all $\theta_1 \in H_1$ and $\theta_2 \in H_2$, then E $\mathcal{C}$-favors H_1 to H_2. In the special case in which $H_1 = \{\theta_1\}$ and $H_2 = \{\theta_2\}$ are simple hypotheses, therefore, E $\mathcal{C}$-favors H_1 to H_2 if* QLL *entails that E favors H_1 to H_2 (i.e., if $E|H_2 \sqsubset E|H_1$).*

The reader should compare theorems 4 and 1. I conjecture the converse of theorem 4 is true, but I do not yet have a proof.

Although theorem 4 makes essential use of C5, it avoids the other problematic coherence constraints of Bayesianism. First, the proof does not require that Axiom C2, and so it does not require that Δ *has* a probability, let alone that its probability is maximal as Axiom C3B asserts. In fact, the conjunction of axioms C0, C1, C3A and C4-C6 is compatible with the claim that there is no maximal degree of conditional probability. Second, although the proof does use Axiom C1, it does so in a rather limited fashion: one need only assume that agents' suppositional beliefs are total over the smallest algebra of events containing $\{E\} \cup H_1 \cup H_2$. In other words, the proof requires only that agents agree on how likely the evidence E given H_1

and H_2, and it does not require that agents agree about the likelihood of other events. Finally, it makes no use of Axiom C7 at all and is therefore compatible with endorsing weak-dominance principles.

3.2 A Qualitative Likelihood Principle

Two experimental outcomes E and F were defined to be Bayesian evidential equivalent if all agents satisfying Bayesian axioms for rationality would update their degrees of belief upon learning E in the same way they would upon learning F. By analogy, let $\mathcal{C}$ be a set containing C0 and some subset of the axioms C1-C7. Say $A, B \in \mathcal{F}_{\Omega_{\mathbb{E}}}$ are **$\mathcal{C}$-evidentially equivalent** if, for all orderings $\leq$ satisfying the axioms in $\mathcal{C}$, we have $\theta|A \leq \eta|A$ if and only if $\theta|B \leq \eta|B$ for all $\theta, \eta \in \Theta$. In other words, A and B are $\mathcal{C}$-evidentially equivalent if all agents satisfying the rationality axioms in $\mathcal{C}$ agree that their judgments about the various hypotheses will be the same whether they learn A or B. The goal is characterize $\mathcal{C}$-evidential equivalence (for various subsets of the axioms C0-C6) using only features of the shared likelihood relation $\sqsubseteq$.

Example 3: Suppose $\mathcal{C}$ is any collection of orderings satisfying C0, C1, C4, and C6. We claim that if $A|\theta \equiv B|\theta$ for all $\theta \in \Theta$, then A and B are $\mathcal{C}$ evidentially equivalent.

Example 4: Suppose $\mathcal{C}$ is the set of *all* orderings satisfying C0, C1, C4, and C6. Suppose further that $A|\theta \sqsubset B|\theta$ and $B|\theta' \sqsubset A|\theta'$. Then A and B are not $\mathcal{C}$-evidentially equivalent.

See the appendix for proofs of the conclusions of both examples. Before proving a qualitative analog of Theorem 3, I must introduce an alternative to Axiom 6, which I think is more intuitive than C6.

Axiom C6': Suppose $C \subseteq B \subseteq A$ and $C' \subseteq B' \subseteq A'$. If $B|A \leq B'|A'$ and $C|B \leq C'|B'$, then $C|A \leq C'|A'$. Further, the inequality in the consequent of that conditional is strict if either inequality in the antecedent is strict.

Axiom C6' is derivable from Axioms C1-C6, and it's just as "axiomatic" as C6. So I will make use of it below. To motivate a qualitative version of LP, suppose that $\{G_\theta\}_{\theta \in \Theta}$ is a collection of events such that

I. $P_\theta(E) = P_\theta(F \cap G_\theta)$ for all θ,

II. F and G_θ are independent with respect to P_θ for all $\theta \in \Theta$, and

III. $P_\theta(G_\theta) = P_\eta(G_\eta) > 0$ for all $\theta, \eta \in \Theta$.

By III, there is a single number $c > 0$ that is equal to $P_\theta(G_\theta)$ for all $\theta \in \Theta$. Assumptions I and II then imply that $P_\theta(E) = P_\theta(F \cap G_\theta) = P_\theta(F) \cdot P_\theta(G_\theta) = c \cdot P_\theta(F)$ for all θ. So by LP, the events E and F are evidentially equivalent. If we can find qualitative analogs of assumptions (I) - (III), then we might likewise be able to find conditions characterizing evidential equivalence for qualitative conditional probabilities satisfying C0-C6.

With this motivation, given an ordering $\leq$ and three events X, Y, and Z, define $Y \perp_Z^{\leq} X$ if $Y|X \cap Z \sim Y|Z$. Here, $X \perp_Z^{\leq} Y$ is analogous to the probabilistic claim that X and Y are conditionally independent given Z.

