



**HAL**  
open science

# The sufficiency of the evidence, the relevancy of the evidence, and quantifying both with a single number

David R. Bickel

## ► To cite this version:

David R. Bickel. The sufficiency of the evidence, the relevancy of the evidence, and quantifying both with a single number. 2019. <hal-01979351>

**HAL Id: hal-01979351**

**<https://hal.science/hal-01979351v1>**

Preprint submitted on 12 Jan 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# The sufficiency of the evidence, the relevancy of the evidence, and quantifying both with a single number

January 12, 2019

David R. Bickel  
Ottawa Institute of Systems Biology  
Department of Biochemistry, Microbiology, and Immunology  
Department of Mathematics and Statistics  
University of Ottawa; 451 Smyth Road; Ottawa, Ontario, K1H 8M5

## Abstract

Consider a data set as a body of evidence that might confirm or disconfirm a hypothesis about a parameter value. If the posterior probability of the hypothesis is high enough, then the truth of the hypothesis is accepted for some purpose such as reporting a new discovery. In that way, the posterior probability measures the sufficiency of the evidence for accepting the hypothesis. It would only follow that the evidence is relevant to the hypothesis if the prior probability were not already high enough for acceptance. A measure of the relevancy of the evidence is the Bayes factor since it is the ratio of the posterior odds to the prior odds.

Measures of the sufficiency of the evidence and measures of the relevancy of the evidence are not mutually exclusive. An example falling in both classes is the likelihood ratio statistic, perhaps based on a pseudolikelihood function that eliminates nuisance parameters. There is a sense in which the likelihood ratio statistic measures both the sufficiency of the evidence and its relevancy. That result is established by representing the likelihood ratio statistic in terms of a conditional possibility measure that satisfies logical coherence rather than probabilistic coherence.

**Keywords:** deductive closure; deductive cogency; general law of likelihood; likelihood paradigm; possibility measure; possibility theory; pure likelihood methods; strength of statistical evidence

A DOG, crossing a bridge over a stream with a piece of flesh in his mouth, saw his own shadow in the water and took it for that of another Dog, with a piece of meat double his own in size. He immediately let go of his own, and fiercely attacked the other Dog to get his larger piece from him. He thus lost both: that which he grasped at in the water, because it was a shadow; and his own, because the stream swept it away.

(*Aesop's Fables*, translated by George Fyler Townsend, Amazon Digital Services, Inc., p. 18)

## 1 Introduction

Scientists seek to publish observations that constitute sufficient evidence to accept a hypothesis as a contribution to scientific knowledge. For measuring the strength of evidence, the likelihood paradigm may have advantages over the frequentist and Bayesian paradigms (Edwards, 1992; Royall, 1997; Blume, 2011; Bickel, 2012; Rohde, 2014).

In this paradigm, the likelihood ratio serves as a measure of the strength of statistical evidence for one hypothesis over another through the lens of a family of distributions (Royall, 1997, 2000a). That differs from the more familiar uses of the likelihood function as a tool for the construction of point estimators,  $p$  values, confidence intervals, and posterior probabilities. It has been used to analyze data both in basic domains such as genetics (Strug and Hodge, 2006a,b; Strug et al., 2007; Hodge et al., 2011; Strug et al., 2010; Strug, 2018) and in more applied domains such as health care (Blume, 2002; Hoch and Blume, 2008). Rohde (2014) provides an accessible exposition of the likelihood paradigm.

The paradigm has roots in the likelihood intervals of R. A. Fisher. In a certain sense, a scalar parameter value  $\theta$  is “consistent with the observations” at some level  $\Lambda$  if and only if  $\theta^-(\Lambda) \leq \theta \leq \theta^+(\Lambda)$ , where  $[\theta^-(\Lambda), \theta^+(\Lambda)]$  is the interval of parameter values with likelihood within a factor of  $\Lambda$  of the maximum likelihood, provided that  $\Lambda > 1$  (Royall, 1997, p. 26). For example, Fisher (1973, pp. 75-76) considered  $\Lambda = 2, 5, 15$ , flagging parameter values outside the  $[\theta^-(\Lambda), \theta^+(\Lambda)]$  intervals as “implausible” and those outside even the  $[\theta^-(15), \theta^+(15)]$  interval as “obviously open to grave suspicion” (cf. Barnard, 1967; Hoch and Blume, 2008). In that context, Fisher (1973, p. 71; cf. 74-75) remarked that the  $p$  value is “not very defensible save as an approximation” (see Bickel and Patriota, 2019). Royall (1997) instead used  $\Lambda = 2^3$  for strong evidence and  $\Lambda = 2^5$  for very strong evidence; see Bickel (2014) for additional gradations. For vector parameters, the level- $\Lambda$  *likelihood set* is the set of parameter values with likelihood within a factor of  $\Lambda$  of the maximum likelihood.

Just as nested confidence sets may be inverted to define a  $p$  value for each parameter value, likelihood sets may be inverted to obtain the likelihood ratio of each parameter value relative to the maximum likelihood. Edwards (1992) and Royall (1997) interpreted the likelihood ratio as the strength of evidence, carefully limiting the scope to comparisons between simple

(point) hypotheses, in which case the Bayes factor is the likelihood ratio. According to the (*special*) *law of likelihood* attributed to Hacking (1965), the likelihood ratio between two simple hypotheses quantifies the strength of evidence of one hypothesis over the other, apart from prior distributions, loss functions, and the sample size (Edwards, 1992; Royall, 1997). This contrasts with the more generally applicable practice of measuring statistical evidence for general hypotheses with the Bayes factor (cf. Jeffreys, 1948).

