Chapter 2
Bayesian and Evidential Paradigms

Abstract The first step is to distinguish two questions:

1. Given the data, what should we *believe*, and to what degree?
2. What kind of *evidence* do the data provide for a hypothesis H_1 as against an alternative hypothesis H_2, and how much?

We call the first the "confirmation", the second the "evidence" question. Many different answers to each have been given. In order to make the distinction between them as intuitive and precise as possible, we answer the first in a Bayesian way: a hypothesis is confirmed to the extent that the data raise the probability that it is true. We answer the second question in a Likelihoodist way, that is, data constitute evidence for a hypothesis as against any of its rivals to the extent that they are more likely on it than on them. These two simple ideas are very different, but both can be made precise, and each has a great deal of explanatory power. At the same time, they enforce corollary distinctions between "data" and "evidence", and between different ways in which the concept of "probability" is to be interpreted. An Appendix explains how our likelihoodist account of evidence deals with composite hypotheses.

Keywords Confirmation · Evidence · Bayesianism · Likelihoods · Interpretations of probability · Absolute and incremental confirmation · Lottery paradox · Composite hypotheses

P.S. Bandyopadhyay et al., *Belief, Evidence, and Uncertainty*,
Philosophy of Science, DOI 10.1007/978-3-319-27772-1_2

Two Basic Questions

Consider two hypotheses, $H_{1,}$ that a patient suffers from tuberculosis, and $H_{2,}$ its denial. Assume that a chest X-ray, administered as a routine test for the presence of tuberculosis, comes out positive. Given the datum that the test is positive, and following the statistician Richard Royall's lead, one could ask three questions:[1]

1. Given the datum, what should we *believe*, and to what degree?
2. What kind of *evidence* does the datum provide for H_1 against an alternative hypothesis $H_{2,}$ and how much?
3. Once questions 1 and 2 have been answered, what should we *do*?

We call the first question the *confirmation question*, the second the *evidence question,* and the third the *decision question*.[2] Like Royall, we think that they are very different questions. Our concern in this monograph is with the first two.[3] A number of answers have been given to each. We want to use two of these answers to make clearer an intuitive distinction between confirmation and evidence which the questions presuppose, and then to show how this distinction both advances our understanding of uncertain inference and provides solutions to notable epistemological puzzles. To this end, and for illustrative purposes, we draw on the Bayesian and Likelihood statistical paradigms, the first to make precise a conception of confirmation, the second of evidence. Each is at least somewhat familiar to many philosophers and scientists, and all statisticians. In Chap. 3, we will show in more

[1]See Royall (1997). That the distinction between belief and evidence questions is pre-theoretically intuitive is underlined by the fact that Royall himself is a Likelihoodist who eschews any reference to an agent's subjective degrees of belief (he is, however, a Bayesian in regard to the decision question). Despite the philosophical differences that one or another of us has with him, our monograph owes a great deal to his work. See in particular Royall (2004).

[2]For a Bayesian response to Royall's three questions, see Bandyopadhyay (2007).

[3]Mark Kaplan (1996) is one of the very few philosophers to take note of the confirmation/evidence distinction (p. 25, footnote 32), but his argument for making it seems to involve no more than a reference to Nelson Goodman (quoted on p. 26, footnote 34), to the effect that "[a]ny hypothesis is 'supported' by its own positive instances; but support … is only one factor in confirmation." Kaplan's own very interesting and extensive account of evidence is itself generally "Bayesian" and makes no use of likelihoods. Goodman thinks that since incompatible and arbitrary hypotheses are "supported by the same evidence," there must be another "linguistic" (data-independent) factor involved in "confirmation." As we will see in Chap. 9, Goodman's well-known "grue paradox," which he uses to argue for this claim, depends on running "confirmation" and "evidence" together. Others who have made a confirmation/evidence distinction include Ellery Eells and Branden Fitelson in their (2000) and Malcolm Forster and Elliott Sober in their (2004). Forster and Sober are neither Bayesians nor Likelihoodists, which fact underlines our claim that the distinction is pre-theoretical, that is to say, statistical-paradigm independent.

detail how confirmation and evidence differ, and as a corollary how "data" are to be distinguished from "evidence." In Chap. 4 we consider, and then reject, four very general objections that have been made to the kind of account we set out in this monograph. In Chaps. 5, 6, 7 and 8 we will discuss some modifications of and alternatives to them.