We're now in a position to prove that there is a qualitative analog of Theorem 3. Recall that $A|\theta \equiv B|\eta$ means $A|\theta \sqsubseteq B|\eta$ and vice versa. So Conditions 1 and 3 in the following theorem concern the shared qualitative likelihoods; only Condition 2 is about qualitative priors.

Theorem 5 *Let $\mathcal{C}$ be a collection of orderings satisfying C0-C2, C4, and C6'. Suppose that $\{C_\theta\}_{\theta \in \Theta}$ is a collection of sets of $\mathcal{F}_{\Omega_{\mathbb{E}}}$ such that*

1. *$A|\theta \equiv B \cap C_\theta|\theta$ for all $\theta \in \Theta$,*

2. *$B \perp_\theta^{\leq} C_\theta$ for all $\theta \in \Theta$ and all $\leq \in \mathcal{C}$ (i.e., all agents in $\mathcal{C}$ regard B and C_θ as conditionally independent given θ),*

3. *$C_\theta|\theta \equiv C_\eta|\eta$ for all $\theta, \eta \in \Theta$.*

Then A and B are $\mathcal{C}$-evidentially equivalent.

See the appendix for a proof. The reader should compare the three conditions of the theorem to conditions I-III above.

Notice that the axioms necessary for Theorem 5 include neither a qualitative additivity axiom (C5) nor the qualitative analog of the probabilistic axiom requiring the entire space to have probability one (C3), and those are the axioms that entail that Bayesian agents are logically omniscient when probabilities must be assigned to formulae in a formal language.

There are at least two disanalogies between theorems 3 and 5. First, theorem 3 is a biconditional. In contrast, theorem 5 provides only sufficient conditions for $\mathcal{C}$-evidential equivalence, and the converse of the theorem is false. Why? The existence of the events $\{C_\theta\}_{\theta \in \Theta}$ in theorem 5 amounts to a "richness" requirement on the algebra of events. In the quantitative case, two pieces of evidence A and B might have proportional likelihood functions without there being a set of events with the properties attributed to $\{C_\theta\}_{\theta\in\Theta}$. However, if one assumes a fairly strong richness requirement on the algebra of events, the converse of theorem 5 is true, but I remain agnostic to its importance.[15]

Second, LP characterizes when pieces of evidence from possibly different experiments are equivalent, whereas theorem 5 concerns the outcomes of a *single* experiment.

4 Conclusions and Future Research

I have argued that LL and LP can be understood as objective evidential principles, in the sense that a community of agents who satisfy weak rationality axioms and who share judgments about likelihood functions can use LL and LP to come to an agreement about (a) when evidence favors one hypothesis over another and (b) when two pieces of evidence are equivalent. My results are preliminary, and so technically-inclined readers, I hope, will be excited by the number of philosophically-motivated and mathematically-tractable questions that are left open by the above results.

For readers who are sympathetic to Bayesian confirmation theory, one might wish to ask whether my framework can be used to motivate particular numerical measures of confirmation in restricted settings. For readers who, like me, think Bayesian coherence norms are implausible, the number of open questions is enormous. What are necessary conditions for $\mathcal{C}$-favoring and $\mathcal{C}$-evidential equivalence for various subsets of the axioms above? What is the quantitative analog of theorem 2, which characterizes when a piece of evidence E favors H_1 over H_2 *more than* F favors H_1 over H_2? Do other common statistical principles (e.g., sufficiency, conditionality, and the stopping rule principle) have qualitative analogs, and what coherence axioms are necessary for deriving them? And can other defenses of the objectivity of Bayesian methodology – for example, those grounded in merging-of-opinion and consistency theorems – likewise be extended to agents meeting weaker coherence constraints?

References

James Orvis Berger and Robert L. Wolpert. The likelihood principle. In *Ims Lecture Notes-Monograph*, volume 6. Institute of Mathematical Statistics, 1988.

[15]Contact the author for a statement and proof.

David R. Bickel. The strength of statistical evidence for composite hypotheses: Inference to the best explanation. *Statistica Sinica*, pages 1147–1198, 2012.

Allan Birnbaum. On the foundations of statistical inference. *Journal of the American Statistical Association*, 57(298):269–306, 1962.

David Blackwell and Lester Dubins. Merging of opinions with increasing information. *The Annals of Mathematical Statistics*, 33(3):882–886, 1962.

George Casella and Roger L. Berger. *Statistical inference*, volume 70. Duxbury Press Belmont, CA, second edition, 2002.

Jake Chandler. Contrastive confirmation: some competing accounts. *Synthese*, 190(1):129–138, 2013.

Bruno De Finetti. Foresight: its logical laws in subjective sources. In H.E. Kyburg and H.E. Smokler, editors, *Studies in Subjective Probability*. 1937.