The primary motivation for the limitation to simple hypotheses was to avoid specifying the prior distributions needed to define a Bayes factor for composite hypotheses, for the Bayes factor is the ratio of the likelihood means with respect to a prior distribution conditional on each hypothesis compared. To achieve applicability to composite hypotheses without a prior distribution of the interest parameter, the prior mean likelihood given each composite hypothesis was replaced with the maximum likelihood over the parameter values of each composite hypothesis. The resulting ratio of maximum likelihoods is interpreted as the weight of the statistical evidence that supports one composite hypothesis over another under the *general law of likelihood* (Bickel, 2012, §2.2.3), applicable to pseudolikelihood as well as to likelihood. For instance, the weight of the evidence substantiating the hypothesis that  $\theta^-(\Lambda) \leq \theta \leq \theta^+(\Lambda)$  over the hypothesis that  $\theta \notin [\theta^-(\Lambda), \theta^+(\Lambda)]$  is

$$W([\theta^-(\Lambda), \theta^+(\Lambda)]; [\theta^-(\Lambda), \theta^+(\Lambda)]^c) = \Lambda,$$

where the complement  $[\theta^-(\Lambda), \theta^+(\Lambda)]^c$  is  $]-\infty, \theta^-(\Lambda)[ \cup ]\theta^+(\Lambda), \infty[$ . With an eye toward clinical trials, Zhang and Zhang (2013a) recommended a special case of the general law for regular models and sufficiently large samples. Motivated by different concerns, Dubois et al. (1997), Walley and Moral (1999), Giang and Shenoy (2005), and Coletti et al. (2009) had previously considered a general form of  $L([\theta^-(\Lambda), \theta^+(\Lambda)])$ , the normalized maximum likelihood of the hypothesis that  $\theta^-(\Lambda) \leq \theta \leq \theta^+(\Lambda)$ .

Although the general law of likelihood overcomes multiplicity paradoxes without resorting to a prior distribution and has been applied to genomics data (Bickel, 2012, 2014) and genetics data (Strug, 2018), it remains controversial. Blume (2013), while advocating the special law of likelihood, does not recognize a need for assigning a strength of evidence to a composite hypothesis, maintaining that the level- $\Lambda$  likelihood set simply indicates which distributions are better supported than the others by the data (cf. Zhang and Zhang, 2013b).

Further, it is often thought that the likelihood ratio cannot be directly compared to a fixed threshold  $\Lambda$  but that it requires calibration (Severini, 2000; Kalbfleisch, 2000; Morgenthaler and Staudte, 2012; Spanos, 2013). For example, Vieland and Seok (2016) made several adjustments to the case of  $L(\bullet)$  defined in Zhang and Zhang (2013a). Frequentist calibrations include those that Bickel (2018a) bases on the fixed-confidence likelihood intervals of Spratt (2000, §5.3), and Patriota (2013) proposed a quantity based on the likelihood ratio test. Frequentist calibration would indeed be needed to achieve specified repeated-sampling coverage rates since a level- $\Lambda$  likelihood set can cover the true value of the parameter with

much less than 95% frequentist probability. Likewise, from a Bayesian perspective, a level- $\Lambda$  likelihood set can have a very low posterior probability.

Largely due to those concerns, the most commonly used extension of the special law of likelihood to composite hypotheses is the Bayes factor rather than the general law's  $W(\bullet; \bullet)$ . Being defined as the posterior odds divided by the prior odds, the Bayes factor captures the intuitive appeal the special law. Indeed, Edwards (1992) commended the special law for its compatibility with data analyses in the presence of priors, and Royall (2000b) interpreted the likelihood ratio as the Bayes factor for the case of comparing two simple hypotheses. To overcome the objection against the Bayes factor as a measure of evidence for composite hypotheses, Bickel (2013a) presented general classes of prior-free approximations to Bayes factors.

In science, the Bayes factor is currently the most prominent measure of the relevancy of the evidence, how relevant the data are when considered as evidence for or against a composite hypothesis. The degree of that relevance is known as the *relevancy* of the evidence to whether some hypothesis is true (Koehler, 2002). The many other proposed measures of the relevancy of the evidence include the *relative belief ratio*, which is the posterior probability of a hypothesis divided by its prior probability (Evans, 2015), and the *relevance measure* of Carnap (1962, §67); see Koscholke (2017).

The data or other evidence can be relevant to the truth of a hypothesis without warranting the conclusion that the hypothesis is true or the conclusion that it is false. That is why the relevancy of the evidence, also called the probative value of the evidence, is distinguished from the *sufficiency* of the evidence to justify drawing a conclusion about the hypothesis (Kaye and Koehler, 2003). While the Bayes factor succeeds in quantifying the relevancy of the data to the truth of the hypothesis, it fails to quantify the sufficiency of the data to warrant a conclusion about the hypothesis (Lavine and Schervish, 1999). Conversely, the posterior probability and the posterior odds of a hypothesis quantify the sufficiency of the data to justify a conclusion but not the relevancy of the data. Fiducial probability defined as an observed confidence level is an alternative measure of the sufficiency of the evidence (Bickel, 2011), as is its generalization in Bickel (2018b).

Nonetheless, the Bayes factor qualifies as a measure of the sufficiency of the data as well as its relevancy when the prior probability of the hypothesis is fixed at 50%, for the Bayes factor is then equal to the posterior odds. The commonly used thresholds for Bayes factors to achieve certain scales of evidence were originally intended for that case (Jeffreys, 1948).

Another measure of both the sufficiency and relevancy of data is the weight of evidence under the general law of likelihood, defined in Section 2. That section also defines a generalization of  $L(\bullet)$  as a particular conditional possibility measure that is dual to a necessity measure, as those terms are used in possibility theory (§2.2). Section 3 derives the general law from idealizations of sufficiency and relevancy as opposed to the idealization of inference to the best explanation found in Bickel (2012).

Appendix A contrasts this paper’s approach to possibility theory with the interpretation of possibility as an upper probability.

## 2 Weight of evidence

### 2.1 Preliminary notation and definitions

Let  $x$  denote an observed scalar, vector, or matrix in some set  $\mathcal{X}$  of possible observations. This  $x$ , a realization of a random element  $X$ , may be a statistic that depends on other observations.

Consider a set  $\Theta$  and a family of density functions  $\{f_{\theta_0} : \theta_0 \in \Theta\}$  such that the parameter is identifiable in the sense that  $f_{\theta_0} \neq f_{\theta}$  for all  $\theta_0, \theta \in \Theta : \theta_0 \neq \theta$ . If the interest parameter value were equal to  $\theta$ , then  $f_{\theta}(x)$  would be the probability density or probability mass of the observation that  $X = x$ . The *likelihood function* is the function  $f_{\bullet}(x)$ , that is,  $f_{\theta}(x)$  as a function of  $\theta$  for all  $\theta \in \Theta$ , and its maximum likelihood estimate (MLE) is  $\hat{\theta} = \arg \sup_{\theta} f_{\theta}(x)$ .