Probabilities and Beliefs

On the Bayesian way of construing what we (and not Royall) call the "confirmation question", the answer has to do with *beliefs* and with *belief probabilities*. On it, *data confirm (disconfirm) a hypothesis just in case they raise (lower) our degree of belief in it.* In John Earman's vocabulary, we "probabilify" hypotheses by the data.[4]

There are at least three reasons for adopting a belief-probabilistic approach, over and above various difficulties with the alternatives to doing so.

First, such an approach reflects the inductive character of the relation between data and hypotheses. As we emphasized in Chap. 1, hypotheses "go beyond" the data enlisted to support them in the sense that all of the data gathered might be correct and yet the hypothesis, on the basis of further such data, be false.[5] One way in which to put this is to say that hypotheses are never more than probable given the data.

Second, the mathematical theory of probabilities is well understood, universally accepted, and precise. It allows for the methodical re-adjustment of the probabilities of hypotheses over time in the light of new or different data, and thus captures the non-monotonic character of inductive inference. In this way it also allows us to rank-order hypotheses with respect to the data that have been gathered or observed.

Third, the use of probabilities affords simple and straightforward measures of *relevance*. Thus data D are relevant to the confirmation of a hypothesis H just in case $\Pr(H \mid D) \neq \Pr(H)$. Similarly, an auxiliary hypothesis H' is useful in predicting data from a target hypothesis H just in case $\Pr(D \mid H \,\&\, H') \neq \Pr(D \mid H)$, and so on for many other examples.

There is more controversy in construing probabilities in this particular context as degrees of "belief", something in the heads of agents, and in this sense "subjective." The construal can be bolstered by two considerations.[6]

---

[4]Earman (1992).

[5]Although most of the hypotheses we will use to illustrate our argument do not take the form of universal conditionals, "All A are B," it is especially clear in their case that the claims they express typically outrun the inevitably finite data gathered to support them.

[6]Understanding probabilities as degrees-of-belief and connecting them to confirmation has a long history. See, for example, (Keynes 1921, pp. 11–12). Carnap, too, thought that inductive probability, i.e., the probability of the conclusion of an inductive argument given its various data premises, is "ascribed to a hypothesis with respect to a body of evidence… To say that the hypothesis h has the probability p (say 3/5) with respect to the evidence e, means that for anyone to

First, confirmation has to do with something like a "logical" (although not deductively valid) relation between data and hypotheses, i.e., it has an inferential structure. But inferential structures are propositional in character, they express relations between something like sentences. Beliefs on the usual account are propositional attitudes, they have the sort of sentential form that allows them to enter into inferences.[7] Beliefs like propositions are the bearers of truth and falsity, and probabilities attach to both.

Second, Royall's "what to do?" question presupposes that the first be expressed in terms of degrees of belief. For on the usual Aristotelian account, we act not on the basis of what is true or somehow established, but on the basis of what we believe to be true or established. For Bayesians, confirmation is linked to action by way of the concept of belief.[8]

To say that this account of confirmation is probabilistic, and that probabilities in connection with it are identified with degrees of belief, is to say that this account of confirmation is generally "Bayesian."[9] What makes it more specifically Bayesian is that central importance is accorded to Bayes Theorem, a way of conditioning

(Footnote 6 continued)

whom this evidence and no other relevant knowledge is available, it would be reasonable to believe in h to the degree p, or, more exactly, it would be unreasonable for him to bet on h at odds higher than $[p(h)/p(1\text{-}h)]$…. Thus inductive probability measures the strength of support given to h by e or the *degree of confirmation* of h on the basis of e (Carnap 1950, p. 441) As Skyrms (1986, p. 167) summarizes the situation, the concepts of inductive and epistemic (which, as we saw in Chap. 1, applies to statements rather than arguments) probabilities were introduced … as numerical measures grading degree of rational belief in a statement and degree of support the premises give its conclusion…. Why should epistemic and inductive probabilities obey the mathematical rules laid down for probabilities and conditional probabilities? One reason that can be given is that these mathematical rules are *required* by the role that epistemic probability plays in rational decision" (our italics). James Hawthorne helped prepare this brief history. See his (2011). For our purposes, it is as important to note that neither Keynes, nor Carnap, nor Skyrms distinguishes between confirmation and evidence (as the title of the selection from Carnap's work in Achinstein 1983, "The Concept of Confirming Evidence," makes clear).