Persi Diaconis and David Freedman. On the consistency of Bayes estimates. *The Annals of Statistics*, pages 1–26, 1986a.

Persi Diaconis and David Freedman. On inconsistent Bayes estimates of location. *The Annals of Statistics*, pages 68–87, 1986b. URL `http://www.jstor.org/stable/2241268`.

Persi W. Diaconis, David Freedman, and others. Consistency of Bayes estimates for nonparametric regression: normal theory. *Bernoulli*, 4(4):411–444, 1998. URL `http://projecteuclid.org/euclid.bj/1173883814`.

Heather Douglas. *Science, policy, and the value-free ideal*. University of Pittsburgh Press, 2009.

Anthony William Fairbank Edwards. *Likelihood*. CUP Archive, 1984.

Michael Evans. *Measuring Statistical Evidence Using Relative Belief*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Chapman and Hall, June 2015.

Peter C. Fishburn. The axioms of subjective probability. *Statistical Science*, 1(3):335–358, 1986. URL `http://projecteuclid.org/euclid.ss/1177013611`.

Ronald A. Fisher. Statistical methods and scientific inference. 1956.

Branden Fitelson. Likelihoodism, Bayesianism, and relational confirmation. *Synthese*, 156(3):473–489, 2007.

D. A. Freedman and Persi Diaconis. On inconsistent M-estimators. *The Annals of Statistics*, pages 454–461, 1982. URL `http://www.jstor.org/stable/2240679`.

David Freedman and Persi Diaconis. On inconsistent Bayes estimates in the discrete case. *The Annals of Statistics*, pages 1109–1118, 1983.

Andrew Gelman and Christian Hennig. Beyond subjective and objective in statistics. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(4):967–1033, 2017.

Ian Hacking. *Logic of statistical inference*. Cambridge University Press, 1965.

Simon M. Huttegger. Merging of opinions and probability kinematics. *The Review of Symbolic Logic*, 8(4):611–648, 2015.

Ehud Kalai and Ehud Lehrer. Weak and strong merging of opinions. *Journal of Mathematical Economics*, 23(1):73–86, 1994.

John Maynard Keynes. *A Treatise on Probability*. Courier Dover Publications, 2004.

Bernard O. Koopman. The axioms and algebra of intuitive probability. *Annals of mathematics*, pages 269–292, 1940a.

Bernard O. Koopman. The bases of probability. *Bulletin of the American Mathematical Society*, 46(10):763–774, 1940b.

Charles H. Kraft, John W. Pratt, and Abraham Seidenberg. Intuitive probability on finite sets. *The Annals of Mathematical Statistics*, pages 408–419, 1959.

David H. Krantz, R. Duncan Luce, Patrick Suppes, and Amos Tversky. *Foundations of Measurement Volume II: Geometrical, Threshold, and Probabilistic Representations*. Dover Publications, Mineola, N.Y, December 2006. ISBN 978-0-486-45315-6.

Subhash Lele. Evidence Functions and the Optimality of the Law of Likelihood. *The nature of scientific evidence: Statistical, philosophical, and empirical considerations*, pages 191–216, 2004.

Elisabeth A. Lloyd. Objectivity and the double standard for feminist epistemologies. *Synthese*, 104(3):351–381, 1995.

Helen E. Longino. *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton University Press, 1990.

A.P. Pedersen. Comparative Expectations. *Unpublished Manuscript*, 2014.

Elizabeth Potter. *Feminism and philosophy of science: An introduction*. Routledge, 2006.

Richard Royall. *Statistical evidence: a likelihood paradigm*, volume 71. Chapman & Hall/CRC, 1997.

L. J. Savage. *The foundation of statistics*. Dover publications, 1972.

Leonard J. Savage. Implications of personal probability for induction. *The Journal of Philosophy*, pages 593–607, 1967. URL `http://www.jstor.org/stable/2024536`.

Mark Schervish and Teddy Seidenfeld. An approach to consensus and certainty with increasing evidence. *Journal of Statistical Planning and Inference*, 25(3):401–414, 1990.

Elliott Sober. *Ockham's razors*. Cambridge University Press, 2015.

Elliott Sober and Mike Steel. Similarities as evidence for common ancestry: a likelihood epistemology. *The British Journal for the Philosophy of Science*, page axv052, 2015.

Jan Sprenger and Stephan Hartmann. *Bayesian Philosophy of Science*. Oxford University Press, 2019. URL `http://www.laeuferpaar.de/BPS/BPS-SprengerHartmann.pdf`.

Rush Stewart and Michael Nielsen. Another Approach to Consensus and Maximally Informed Opinions with Increasing Evidence. 2018.

William Talbott. Bayesian Epistemology. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Fall 2013 edition, 2013. URL `http://plato.stanford.edu/archives/fall2013/entries/epistemology-bayesian/`.