The function  $f_{\bullet}(x)$  may be any pseudo-likelihood function such that  $f_{\theta}(x)$  approximates a probability density for every  $\theta \in \Theta$ . Thus,  $f_{\bullet}(x)$  may be a marginal, conditional, estimated, or integrated likelihood, eliminating a nuisance parameter. If the profile likelihood does not approximate a density for a particular model, it may nevertheless be corrected to approximate a conditional or marginal likelihood in certain cases (Severini, 2000, pp. 310-312, 323). The prefix “pseudo” is somewhat misleading: even the “true” likelihood function might be considered a pseudo-likelihood function since a statistical model cannot completely capture the data-generation process (Lindsey, 1996, §6.5).

Each hypothesis about  $\theta$  may be expressed as “ $\theta \in \mathcal{H}$ ” for an  $\mathcal{H} \subset \Theta$ . Thus, all possible hypotheses about  $\theta$  correspond to members of  $\mathfrak{H}$ , a set of subsets of  $\Theta$ . For example, if  $\Theta$  is the real line,  $\mathfrak{H}$  is the set of Borel subsets of  $\Theta$ , and  $\overline{\{0\}}$  is the complement  $\Theta \setminus \{0\}$  of  $\{0\}$ , then the hypothesis that  $\theta \neq 0$  is the hypothesis that  $\theta \in \overline{\{0\}}$ , corresponding to the subset  $\overline{\{0\}}$ , which is a member of  $\mathfrak{H}$ .

### 2.2 Likeliness and unlikeliness

For any  $\mathcal{H} \in \mathfrak{H}$  and  $\mathcal{R} \in \mathfrak{H} \setminus \{\emptyset\}$ , call

$$L(\mathcal{H}) = \frac{\sup_{\theta \in \mathcal{H}} f_{\theta}(x)}{\sup_{\theta \in \Theta} f_{\theta}(x)} \tag{1}$$

the *marginal likeliness* of the hypothesis that  $\theta \in \mathcal{H}$  and, if  $L(\mathcal{R}) > 0$ ,

$$L(\mathcal{H}|\mathcal{R}) = \frac{L(\mathcal{H} \cap \mathcal{R})}{L(\mathcal{R})} \tag{2}$$

the *conditional likeliness* of the hypothesis that  $\theta \in \mathcal{H}$  given  $\theta \in \mathcal{R}$  (Bickel, 2014). Here, the supremum is the least upper bound in  $[0, \infty[$ ,  $\sup \emptyset \equiv 0$ . It follows that  $L(\mathcal{H}) \in [0, 1]$  and  $L(\mathcal{H}|\mathcal{R}) \in [0, 1]$ .

The likeliness of a hypothesis is insufficient as a measure of its strength of evidence since the likeliness of the hypothesis's alternative must also be considered. For that reason, it is convenient to define the *marginal unlikeliness* of the hypothesis that  $\theta \in \mathcal{H}$  as  $U(\mathcal{H}) = L(\overline{\mathcal{H}})$  and the *conditional unlikeliness* of the hypothesis that  $\theta \in \mathcal{H}$  given  $\theta \in \mathcal{R}$  as  $U(\mathcal{H}|\mathcal{R}) = L(\overline{\mathcal{H}}|\mathcal{R})$ , where  $\overline{\mathcal{H}}$  is the complement of  $\mathcal{H}$ . The likeliness and unlikeliness of a hypothesis are combined into a single measure of evidence in Section 2.4. According to possibility theory,  $L(\bullet)$  is a possibility measure, and  $1 - U(\bullet)$  is a necessity measure (Bickel, 2014).

### 2.3 Marginal and conditional weight of evidence

Suppose  $\mathcal{H}_1, \mathcal{H}_2 \in \mathfrak{H}$ . According to the general law of likelihood (§1), the *weight of evidence* in the observation that  $X = x$  substantiating the hypothesis that  $\theta \in \mathcal{H}_1$  as opposed to the hypothesis that  $\theta \in \mathcal{H}_2$  is the extended real number

$$W(\mathcal{H}_1; \mathcal{H}_2) = \frac{\sup_{\theta \in \mathcal{H}_1} f_\theta(x)}{\sup_{\theta \in \mathcal{H}_2} f_\theta(x)} \quad (3)$$

with the conventions that  $0/0 = 1$ ,  $1/\infty = 0$ , and  $1/0 = \infty$ . That will be called the *marginal weight of evidence* to distinguish it from the conditional weight of evidence, defined below. If  $f_\bullet(x)$  is a profile likelihood function and  $W(\mathcal{H}_1; \mathcal{H}_2) \neq 1$ , then  $W(\mathcal{H}_1; \mathcal{H}_2)$  reduces to the quantity considered by Zhang and Zhang (2013a), as discussed in Bickel (2013b).

The *conditional weight of evidence* in the observation that  $X = x$  substantiating the hypothesis that  $\theta \in \mathcal{H}_1$  as opposed to the hypothesis that  $\theta \in \mathcal{H}_2$  given  $\theta \in \mathcal{R}$  is

$$W(\mathcal{H}_1; \mathcal{H}_2|\mathcal{R}) = W(\mathcal{H}_1 \cap \mathcal{R}; \mathcal{H}_2 \cap \mathcal{R}) \quad (4)$$

for all  $\mathcal{H}_1, \mathcal{H}_2, \mathcal{R} \in \mathfrak{H}$  such that  $L(\mathcal{R}) > 0$ . This is connected to the likeliness of Section 2.2 as follows.

**Theorem 1.** For any  $\mathcal{H}_1, \mathcal{H}_2, \mathcal{R} \in \mathfrak{H}$  such that  $L(\mathcal{R}) > 0$ ,

$$W(\mathcal{H}_1; \mathcal{H}_2|\mathcal{R}) = \frac{L(\mathcal{H}_1|\mathcal{R})}{L(\mathcal{H}_2|\mathcal{R})}. \quad (5)$$

For any set  $\mathfrak{H}_0 \subset \mathfrak{H}$  such that  $\bigcup_{\mathcal{H}_0 \in \mathfrak{H}_0} \mathcal{H}_0 = \mathcal{H}$ ,

$$L(\mathcal{H}|\mathcal{R}) = \frac{\sup_{\theta \in \mathcal{H} \cap \mathcal{R}} f_\theta(x)}{\sup_{\theta \in \mathcal{R}} f_\theta(x)} = \sup_{\mathcal{H}_0 \in \mathfrak{H}_0} L(\mathcal{H}_0|\mathcal{R}). \quad (6)$$