[7] Although it is not needed for our argument, it is worth mentioning that a confirmation-generalization of logical entailment has been worked out by Crupi et al.(2013).

[8] On the traditional account of voluntary action, an action is voluntary just in case it is performed by a rational agent on the basis of her desires and beliefs. For Skyrms and other Bayesians, "rational agency" requires at a minimum that the agent's beliefs conform to the rules of the theory of probability.

[9] There are other ways in which to model belief within the context of a confirmation theory, for example, the Dempster-Shafer belief function. See Shafer (1976). Since the probability-based account is well-known and has a long tradition, we are resorting to it.

degrees of belief on the data gathered. At any time t_i, Bayes Theorem (to be described very shortly) tells us to what extent it is reasonable to believe a particular hypothesis given the data. On this account of confirmation, an agent should change her degree of belief in a hypothesis H from t_i to t_{i+1} by the amount equal to the difference between the posterior probability of H, $\Pr(H \mid D)$, and $\Pr(H)$, its prior probability.

Bayesian confirmation theory is "subjective" in the sense that all probabilities are identified with degrees of belief and confirmation with the way in which new data raise the probability of initially-held or prior beliefs. That said, many philosophers believe that some Bayesian approaches are more "objective" than others, depending on the constraints put on the determination of prior probabilities.[10] Those who reject a Bayesian approach in toto do so entirely on the basis of such "subjectivity", large or small, which in their view is incompatible with the objectivity of science, although they also allege that subjective belief leads to paradox. We will try to unravel the alleged paradoxes later in this chapter and in Chap. 9. But we will not try to defend the Bayesian approach against every challenge. We think that the rule of conditional probability on which it rests makes precise the notion of learning from experience, and that learning from experience is the intuitive basis of all confirmation. Our use of the Bayesian approach in this monograph, however, is purely instrumental, and is intended to make the distinction between confirmation and evidence as sharp as possible.

Bayesian Confirmation

The account of confirmation we take as paradigm involves a relation, $C(D, H, B)$ among data D, hypothesis H, and the agent's background information B.[11] However it is further specified, it is modeled on the basic rules of probability theory including the rule of conditional probability, together with some reasonable constraints on one's a priori degree of belief in whatever empirical hypothesis is under consideration. If learning from experience is to be possible, one of these constraints is that the agent should not have an a priori belief that an empirical hypothesis is true to degree 1, i.e., with full certainty, or 0, in which case it would have to be self-contradictory. This said, the agent learns from experience by up-dating her

[10]See Bandyopadhyay and Brittan (2010).

[11]Except when it is important, we leave out reference to the background information in what follows.

degrees of belief that hypotheses are true by conditionalizing on the data as she gathers them, i.e., in accord with the principle, derivable from probability theory, that $\Pr(H \mid D) = \Pr(H)\Pr(H \,\&\, D)/\Pr(D)$. Assuming that $\Pr(D) \neq 0$, her degree of belief in H after the data are known is given by $\Pr(H \mid D)$. Thus, D confirm H if and only if $\Pr(H \mid D) > \Pr(H)$. Call this the *Confirmation Condition.*[12] It is qualitative, i.e., compares the probabilities of a hypothesis before and after data have been collected, the intuitional basis of this conception of confirmation. This definition rests, as do most Bayesian conceptions of confirmation as a probability measure, on the following principle: for any H, D_1, D_2,, the confirmation (disconfirmation) of H in the light of D_1 is greater (less) than the confirmation(disconfirmation) of H in the light of D_2 just in case $\Pr(H \mid D_1) > (<) \Pr(H \mid D_2)$.[13] This principle makes explicit that as the probability of the hypothesis increases (decreases) as a result of further data-gathering, so too does its degree of confirmation (disconfirmation). A quantitative notion of confirmation of a hypothesis at any given time, is measured, for instance, in terms of the difference between its prior and posterior probabilities.[14] A hypothesis is always confirmed to some degree if the confirmation condition is satisfied. Whether it is "low" or "high" depends on the particular confirmation measure chosen,[15] the implicit standards of particular scientific communities, and the purposes of the investigator. On its Bayesian reading, the posterior probability of a hypothesis H equals its prior probability multiplied by the probability of D given H, $\Pr(D \mid H)$, divided by the marginal probability of D, $\Pr(D)$:

$$\Pr(H|D) = \Pr(H)\Pr(D|\mathrm{H})/\Pr(D) \tag{1}$$

[12]Or as it is sometimes called, "the positive relevance condition." See Salmon (1983) for an extended argument in behalf of the primacy of this condition in an analysis of confirmation.