Veronica J. Vieland. Measurement of Statistical Evidence: Picking Up Where Hacking and Others Left Off. *Philosophy of Science*, 84(5):853–865, 2017.

Larry Wasserman. Statistical Principles?, July 2012. URL https://normaldeviate.wordpress.com/2012/07/28/statistical-principles/.

5 Appendix

5.1 Quantitative versions of LL and LP

Proof of Theorem 1: Suppose that $H_1, H_2 \subseteq \Theta$ are two (possibly composite hypotheses) such that $H_1 \cap H_2 = \emptyset$. Our goal is to show the following two assertions are equivalent:

1. $Q(H_1|E \cap (H_1 \cup H_2)) > Q(H_1|H_1 \cup H_2)$ for all priors Q that match likelihoods, and
2. $\min_{\theta \in H_1} P_\theta(E) > \max_{\theta \in H_2} P_\theta(E)$.

Proof of 1 ⇒ 2: Let $\theta_1 \in H_1$ be such that $P_{\theta_1}(E) \leq P_\theta(E)$ for all $\theta \in H_1$. Similarly, let $\theta_2 \in H_2$ be such that $P_{\theta_2}(E) \geq P_\theta(E)$ for all $\theta \in H_2$. Consider the prior Q such that $Q(\theta_1) = Q(\theta_2) = 1/2$. Thus, $Q(\{\theta_1, \theta_2\}) = 1$ and so

$$\begin{aligned} Q(H_i) &= Q(\theta_i) \text{ for } i = 1, 2 \\ Q(\theta_1|E \cap \{\theta_1, \theta_2\}) &= Q(\theta_1|E) \text{ and} \\ Q(\theta_1|\{\theta_1, \theta_2\}) &= Q(\theta_1) = 1/2 \end{aligned}$$

Thus, by assumption it follows that

$$\dagger \quad Q(\theta_1|E) = Q(H_1|E \cap (H_1 \cup H_2)) > Q(H_1|H_1 \cup H_2) = 1/2.$$

Now

$$\begin{aligned} Q(H_1|E) &= Q(\theta_1|E) \\ &= \frac{Q(E|\theta_1) \cdot Q(\theta_1)}{Q(E)} && \text{by Bayes' theorem} \\ &= \frac{Q(E|\theta_1) \cdot Q(\theta_1)}{Q(E|\theta_1) \cdot Q(\theta_1) + (E|\theta_2) \cdot Q(\theta_2)} && \text{because } Q(\{\theta_1, \theta_2\}) = 1 \\ &= \frac{P_{\theta_1}(E) \cdot Q(\theta_1)}{P_{\theta_1}(E) \cdot Q(\theta_1) + P_{\theta_2}(E) \cdot Q(\theta_2)} && \text{because } Q \text{ matches likelihoods} \\ &= \frac{P_{\theta_1}(E)}{P_{\theta_1}(E) + P_{\theta_2}(E)} && \text{because } Q(\theta_1) = Q(\theta_2) \end{aligned}$$

The last line and the equation labeled † together entail:

$$\frac{P_{\theta_1}(E)}{P_{\theta_1}(E) + P_{\theta_2}(E)} > 1/2$$

which is equivalent to $P_{\theta_1}(E) > P_{\theta_2}(E)$, as desired.

Proof of 2 ⇒ 1: Left to reader. See [Chandler, 2013], Observation 5 for a nearly identical claim.

□

Proof of Theorem 3: The proof of 1 ⇒ 2 is in the body of the paper. We now prove 2 ⇒ 1 by contraposition. Suppose there is no $c > 0$ such that $P_\theta^{\mathbb{E}_1}(E) = c \cdot P_\theta^{\mathbb{E}_2}(F)$ for all $\theta \in \Theta$. We must find a prior Q and a hypothesis $H \subseteq \Theta$ such that $Q(H|E) \neq Q(H|F)$.

We first claim there is some $\theta \in \Theta$ such that $P_\theta^{\mathbb{E}_1}(E) > 0$ or $P_\theta^{\mathbb{E}_2}(F) > 0$. Otherwise, $P_\theta^{\mathbb{E}_1}(E) = P_\theta^{\mathbb{E}_2}(F) = 0$ for all $\theta \in \Theta$, and so $P_\theta^{\mathbb{E}_1}(E) = c \cdot P_\theta^{\mathbb{E}_2}(F)$ for all $c > 0$, contradicting assumption.

So let $\upsilon_1 \in \Theta$ be such that either $P_{\upsilon_1}^{\mathbb{E}_1}(E) > 0$ or $P_{\upsilon_1}^{\mathbb{E}_2}(F) > 0$. Without loss of generality, assume that $P_{\upsilon_1}^{\mathbb{E}_2}(F) > 0$, and let $d = P_{\upsilon_1}^{\mathbb{E}_1}(E)/P_{\upsilon_1}^{\mathbb{E}_2}(F)$. Clearly, $d \geq 0$. We consider two cases.