As Bickel (2014) claimed, for any partition  $\mathfrak{P} \subset \mathfrak{H}$  of  $\Theta$  such that  $L(\mathcal{R}) > 0$  for all  $\mathcal{R} \in \mathfrak{P}$ ,

$$L(\mathcal{H}) = \sup_{\mathcal{R} \in \mathfrak{P}} L(\mathcal{R}) L(\mathcal{H}|\mathcal{R}). \quad (7)$$

*Proof.* By equations (1), (4), and (3),

$$W(\mathcal{H}_1; \mathcal{H}_2 | \mathcal{R}) = \frac{\sup_{\theta \in \mathcal{H}_1 \cap \mathcal{R}} f_\theta(x)}{\sup_{\theta \in \mathcal{H}_2 \cap \mathcal{R}} f_\theta(x)} = \frac{\sup_{\theta \in \mathcal{H}_1 \cap \mathcal{R}} f_\theta(x) / \sup_{\theta \in \mathcal{R}} f_\theta(x)}{\sup_{\theta \in \mathcal{H}_2 \cap \mathcal{R}} f_\theta(x) / \sup_{\theta \in \mathcal{R}} f_\theta(x)},$$

which is the right-hand side of equation (5) according to equation (2). Equation (1) and (2) imply that

$$L(\mathcal{H} | \mathcal{R}) = L\left(\bigcup_{\mathcal{H}_0 \in \mathfrak{H}_0} \mathcal{H}_0 | \mathcal{R}\right) = \frac{\sup_{\mathcal{H}_0 \in \mathfrak{H}_0} \sup_{\theta \in \mathcal{H}_0 \cap \mathcal{R}} f_\theta(x)}{\sup_{\theta \in \mathcal{R}} f_\theta(x)} = \sup_{\mathcal{H}_0 \in \mathfrak{H}_0} \frac{\sup_{\theta \in \mathcal{H}_0 \cap \mathcal{R}} f_\theta(x)}{\sup_{\theta \in \mathcal{R}} f_\theta(x)},$$

yielding  $L(\mathcal{H} | \mathcal{R}) = \sup_{\mathcal{H}_0 \in \mathfrak{H}_0} L(\mathcal{H}_0 | \mathcal{R})$ . The other portion of formula (6) is established by substituting  $\{\{\theta\} : \theta \in \mathcal{H}\}$  for  $\mathfrak{H}_0$ . Since  $\mathfrak{P} \subset \mathfrak{H}$  is a partition,

$$L(\mathcal{H}) = L(\mathcal{H} \cap \Theta) = L\left(\mathcal{H} \cap \bigcup_{\mathcal{R} \in \mathfrak{P}} \mathcal{R}\right) = L\left(\bigcup_{\mathcal{R} \in \mathfrak{P}} (\mathcal{H} \cap \mathcal{R})\right) = L\left(\bigcup_{\mathcal{R}_0 \in \mathfrak{P}(\mathcal{H})} \mathcal{R}_0\right),$$

where  $\mathfrak{P}(\mathcal{H}) = \{\mathcal{R} \in \mathfrak{P} : \mathcal{R} \subseteq \mathcal{H}\}$ . Thus, using equation (6),

$$L(\mathcal{H}) = \sup_{\mathcal{R}_0 \in \mathfrak{P}(\mathcal{H})} L(\mathcal{R}_0) = \sup_{\mathcal{R} \in \mathfrak{P}} L(\mathcal{H} \cap \mathcal{R}) = \sup_{\mathcal{R} \in \mathfrak{P}} L(\mathcal{R}) L(\mathcal{H} | \mathcal{R}),$$

with the last equality following from equation (2).  $\square$

Equation (6) is the foundation of the multiple hypothesis method of Bickel (2014).

## 2.4 Absolute weight of evidence

The strength of evidence favoring the hypothesis that  $\theta \in \mathcal{H}$  can also be quantified without explicit reference to a second hypothesis by taking that second hypothesis to be the alternative to the first (i.e., the hypothesis that  $\theta \notin \mathcal{H}$ ). The *conditional weight of evidence* in the observation that  $X = x$  substantiating the hypothesis that  $\theta \in \mathcal{H}$  given  $\theta \in \mathcal{R}$  is  $W(\mathcal{H} | \mathcal{R}) = W(\mathcal{H}; \overline{\mathcal{H}} | \mathcal{R})$ . Likewise, the *marginal weight of evidence* in the observation that  $X = x$  substantiating the hypothesis that  $\theta \in \mathcal{H}$  is  $W(\mathcal{H}) = W(\mathcal{H} | \Theta)$ .

**Corollary 1.** *Under the assumptions of Theorem 1, for any  $\mathcal{H}, \mathcal{R} \in \mathfrak{H}$ ,*

$$W(\mathcal{H} | \mathcal{R}) = \frac{L(\mathcal{H} | \mathcal{R})}{U(\mathcal{H} | \mathcal{R})} = \frac{L(\mathcal{H} \cap \mathcal{R})}{L(\mathcal{R} \setminus \mathcal{H})} = \frac{\sup_{\theta \in \mathcal{H} \cap \mathcal{R}} f_\theta(x)}{\sup_{\theta \in \mathcal{R} \setminus \mathcal{H}} f_\theta(x)} \quad (8)$$

$$W(\mathcal{H} | \mathcal{R}) = \frac{\sup_{\mathcal{R} \in \mathfrak{P}} L(\mathcal{R}) L(\mathcal{H} | \mathcal{R})}{\sup_{\mathcal{R} \in \mathfrak{P}} L(\mathcal{R}) L(\overline{\mathcal{H}} | \mathcal{R})}. \quad (9)$$

*Proof.* The claims follow directly from  $U(\mathcal{H}|\mathcal{R}) = L(\overline{\mathcal{H}}|\mathcal{R})$  and from equations (1), (5), and (7).  $\square$

Equation (8) (Bickel, 2014) indicates that  $W(\mathcal{H}|\mathcal{R})$  is a coherent measure of evidence in the sense to be defined in Section 3. As will be seen, that property supports calling  $W(\mathcal{H}|\mathcal{R})$  the weight of evidence.

## 2.5 Likelihood and unlikelihood from the weight of evidence

While the weight of evidence is the ratio of likelihood to the unlikelihood (8), it is convenient in some applications to derive the likelihood and unlikelihood from the weight of evidence.