[13]Following Crupi et al. (2013). The article includes a long list of Bayesians who subscribe to this principle.

[14]Clark Glymour (in an e-mail comment to us) and Peter Achinstein (2001, especially Chap. 4) object that this sort of account has a counter-intuitive consequence, that the same data could confirm incompatible hypotheses to different degrees. But so long as our assignments of degrees of belief are consistent, i.e., do not violate the rules of probability theory, it is possible to be rationally justified in believing incompatible hypotheses to different degrees on the basis of the same data.

[15]That is to say, we use this as an exemplary measure of degree of confirmation. Many others are possible. See Fitelson (1999), for a discussion of the sensitivity of confirmational values to the measure used. We believe that the choice of a specific confirmation measure depends on the type of question one is asking. The same idea, that the type of question asked determines the measure chosen, applies to the evidence question as well. Although we have adopted the likelihood ratio to weigh evidence, different evidential measures would be required *if* we were to ask a different set of evidential questions. See Taper et al. (2008) and Chap. 5 for further discussion of alternative measures.

$\Pr(H \mid D)$ is also called the conditional probability of H given D. The prior probability of a hypothesis represents the agent's degree of belief that the hypothesis is true before (i.e., prior to) new data bearing on the hypothesis have been gathered. This agent-relative prior probability component of the definition is the most controversial element in the application of Bayes' Theorem, and will be discussed in more detail later.

The quantity $\Pr(D \mid H)$ is often referred to in the *philosophical* literature as the likelihood. While numerically the likelihood of the hypothesis given the data is equal to the probability of the data given the hypothesis, likelihood and probability are not the same thing; likelihood is considered a function of the hypothesis, whereas the probability is considered a function of the data. We here adopt the common philosophical notation of denoting the likelihood by $\Pr(D \mid H)$ rather than the common statistical notation of L $(H; D)$,[16] but do not mean to imply that the hypothesis H *needs* to be considered a random variable.

The likelihood function provides a tool, through the likelihood ratio, to answer the question, "How much support for a hypothesis is there in the data relative to another hypothesis?" The likelihood function is an important tool for Bayesians and non-Bayesians alike, but too rarely accorded the kind of importance that we do here.

The final element of Eq. 1 is Pr(D). This is calculated as the marginal probability of the data over the alternative hypotheses, that is, the probability that D would obtain, averaged over H and ~H:[17]

$$\Pr(D) = \Pr(H)\Pr(D|H) + \Pr(\sim H)\Pr(D|\sim H). \tag{2}$$

[16]In the eyes of many statisticians this notation signals the difference between "probability" and "likelihood," as two different concepts. More than a simple notational difference is involved. The " $\mid$ " notation indicates conditioning on a random variable, i.e., in the case of $\Pr(D \mid H)$ the hypothesis is a random variable, while $\Pr(D; H)$ indicates that the data are conditioned on a variable that is considered fixed. The first is fundamentally Bayesian, the second is fundamentally evidentialist.

[17]For convenience, hypotheses are most often presented as exhaustive pairs, H and $\sim H$, but in theory the list of hypotheses considered is not limited to such pairs. It is difficult, among other reasons, to compute the probability of the data under the catch-all hypothesis [$\Pr(D \mid \sim H)$, and in consequence it is difficult to calculate the posterior probability of the catch-all. This difficulty then extends to comparing the posterior probability of the catch-all with the posterior probabilities of other hypotheses. We avoid such difficulties by confining our discussion to *simple* hypotheses. It might be added here that on the present account of what is often called "incremental confirmation," the Special Consequence Condition does not hold see Salmon (1983, p. 106). To handle objections in this connection, Kotzen has produced a principle which he calls "Confirmation of Dragged Consequences:" If [$\Pr(H \mid D) > \Pr(H)$, and H_1 entails H_2, and $\Pr(H_2) < \Pr(H_1 \mid D)$] then $\Pr(H_2 \mid D) > \Pr(H_2)$. See Kotzen (2012).

Evidence and Likelihood Ratios

As we understand it, evidence invariably involves a comparison of the merits of two hypotheses,[18] H_1 and H_2 (possibly, but not necessarily, ~H_1) relative to the data D, background information B, and auxiliaries A.[19] This is to say that we distinguish "evidence" from "data." All evidence depends on data. But data constitute evidence only when, in context, they serve to distinguish and compare hypotheses. Four preliminary points might be made in this connection.