Case 1 ($d = 0$): Thus, $P_{\upsilon_1}^{\mathbb{E}_1}(E) = 0$. So define Q to be the probability measure such $Q(\upsilon_1) = 1$ and such that Q matches likelihoods. Then Q is clearly well-defined, and $Q(\upsilon_1|E)$ is undefined (since $Q(E) = Q(E|\upsilon_1) = P_{\upsilon_1}(E) = 0$), whereas $Q(\upsilon_1|F) = 1$ because $Q(\upsilon_1) = 1$. Thus, $Q(\upsilon_1|E) \neq Q(\upsilon_1|F)$, as desired.

Case 2 ($d > 0$): Thus, $P_{\upsilon_1}^{\mathbb{E}_1}(E) > 0$. Because there is no $c > 0$ such that $P_\theta^{\mathbb{E}_1}(E) = c \cdot P_\theta^{\mathbb{E}_2}(F)$ for all $\theta \in \Theta$, it follows there is some $\upsilon_2 \in \Theta$ such that $P_{\upsilon_2}^{\mathbb{E}_1}(E) \neq dP_{\upsilon_2}^{\mathbb{E}_2}(F)$. Since $d > 0$, either $P_{\upsilon_2}^{\mathbb{E}_1}(E) > 0$ or $P_{\upsilon_2}^{\mathbb{E}_2}(F) > 0$ (or both). Consider two subcases.

Case 2a: Suppose $P_{\upsilon_2}^{\mathbb{E}_2}(F) = 0$. Then $P_{\upsilon_2}^{\mathbb{E}_1}(E) > 0$. So define Q so that $Q(\upsilon_2) = 1$ and so that Q matches likelihoods. Again, $Q(\upsilon_2|F)$ is undefined (since $Q(F) = Q(F|\upsilon_2) = P_{\upsilon_2}^{\mathbb{E}_2}(F) = 0$), whereas $Q(\upsilon_2|E) = Q(\upsilon_2) = 1$. So $Q(\upsilon_2|E) \neq Q(\upsilon_2|F)$, as desired.

Case 2b: Suppose $P_{\upsilon_2}^{\mathbb{E}_1}(F) > 0$. For brevity, let $a_i = P_{\upsilon_i}^{\mathbb{E}_1}(E)$ and let $b_i = P_{\upsilon_i}^{\mathbb{E}_2}(F)$ for $i \in \{1, 2\}$. By assumption, $b_1, b_2 > 0$. Pick any r such that

$0 < r < 1$. Let Q be any measure that matches likelihoods and such that $r = Q(\upsilon_1) = 1 - Q(\upsilon_2)$. We claim that $Q(\upsilon_1|E) \neq Q(\upsilon_1|F)$, as desired.

Suppose for the sake of contradiction that $Q(\upsilon_1|E) = Q(\upsilon_1|F)$. By Bayes theorem:

$$Q(\upsilon_1|E) = \frac{a_1 r}{a_1 r + a_2(1-r)} \tag{1}$$

$$Q(\upsilon_1|F) = \frac{b_1 r}{b_1 r + b_2(1-r)}. \tag{2}$$

Note the denominator of $Q(\upsilon_1|E)$ is positive because $a_1, r > 0$ (by the assumption of Case 2), and the denominator of $Q(\upsilon_1|F)$ is positive because $b_1, r > 0$ (by the assumption of Case 2b). Thus, because $Q(\upsilon_1|E) = Q(\upsilon_1|F)$, we obtain that:

$$a_1 r \left(b_1 r + b_2(1-r)\right) = r b_1 \left(a_1 r + a_2(1-r)\right). \tag{3}$$

Because $b_1, b_2, r, 1-r > 0$, we can simplify (3) to show that:

$$a_1/b_1 = a_2/b_2.$$

In other words, $P^{\mathbb{E}_1}_{\upsilon_2}(E)/P^{\mathbb{E}_2}_{\upsilon_2}(F) = P^{\mathbb{E}_1}_{\upsilon_2}(E)/P^{\mathbb{E}_1}_{\upsilon_1}(F) = d$, contradicting choice of υ_2.