**Lemma 1.** *Given  $\mathcal{H}, \mathcal{R} \in \mathfrak{H}$  such that  $L(\mathcal{R}) > 0$ , it follows that  $L(\mathcal{H}|\mathcal{R}) = 1$  and  $U(\mathcal{H}|\mathcal{R}) = 1/W(\mathcal{H}|\mathcal{R})$  if  $W(\mathcal{H}|\mathcal{R}) \geq 1$  but that  $L(\mathcal{H}|\mathcal{R}) = W(\mathcal{H}|\mathcal{R})$  and  $U(\mathcal{H}|\mathcal{R}) = 1$  if  $W(\mathcal{H}|\mathcal{R}) < 1$ .*

Bickel (2014) proves the result and relates it to the theory of ranking functions treated in Spohn (2012, §5.2).

## 3 Derivation from coherence and Bayes compatibility

Let  $P$  stand for a probability measure on  $(\Theta \times \mathcal{X}, \mathfrak{H} \otimes \mathfrak{X})$ , where  $\mathfrak{X}$  is a  $\sigma$ -algebra of subsets of  $\mathcal{X}$ . Consider a random parameter  $\vartheta$  of prior distribution  $P_0 = P(\bullet \times \mathcal{X})$  on  $(\Theta, \mathfrak{H})$  such that the posterior probability that  $\vartheta \in \mathcal{H}$  is

$$P(\vartheta \in \mathcal{H}|x) = \frac{P_0(\vartheta \in \mathcal{H}) \int_{\mathcal{H}} f_{\theta}(x) dP_0(\theta|\mathcal{H})}{\int f_{\theta}(x) dP_0(\theta)}.$$

This is considered a function of  $P$  such that, if  $Q$  were the joint distribution on  $(\Theta \times \mathcal{X}, \mathfrak{H} \otimes \mathfrak{X})$ , the posterior distribution would be  $Q(\vartheta \in \mathcal{H}|x)$  with  $Q$  in place of  $P$  and  $Q_0 = Q(\bullet \times \mathcal{X})$  in place of  $P_0$ . The *increase in the odds ratio* due to the observation that  $X = x$  in favor of the hypothesis that  $\theta \in \mathcal{H}$  given  $\theta \in \mathcal{R}$  is the ratio of the conditional posterior odds to the conditional prior odds:

$$\Delta(\mathcal{H}; P|\mathcal{R}) = \frac{P(\vartheta \in \mathcal{H}|x, \mathcal{R})/P(\vartheta \notin \mathcal{H}|x, \mathcal{R})}{P_0(\vartheta \in \mathcal{H}|\mathcal{R})/P_0(\vartheta \notin \mathcal{H}|\mathcal{R})} = \frac{\int_{\mathcal{H} \cap \mathcal{R}} f_{\theta}(x) dP_0(\theta|\mathcal{H}, \mathcal{R})}{\int_{\mathcal{R} \setminus \mathcal{H}} f_{\theta}(x) dP_0(\theta|\overline{\mathcal{H}}, \mathcal{R})}. \quad (10)$$

The conditional Bayes factor  $B(\mathcal{H}|\mathcal{R})$  as a function of  $\mathcal{H}$  and  $\mathcal{R}$ , is defined such that  $B(\mathcal{H}|\mathcal{R}) = \Delta(\mathcal{H}; P|\mathcal{R})$  for some fixed probability measure  $P$  on  $(\Theta \times \mathcal{X}, \mathfrak{H} \otimes \mathfrak{X})$ .

The requirement of Edwards (1992) that a measure of support for one hypothesis over another be compatible with Bayes's theorem is generalized to composite hypotheses by the

following definition, differing from the generalization that often forbids accepting a hypothesis of sufficiently high weight of evidence (Bickel, 2013a). Any function  $u : \mathfrak{H}^2 \rightarrow [0, \infty]$  measures the *odds ratio increase* due to the observation that  $X = x$  if, for every  $\mathcal{H} \in \mathfrak{H}$ , there is a probability measure  $P_{\mathcal{H}}$  on  $(\Theta, \mathfrak{H})$  such that

$$u(\mathcal{H}|\mathcal{R}) = \Delta(\mathcal{H}; P_{\mathcal{H}}|\mathcal{R}) \quad (11)$$

for all  $\mathcal{R} \in \mathfrak{H}$  satisfying  $L(\mathcal{R}) > 0$ . Functions that measure the odds ratio increase quantify the *relevancy* of the body of evidence to whether or not a hypothesis is true.

On the other hand, a property of a measure of the *sufficiency* of the body of evidence for concluding that a hypothesis is true is the avoidance of asserting that contradictory statements are individually supported by the evidence (Schervish, 1996; Lavine and Schervish, 1999; Zhang and Zhang, 2013a). More generally, the functions  $v : \mathfrak{H} \rightarrow [0, \infty]$  and  $v : \mathfrak{H}^2 \rightarrow [0, \infty]$  are *logically coherent* if

$$v(\mathcal{H}_0) \leq v(\mathcal{H}_1) \iff (\theta \in \mathcal{H}_0 \implies \theta \in \mathcal{H}_1) \quad (12)$$

$$v(\mathcal{H}_0|\mathcal{R}) \leq v(\mathcal{H}_1|\mathcal{R}) \iff (\theta \in \mathcal{H}_0 \cap \mathcal{R} \implies \theta \in \mathcal{H}_1 \cap \mathcal{R}) \quad (13)$$

for all  $\mathcal{H}_0, \mathcal{H}_1 \in \mathfrak{H}$  and all  $\mathcal{R} \in \mathfrak{H}$ . The main Bayesian logically coherent measure is the posterior probability. A frequentist logically coherent measure is the compatibility or  $c$  value, a generalization of the  $p$  value (Bickel and Patriota, 2019).

In short, whereas the odds ratio increase quantifies the relevancy of the evidence, logical coherence is a minimal requirement of a measure of the sufficiency of evidence. Putting them together leads to the following definition and theorem.

**Definition 1.** A function  $w : \mathfrak{H}^2 \rightarrow [0, \infty]$  measures the *relevancy and sufficiency of the evidence* if it both measures the odds ratio increase and is logically coherent.