Next, we prove $2 \Rightarrow 3$. Suppose $Q(H|E) = Q(H|F)$ for all hypotheses H and Q matching likelihoods. We must show that $Q(H_1|E \cap (H_1 \cup H_2)) = Q(H_1|F \cap (H_1 \cup H_2))$ for all disjoint hypotheses $H_1, H_2 \subseteq \Theta$ and all Q matching likelihoods. To do so, let H_1, H_2 and Q be given. Then:

$$
\begin{aligned}
Q(H_1|E \cap (H_1 \cup H_2)) &= \frac{Q(H_1 \cap (E \cap (H_1 \cup H_2))}{Q(E \cap (H_1 \cup H_2))} \\
&= \frac{Q(E \cap H_1)}{Q(E \cap H_1) + Q(E \cap H_2)} \\
&\text{as } H_1 \cap H_2 = \emptyset \\
&= \frac{Q(H_1|E) \cdot Q(E)}{Q(H_1|E) \cdot Q(E) + Q(H_2|E) \cdot Q(E)} \\
&= \frac{Q(H_1|E)}{Q(H_1|E) + Q(H_2|E)} \\
&\text{by canceling } Q(E) \\
&= \frac{Q(H_1|F)}{Q(H_1|F) + Q(H_2|F)} \\
&\text{as } E \text{ and } F \text{ are Bayesian posterior equivalent} \\
&= Q(H_1|F \cap (H_1 \cup H_2)) \\
&\text{reversing the above calculations}
\end{aligned}
$$

Finally, we prove $3 \Rightarrow 2$. Suppose that $Q(H_1|E \cap (H_1 \cup H_2)) = Q(H_1|F \cap (H_1 \cup H_2))$ for all disjoint hypotheses $H_1, H_2 \subseteq \Theta$ and all Q matching likelihoods. We must show that E and F are Bayesian posterior equivalent, i.e., that $Q(H|E) = Q(H|F)$ for all hypotheses H and all Q matching likelihoods. To do so, let H and Q be given. Then define $H_1 = H$ and $H_2 = \Theta \setminus H$. Then

$$Q(H|E) = Q(H_1|E) = Q(H_1|E \cap \Theta) = Q(H_1|E \cap (H_1 \cup H_2)) = Q(H_1|F \cap (H_1 \cup H_2)) = Q(H|F)$$

as desired.

$\square$

5.2 Qualitative Theorems

5.2.1 Examples

For brevity, I abbreviate $X|\Delta \leq Y|\Delta$ by $X \leq Y$ below.

Lemma 1 *Let $\leq$ be a relation satisfying C1, C4, and C6'. If $Y \notin \mathcal{N}$, then $X|Y \leq W|Y$ if and only if $X \cap Y \leq W \cap Y$.*

Proof: To prove the left to right direction, we apply Axiom C6. Let $A = A' = \Delta$; $B = B' = Y$; $C = X \cap Y$ and $C' = W \cap Y$. Notice that $C \subseteq B \subseteq A$ and $C' \subseteq B' \subseteq A'$. Because $A = A'$ and $B = B'$, it immediately follows from the reflexivity of $\leq$ that $B|A \leq B'|A'$. Further:

$$\begin{aligned} C|B &= X \cap Y|Y \text{ by definition of } C\&B \\ &\sim X|Y \text{ by Axiom C4} \\ &\leq W|Y \text{ by assumption} \\ &\sim W \cap Y|Y \text{ by Axiom C4} \\ &= C'|B' \text{ by definition of } C'\&B' \end{aligned}$$

So by Axiom C6', it follows that $C|A \leq C'|A'$, i.e., that $X \cap Y|\Omega \leq W \cap Y|\Omega$, i.e., that $X \cap Y \leq W \cap Y$, as desired.

In the right to left direction, we prove the contrapositive. That is, suppose that $X|Y \not\leq W|Y$. Because the relation $\leq$ is total, it follows that $W|Y < X|Y$. One can then apply Axiom C6 in the same way we just did to show that $W \cap Y < X \cap Y$. By the definition of the $<$, it follows that $X \cap Y \not\leq W \cap Y$, as desired.

□

Proof of Example 3: For any H, C0 entails that $A|H \equiv B|H$ if and only if $A|H \sim B|H$. So by lemma 1 (which requires C4 and C6'), we knew † $A|H \equiv B|H$ if and only if $A \cap H \sim B \cap H$. Thus:

$$\begin{aligned} H|A \leq H'|A &\Leftrightarrow H \cap A \leq H' \cap A \text{ by Lemma 1} \\ &\Leftrightarrow H \cap B \leq H' \cap B \text{ by † and C1} \\ &\Leftrightarrow H|B \leq H'|B \text{ by Lemma1} \end{aligned}$$

Proof of Example 4: Let $\leq$ be an ordering such that $A \cap H < B \cap H' < B \cap H < A \cap H'$. Then lemma 1 entails $H|A < H'|A$ and $H'|B < H|B$.

5.3 Qualitative LL

Lemma 2 *Assume Axioms C1 and C3-C6. If $E|H_2 < E|H_1$ and $H_1 \cap H_2 = \emptyset$, then $E|H_1 \cup H_2 < E|H_1$.*

Proof: Suppose for the sake of contradiction that $E|H_1 \cup H_2 \not< E|H_1$. Since $\leq$ is total, that entails that $E|H_1 \cup H_2 \geq E|H_1$. Define:

$$\begin{array}{ll} A = H_1 \cup H_2 & A' = H_1 \cup H_2 \\ B = E \cap (H_1 \cup H_2) & B' = H_1 \\ C = E \cap H_1 & C' = E \cap H_1 \end{array}$$

Now $B|A = E \cap (H_1 \cup H_2)|H_1 \cup H_2 \sim E|H_1 \cup H_2$ and $C'|B' = E \cap H_1|H_1 \sim E|H_1$ by Axiom C4. Hence, $B|A > C'|B'$ by assumption of the reductio and the transitivity of $\leq$ (which is part of Axiom C1). Thus, if $C|B > B'|A'$, then by Axiom C6, it would follows that $C|A > C'|A' = C|A$, contradicting the reflexivity of $\leq$ (which is part of Axiom C1). Thus, the totality of $\leq$ entails that $C|B \leq B'|A'$, i.e., that $(*)$ $E \cap H_1|E \cap (H_1 \cup H_2) \leq H_1|H_1 \cup H_2$.

Using the assumption that $E|H_1 > E|H_2$, analogous reasoning shows that $(**)$ $E \cap H_2|H_1 \cup H_2 < H_2|H_1 \cup H_2$. Notice the inequality here is *strict*. In greater detail, let

$$\begin{array}{ll} A = H_1 \cup H_2 & A' = H_1 \cup H_2 \\ B = E \cap (H_1 \cup H_2) & B' = H_2 \\ C = E \cap H_2 & C' = E \cap H_2 \end{array}$$

Again, $B|A = E \cap (H_1 \cup H_2)|H_1 \cup H_2 \sim E|H_1 \cup H_2$ and $C'|B' = E \cap H_2|H_2 \sim E|H_2$ by Axiom C4. Now $E|H_1 \cup H_2 \geq E|H_1$ by assumption of the reductio, and $E|H_1 > E|H_2 \sim C'|B'$ by assumption. Hence, $B|A > C'|B'$ by the transitivity of $\leq$ (which is part of Axiom C1). Thus, if $C|B \geq B'|A'$, then by Axiom C6, it would follows that $C|A > C'|A' = C|A$, contradicting the reflexivity of $\leq$ (which is part of Axiom C1). Thus, the totality of $\leq$ entails that $C|B < B'|A'$, i.e., that $(**) E \cap H_2|E \cap (H_1 \cup H_2) < H_2|H_1 \cup H_2$.

Recall that $H_1 \cap H_2 = \emptyset$ by assumption, and so Axiom C5 and the equations labeled $(*)$ and $(**)$ together entail that

$$(E \cap H_1) \cup (E \cap H_2)|E \cap (H_1 \cup H_2) < H_1 \cup H_2|H_1 \cup H_2.$$

But Axiom C3 entails that

$$(E \cap H_1) \cup (E \cap H_2)|E \cap (H_1 \cup H_2) \sim H_1 \cup H_2|H_1 \cup H_2$$

and so we've derived a contradiction.

$\square$

Lemma 3

Lemma 4 *If $E|H_1, E|H_2 < E|H_3$ and $H_1 \cap H_2 = \emptyset$, then $E|H_1 \cup H_2 < E|H_3$.*

Proof: If $E|H_1 < E|H_2$, then $E|H_1 \cup H_2 < E|H_2$, by lemma 2. So $E|H_1 \cup H_2 < E|H_3$ by transitivity and assumption that $E|H_2 < E|H_3$. Similarly, if $E|H_2 < E|H_1$, then $E|H_1 \cup H_2 < E|H_1 < E|H_3$, by lemma 2 and assumption.

Because $\leq$ is total, we know that one of the following three conditions hold: (i) $E|H_1 < E|H_2$, (ii) $E|H_2 < E|H_1$, or (iii) $E|H_1 \sim E|H_2$. Since we've derived the conclusion from the first two assumptions, it remains to be shown that iii and the assumption entails that $E|H_1 \cup H_2 < E|H_3$.

Lemma 5 *Assume Axioms C1 and C3-C6. If $E|H_2 < E|H_1$ and $H_1 \cap H_2 = \emptyset$, then $H_1|E \cap (H_1 \cup H_2) > H_1|H_1 \cup H_2$.*

Proof: Suppose for the sake of contradiction that $H_1|E \cap (H_1 \cup H_2) \not> H_1|H_1 \cup H_2$. Then by the totality of $\leq$ (i.e., Axiom C1), it follows that

$$\dagger \quad H_1|E \cap (H_1 \cup H_2) \leq H_1|H_1 \cup H_2$$

Define:

$$\begin{array}{ll} A = H_1 \cup H_2 & A' = H_1 \cup H_2 \\ B = H_1 & B' = E \cap (H_1 \cup H_2) \\ C = E \cap H_1 & C' = E \cap H_1 \end{array}$$