**Theorem 2.** If  $\hat{\theta}_{\mathcal{H} \cap \mathcal{R}} = \arg \sup_{\theta \in \mathcal{H} \cap \mathcal{R}} f_{\theta}(x)$  and  $\hat{\theta}_{\mathcal{R} \setminus \mathcal{H}} = \arg \sup_{\theta \in \mathcal{R} \setminus \mathcal{H}} f_{\theta}(x)$  are unique, then a function  $w : \mathfrak{H}^2 \rightarrow [0, \infty]$  measures the *relevancy and sufficiency of the evidence* if and only if it is the *weight of evidence function*  $W(\bullet|\bullet)$ .

*Proof.* ( $\Leftarrow$ ). The following statements apply for all  $\mathcal{H}, \mathcal{H}_0, \mathcal{H}_1 \in \mathfrak{H}$  and all  $\mathcal{R} \in \mathfrak{H}$ . Let  $\delta(\bullet; \hat{\theta}_{\mathcal{H} \cap \mathcal{R}})$  and  $\delta(\bullet; \hat{\theta}_{\mathcal{R} \setminus \mathcal{H}})$  denote the Dirac probability measures on  $(\Theta, \mathfrak{H})$  with mass at  $\hat{\theta}_{\mathcal{H} \cap \mathcal{R}}$  and  $\hat{\theta}_{\mathcal{R} \setminus \mathcal{H}}$ , respectively. By equation (8),

$$W(\mathcal{H}|\mathcal{R}) = \frac{\int f_{\theta}(x) d\delta(\theta; \hat{\theta}_{\mathcal{H} \cap \mathcal{R}})}{\int f_{\theta}(x) d\delta(\theta; \hat{\theta}_{\mathcal{R} \setminus \mathcal{H}})}. \quad (14)$$

There is a probability measure  $P$  on  $(\Theta \times \mathcal{X}, \mathfrak{H} \otimes \mathfrak{X})$  such that  $P_0(\bullet|\mathcal{H}, \mathcal{R}) = \delta(\bullet; \widehat{\theta}_{\mathcal{H} \cap \mathcal{R}})$  and  $P_0(\bullet|\overline{\mathcal{H}}, \mathcal{R}) = \delta(\bullet; \widehat{\theta}_{\mathcal{R} \setminus \mathcal{H}})$ , in which case equations (14) and (10) imply that  $W(\mathcal{H}|\mathcal{R}) = \Delta(\mathcal{H}; P|\mathcal{R})$ . Thus,  $W(\bullet|\bullet)$  measures the odds ratio increase. The fact that  $W(\mathcal{H}_0, \mathcal{R}) \leq W(\mathcal{H}_1, \mathcal{R})$  if and only if  $\mathcal{H}_0 \subseteq \mathcal{H}_1$  demonstrates equation (13). Therefore,  $W(\bullet|\bullet)$  is logically coherent. Thus, both criteria of Definition 1 are satisfied. ( $\implies$ ). Let  $w : \mathfrak{H}^2 \rightarrow [0, \infty]$  denote a function that measures the relevancy and sufficiency of the evidence. By Definition 1,  $w$  both measures the odds ratio increase and is logically coherent. Assume that there are  $\mathcal{H} \in \mathfrak{H}$  and  $\mathcal{R} \in \mathfrak{H}$  and, contrary to the  $w = W$  claim and equation (8), such that

$$w(\mathcal{H}|\mathcal{R}) \neq \frac{\sup_{\theta \in \mathcal{H} \cap \mathcal{R}} f_{\theta}(x)}{\sup_{\theta \in \mathcal{R} \setminus \mathcal{H}} f_{\theta}(x)} \quad (15)$$

in order to prove the claim by contradiction. Since  $w = v$ , equations (10), (11), and (13) yield

$$\int_{\mathcal{H}_0 \cap \mathcal{R}} f_{\theta}(x) dP_0(\theta|\mathcal{H}_0, \mathcal{R}) \leq \int_{\mathcal{H}_1 \cap \mathcal{R}} f_{\theta}(x) dP_0(\theta|\mathcal{H}_1, \mathcal{R}) \iff \mathcal{H}_0 \subseteq \mathcal{H}_1.$$

Since  $\{\widehat{\theta}_{\mathcal{H} \cap \mathcal{R}}\} \subseteq \mathcal{H} \cap \mathcal{R}$ ,

$$\begin{aligned} \int_{\mathcal{H} \cap \mathcal{R}} f_{\theta}(x) dP_0(\theta|\mathcal{H}, \mathcal{R}) &\geq \int_{\{\widehat{\theta}_{\mathcal{H} \cap \mathcal{R}}\}} f_{\theta}(x) dP_0(\theta|\{\widehat{\theta}_{\mathcal{H} \cap \mathcal{R}}\}, \mathcal{R}) \\ &= \int f_{\theta}(x) d\delta(\theta; \widehat{\theta}_{\mathcal{H} \cap \mathcal{R}}) = \sup_{\theta \in \mathcal{H} \cap \mathcal{R}} f_{\theta}(x), \end{aligned}$$

but that requires that  $\int_{\mathcal{H} \cap \mathcal{R}} f_{\theta}(x) dP_0(\theta|\mathcal{H}, \mathcal{R}) = \sup_{\theta \in \mathcal{H} \cap \mathcal{R}} f_{\theta}(x)$  (cf. Coletti et al., 2009). Analogous reasoning leads to  $\int_{\mathcal{R} \setminus \mathcal{H}} f_{\theta}(x) dP_0(\theta|\mathcal{H}, \mathcal{R}) = \sup_{\theta \in \mathcal{R} \setminus \mathcal{H}} f_{\theta}(x)$ . Thus, equations (10), (11), and (13) establish equation (8), contradicting equation (15), thereby proving the  $w = W$  claim.  $\square$

Theorem 2 says the weight of evidence uniquely measures both the relevancy and the sufficiency of the evidence. That raises questions about the senses in which the posterior probability and the Bayes factor fall short as measures of the relevancy and sufficiency of the evidence. Lavine and Schervish (1999) demonstrated that the posterior probability but not the Bayes factor is coherent as a measure of evidence. This is restated in the following corollaries in addition to whether each measures the odds ratio increase.

**Corollary 2.** *Given any probability measure  $P$  on  $(\Theta \times \mathcal{X}, \mathfrak{H} \otimes \mathfrak{X})$  satisfying the above conditions, the conditional Bayes factor function  $B$  measures the odds ratio increase but is not necessarily logically coherent.*

*Proof.* By the definition of the conditional Bayes factor  $B(\mathcal{H}|\mathcal{R}) = \Delta(\mathcal{H}; P|\mathcal{R})$ . Thus,  $u = B$  yields equation (11), establishing the first claim. The second claim is established by noting that, according to Theorem 2,  $B$  is only logically coherent in the special case that  $B(\mathcal{H}|\mathcal{R}) = W(\mathcal{H}|\mathcal{R})$  for all  $\mathcal{H} \in \mathfrak{H}$  and all  $\mathcal{R} \in \mathfrak{R}$ .  $\square$

The next corollary uses the *posterior odds function*, the function  $\text{odds}(\bullet|x, \bullet)$  defined on  $\mathfrak{H}^2$  such that  $\text{odds}(\mathcal{H}|x, \mathcal{R}) = P(\vartheta \in \mathcal{H}|x, \mathcal{R}) / P(\vartheta \notin \mathcal{H}|x, \mathcal{R})$  for all  $\mathcal{H}, \mathcal{R} \in \mathfrak{H}$ .

**Corollary 3.** *Given any probability measure  $P$  on  $(\Theta \times \mathcal{X}, \mathfrak{H} \otimes \mathfrak{X})$  satisfying the above conditions, the posterior odds function is logically coherent but does not necessarily measure the odds ratio increase.*

*Proof.* Consider an  $\mathcal{R} \in \mathfrak{R}$  that satisfies  $L(\mathcal{R}) > 0$  and  $\mathcal{H}_0, \mathcal{H}_1 \in \mathfrak{H}$  such that  $\mathcal{H}_0 \subseteq \mathcal{H}_1$ . By the additivity of probability measures,  $P(\vartheta \in \mathcal{H}_0|x, \mathcal{R}) \leq P(\vartheta \in \mathcal{H}_1|x, \mathcal{R})$ , from which  $\text{odds}(\mathcal{H}_0|x, \mathcal{R}) \leq \text{odds}(\mathcal{H}_1|x, \mathcal{R})$  follows. Likewise, any  $\mathcal{H}_0, \mathcal{H}_1 \in \mathfrak{H}$  such that  $\text{odds}(\mathcal{H}_0|x, \mathcal{R}) \leq \text{odds}(\mathcal{H}_1|x, \mathcal{R})$  are related by  $\mathcal{H}_0 \subseteq \mathcal{H}_1$ . Thus,  $v = \text{odds}(\bullet|x, \bullet)$  yields equation (13), establishing the first claim. The second claim is established by noting that, according to Theorem 2,  $\text{odds}(\bullet|x, \bullet)$  only measures the odds ratio increase in the special case that  $\text{odds}(\mathcal{H}|x, \mathcal{R}) = W(\mathcal{H}|\mathcal{R})$  for all  $\mathcal{H} \in \mathfrak{H}$  and all  $\mathcal{R} \in \mathfrak{R}$  satisfying  $L(\mathcal{R}) > 0$ .  $\square$

Lavine and Schervish (1999) likewise argued that the posterior probability is coherent as a measure of evidence. In short, while the Bayes factor measures the odds ratio increase and the posterior probability is logically coherent, the weight of evidence is the only quantity with both properties.

## Acknowledgments

This research was partially supported by the Natural Sciences and Engineering Research Council of Canada (RGPIN/356018-2009), by the Canada Foundation for Innovation (CFI16604), by the Ministry of Research and Innovation of Ontario (MRI16604), and by the Faculty of Medicine of the University of Ottawa.

## References

- Barnard, G. A., 1967. The use of the likelihood function in statistical practice. Proc. 5th Berkeley Symp. on Math. Stat. Prob. Vol. I, 27–40.
- Bickel, D. R., 2011. Estimating the null distribution to adjust observed confidence levels for genome-scale screening. *Biometrics* 67, 363–370.

- Bickel, D. R., 2012. The strength of statistical evidence for composite hypotheses: Inference to the best explanation. *Statistica Sinica* 22, 1147–1198.
- Bickel, D. R., 2013a. Minimax-optimal strength of statistical evidence for a composite alternative hypothesis. *International Statistical Review* 81, 188–206.
- Bickel, D. R., 2013b. Pseudo-likelihood, explanatory power, and Bayes's theorem [comment on "A likelihood paradigm for clinical trials"]. *Journal of Statistical Theory and Practice* 7, 178–182.
- Bickel, D. R., 2014. Model fusion and multiple testing in the likelihood paradigm: Shrinkage and evidence supporting a point null hypothesis. Working Paper, University of Ottawa, deposited in uO Research at <http://hdl.handle.net/10393/31897>.
- Bickel, D. R., 2018a. Bayesian revision of a prior given prior-data conflict, expert opinion, or a similar insight: A large-deviation approach. *Statistics* 52, 552–570.
- Bickel, D. R., 2018b. Confidence distributions and empirical Bayes posterior distributions unified as distributions of evidential support, working paper, DOI: 10.5281/zenodo.2529438.  
URL <https://doi.org/10.5281/zenodo.2529438>
- Bickel, D. R., Patriota, A. G., 2019. Self-consistent confidence sets and tests of composite hypotheses applicable to restricted parameters. *Bernoulli* 25 (1), 47–74.
- Blume, J., 2013. Likelihood and composite hypotheses [comment on "A likelihood paradigm for clinical trials"]. *Journal of Statistical Theory and Practice* 7 (2), 183–186.
- Blume, J. D., 2002. Likelihood methods for measuring statistical evidence. *Statistics In Medicine* 21, 2563–2599.
- Blume, J. D., 2011. Likelihood and its evidential framework. In: Bandyopadhyay, P. S., Forster, M. R. (Eds.), *Philosophy of Statistics*. North Holland, Amsterdam, pp. 493–512.
- Carnap, R., 1962. *Logical Foundations of Probability*. University of Chicago Press, Chicago.
- Coletti, G., Scozzafava, R., Vantaggi, B., 2009. Integrated likelihood in a finitely additive setting. In: *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*. Vol. 5590 of *Lecture Notes in Comput. Sci.* Springer, Berlin, pp. 554–565.
- de Finetti, B., 1970. *Theory of Probability: A Critical Introductory Treatment*, 1st Edition. John Wiley & Sons, Ltd., New York.