It follows that:

$$\begin{aligned} B|A &= H_1|H_1 \cup H_2 & \\ &\geq H_1|E \cap (H_1 \cup H_2) & \text{by } \dagger \\ &\sim H_1 \cap (E \cap (H_1 \cup H_2))|E \cap (H_1 \cup H_2) & \text{by Axiom C4} \\ &= E \cap H_1|E \cap (H_1 \cup H_2) & \text{because } H_1 \cap H_2 = \emptyset \\ &= C'|B' & \end{aligned}$$

By the transitivity of $\leq$ (i.e., Axiom C1), we have $B|A > C'|B'$. Further:

$$\begin{aligned} C|B &= E \cap H_1|H_1 & \\ &\sim E|H_1 & \text{by Axiom C4} \\ &> E|H_1 \cup H_2 & \text{by Lemma 2} \\ &\sim E \cap (H_1 \cup H_2)|H_1 \cup H_2 & \text{by Axiom C4} \\ &= B'|A' & \end{aligned}$$

Again, by the transitivity of $\leq$ (i.e., Axiom C1), we have $C|B > B'|A'$. By Axiom C6, it follows that $C|A > C'|A' = C|A$, contradicting the reflexivity of $\leq$ (i.e., Axiom C1). So we're done.

$\square$

Lemma 6 *Suppose H_1 and H_2 are finite. If $E|\theta_2 \sqsubset E|\theta_1$ for all $\theta_1 \in H_1$ and $\theta_2 \in H_2$, then $E|H_2 < E|H_1$*

5.4 Qualitative LP

Lemma 7 *Let $\leq$ be an ordering satisfying C0, C1, C4, and C6'. Suppose (1) $Y_1|Z_1 \sim Y_2|Z_2$, (2) $Y_1 \perp^{\leq}_{Z_1} X$, and (3) $Y_2 \perp^{\leq}_{Z_2} X$. Then $X \cap Z_1 \leq X \cap Z_2$ if and only if $X \cap Y_1 \cap Z_1 \leq X \cap Y_2 \cap Z_2$.*

Proof: In the left to right direction, suppose $X \cap Z_1 \leq X \cap Z_2$. We'll apply Axiom C6 as follows. Let $A = A' = \Delta$, $B = X \cap Z_1$, $B' = X \cap Z_2$, $C = X \cap Y_1 \cap Z_1$ and $C' = X \cap Y_2 \cap Z_2$. Then $B|A = X \cap Z_1|\Delta$ and $B'|A' = X \cap Z_2|\Delta$, and so by assumption, $B|A \leq B'|A'$. Further:

$$\begin{aligned}
C|B &= X \cap Y_1 \cap Z_1 | X \cap Z_1 \text{ by definition of } C\&B \\
&\sim Y_1 | X \cap Z_1 \text{ by Axiom C4} \\
&\sim Y_1 | Z_1 \text{ as } Y_1 \perp^{\leq}_{Z_1} X \\
&\sim Y_2 | Z_2 \text{ by assumption (1)} \\
&\sim Y_2 | X \cap Z_2 \text{ as } Y_2 \perp^{\leq}_{Z_2} X \\
&\sim X \cap Y_2 \cap Z_2 | X \cap Z_2 \text{ by Axiom C4} \\
&= C'|B' \text{ by definition of } C'\&B'
\end{aligned}$$

Thus, by Axiom C6, it follows that $C|A \leq C'|A'$, i.e., $X \cap Y_1 \cap Z_1 \leq X \cap Y_2 \cap Z_2$ as desired.

In the right to left direction, we prove the contrapositive. Suppose that $X \cap Z_1 \not\leq X \cap Z_2$. Then by the totality of $\leq$, it follows that $X \cap Z_2 < X \cap Z_1$. Then just as in the left-to-right direction, we apply Axiom C6 to show that $X \cap Y_2 \cap Z_2 < X \cap Y_1 \cap Z_1$. From that it follows that $X \cap Y_1 \cap Z_1 \not\leq X \cap Y_2 \cap Z_2$, as desired.

$\square$

Proof of theorem 5: Let $\leq$ be any ordering in $\mathcal{C}$ and $\sim$ its corresponding equivalence relation. Lemma 1, Assumption 1, and C0 together entail that

$$\dagger\ A \cap \theta \sim B \cap C_\theta \cap \theta \text{ and } A \cap \eta \sim B \cap C_\eta \cap \eta$$

Thus:

$$\begin{aligned}
\theta|A \leq \eta|A \quad &\Leftrightarrow \quad A \cap \theta \leq A \cap \eta \text{ by Lemma 1} \\
&\Leftrightarrow \quad B \cap C_\theta \cap \theta \leq B \cap C_\eta \cap \eta \text{ by } \dagger \\
&\Leftrightarrow \quad B \cap \theta \leq B \cap \eta \text{ by Lemma 7} \\
&\Leftrightarrow \quad \theta|B \leq \eta|B \text{ by Lemma 1}
\end{aligned}$$

$\square$