- Dubois, D., Moral, S., Prade, H., Jan. 1997. A semantics for possibility theory based on likelihoods. *Journal of Mathematical Analysis and Applications* 205 (2), 359–380.
- Edwards, A. W. F., 1992. *Likelihood*. Johns Hopkins Press, Baltimore.
- Evans, M., 2015. *Measuring Statistical Evidence Using Relative Belief*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press, New York.
- Fisher, R. A., 1973. *Statistical Methods and Scientific Inference*. Hafner Press, New York.
- Fossaluzza, V., Izbicki, R., da Silva, G., Esteves, L., 2017. Coherent hypothesis testing. *American Statistician* 71 (3), 242–248.
- Giang, P. H., Shenoy, P. P., 2005. Decision making on the sole basis of statistical likelihood. *Artificial Intelligence* 165, 137–163.
- Hacking, I., 1965. *Logic of Statistical Inference*. Cambridge University Press, Cambridge.
- Hoch, J. S., Blume, J. D., 2008. Measuring and illustrating statistical evidence in a cost-effectiveness analysis. *Journal of Health Economics* 27, 476–495.
- Hodge, S. E., Baskurt, Z., Strug, L. J., 2011. Using parametric multipoint lods and mods for linkage analysis requires a shift in statistical thinking. *Human Heredity* 72 (4), 264–275.
- Jeffreys, H., 1948. *Theory of Probability*. Oxford University Press, London.
- Kalbfleisch, J. D., 2000. Comment on R. Royall, "On the probability of observing misleading statistical evidence". *Journal of the American Statistical Association* 95, 770–771.
- Kaplan, M., 1996. *Decision Theory As Philosophy*. Cambridge University Press, Cambridge.
- Kaye, D., Koehler, J., 2003. The misquantification of probative value. *Law and Human Behavior* 27 (6), 645–659.
- Koehler, J. J., 2002. When do courts think base rate statistics are relevant? *Jurimetrics*, 373–402.
- Koscholke, J., 2017. Carnap's relevance measure as a probabilistic measure of coherence. *Erkenntnis* 82 (2), 339–350.
- Lavine, M., Schervish, M. J., 1999. Bayes factors: What they are and what they are not. *American Statistician* 53, 119–122.

- Lindsey, J., 1996. *Parametric Statistical Inference*. Oxford Science Publications. Clarendon Press, Oxford.
- Morgenthaler, S., Staudte, R. G., 2012. Advantages of Variance Stabilization. *Scandinavian Journal of Statistics* 39 (4), 714–728.
- Patriota, A. G., 2013. A classical measure of evidence for general null hypotheses. *Fuzzy Sets and Systems* 233, 74 – 88.
- Rohde, C. A., 2014. *Pure likelihood methods*. Springer International Publishing, New York, Ch. 18, pp. 197–209.
- Royall, R., 1997. *Statistical Evidence: A Likelihood Paradigm*. CRC Press, New York.
- Royall, R., 2000a. On the probability of observing misleading statistical evidence. *Journal of the American Statistical Association* 95, 760–768.
- Royall, R., 2000b. On the probability of observing misleading statistical evidence (with discussion). *Journal of the American Statistical Association* 95, 760–780.
- Schervish, M. J., 1996. P values: What they are and what they are not. *American Statistician* 50, 203–206.
- Severini, T., 2000. *Likelihood Methods in Statistics*. Oxford University Press, Oxford.
- Spanos, A., 2013. Revisiting the likelihoodist evidential account [comment on "A likelihood paradigm for clinical trials"]. *Journal of Statistical Theory and Practice* 7 (2), 187–195.
- Spohn, W., 2012. *The Laws of Belief: Ranking Theory and Its Philosophical Applications*. Oxford University Press.
- Sprott, D. A., 2000. *Statistical Inference in Science*. Springer, New York.
- Strug, L., 2018. The evidential statistical paradigm in genetics. *Genetic Epidemiology*. URL <https://doi.org/10.1002/gepi.22151>
- Strug, L., Hodge, S., Chiang, T., Pal, D., Corey, P., Rohde, C., 2010. A pure likelihood approach to the analysis of genetic association data: An alternative to Bayesian and frequentist analysis. *European Journal of Human Genetics* 18, 933–941.
- Strug, L. J., Hodge, S. E., 2006a. An alternative foundation for the planning and evaluation of linkage analysis i. decoupling 'error probabilities' from 'measures of evidence'. *Human Heredity* 61, 166–188.

- Strug, L. J., Hodge, S. E., 2006b. An alternative foundation for the planning and evaluation of linkage analysis. ii. implications for multiple test adjustments. *Human Heredity* 61, 200–209.
- Strug, L. J., Rohde, C. A., Corey, P. N., 2007. An introduction to evidential sample size calculations. *American Statistician* 61, 207–212.
- Vieland, V. J., Seok, S.-C., 2016. Statistical evidence measured on a properly calibrated scale for multinomial hypothesis comparisons. *Entropy* 18 (4), 114.
- Walley, P., 1991. *Statistical Reasoning with Imprecise Probabilities*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, Boca Raton, Florida.
- Walley, P., Moral, S., 1999. Upper probabilities based only on the likelihood function. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 61, 831–847.
- Zhang, Z., Zhang, B., 2013a. A likelihood paradigm for clinical trials. *Journal of Statistical Theory and Practice* 7, 157–177.
- Zhang, Z., Zhang, B., 2013b. Rejoinder [on "A likelihood paradigm for clinical trials"]. *Journal of Statistical Theory and Practice* 7, 196–203.

## A Imprecise probability as an alternative interpretation of possibility

The derivation in Section 3 agrees with the general law’s interpretation in terms of logical consistency (Bickel, 2012; Zhang and Zhang, 2013a; cf. Fossaluza et al., 2017), which is closely related to the deductive closure condition in the deductive cogency theory of Kaplan (1996). That theory is essentially possibilistic rather than probabilistic (Bickel and Patriota, 2019), and  $L(\bullet)$  is in fact a possibility measure. The conditional possibility measure generalization of  $L(\bullet)$  is dual to a necessity measure as defined in possibility theory (§2.5). While any possibility measure is mathematically an upper probability function (Walley and Moral, 1999), interpreting  $L(\bullet)$  under the probabilistic coherence theories such as those of de Finetti (1970) and Walley (1991) differs from interpreting it on the basis of logical consistency (cf. Spohn, 2012; Bickel, 2012; Zhang and Zhang, 2013a). For that reason, arguments that  $L(\bullet)$  performs poorly as an upper subjective probability are not necessarily relevant to its use in Section 3.