First, it is a commonplace that not all data constitute evidence. Whether they do so or not depends on the hypothesis being tested, whether the data "fit" or are relevant to appraising the hypothesis. Whether data constitute evidence is a matter of context. Data themselves, which are often taken as a first approximation as the reports of observations and the results of experiments, are in this sense context-free.

Second, data are paradigmatically invoked as evidence in the history of science in a comparative context; Galileo's sighting of Jupiter's moons was rightly taken as evidence for the truth of the Copernican as against the Ptolemaic hypothesis.

Third, a comparative account of evidence meets the demand of one key goal of science. The goal is to understand and explain natural phenomena. To do so, scientists propose descriptive-explanatory models of these phenomena, and seek better and better models as closer and closer approximations to the truth. We think that there is in some meaning of the words, truth or reality which we study with the help of models. But, none of the models are true since they all contain idealizations which are false. Our evidential framework is designed to quantify in what sense and by how much one model is a superior approximation to the truth about natural phenomena than another.

Fourth, a distinction between data and evidence first allows us to understand and then helps resolve many if not all of the difficulties that beset traditional theories of confirmation. In Chap. 9, for example, we take up "the old evidence problem" for subjective Bayesian accounts of confirmation. Unless a distinction between data and evidence is made, the problem is trivial. To the question "does old evidence provide evidence for a new theory?" the obvious answer is "yes, of course." But the

[18]Statisticians prefer to use the term "models," by which they mean statistical models that allow for quantitative testing vis-à-vis the data, rather than "hypotheses" in this connection. Although we mostly stick to the more general (albeit vaguer) philosophical usage, the difference in meaning between the terms is important. As we emphasize at the end of this chapter in connection with various interpretations of probability, a hypothesis is a verbal statement about a natural state or process, a model is a mathematical abstraction that captures some of the potential of the hypothesis. Although we occasionally use the terms interchangeably, it matters a great deal whether one has models or hypotheses in mind when it comes to the correct characterization of the hypothesis-data and model-data relationships.

[19]Again in what follows, and except where it matters, we will not include reference to A or B in our formulations.

question is no longer trivial when it is re-phrased: "do old data provide evidence for a new theory?" The answer to this second question requires serious reflection.[20]

On our account of evidence, a model or hypothesis is a defined data-generating mechanism. Observed data support model_1 over model_2 if the data that would be generated by model_1's mechanism are by some measure more similar to the observed data than the data that would be generated by model_2. Thus on our account, data provide evidence for one hypothesis against its alternative *independent of what* the agent *knows or believes* about either the available data or the hypotheses being tested.[21] A subtle but important distinction in this connection is between knowing that the data are available and knowing the probability of the data. One can know how probable observations of the data are without knowing whether the data have actually been observed. For example, one could calculate the probability of obtaining 100 tails out of 100 flips of a coin on the hypothesis that the coin is biased towards tails with a probability of 0.9. This differs from asserting that we know on the basis of the data that out of 100 flips a coin has landed tail-side up 90 times.

One final preliminary. We have assumed for the sake of clarity and convenience that the hypotheses in our schematic examples are simple and not complex. Error-statisticians object that clarity and convenience have little to do with it; we are forced to make this assumption in the case of our account of evidence because it cannot deal with composite hypotheses.[22] The issues involved are technical, and for that reason we have put our discussion of them in an Appendix to this chapter. Suffice it to say here that this objection can be met.

The Evidential Condition

Now back to our characterization of evidence. It is made precise in the following equation:[23]

$$D \text{ is evidence for } H_1\&B \text{ as against } H_2\&B \text{ if and only if } \mathrm{LR}_{1,2} = \left[\frac{\Pr(D|H1\&B)}{\Pr(D|H2\&B)}\right] > 1$$

[20]Robert Boik pointed this out to us.

[21]It thus stands in sharp contrast to the well-known position of Williamson (2000), whose summarizing slogan is "a subject's evidence consists of all and only the propositions that the subject knows." Williamson is not analyzing the concept of "evidence," or more precisely "evidence for a hypothesis or theory," but the concept of "evidence for a subject." This concept is important in classical epistemology, but has little use, or so we believe, as applied to scientific methodology or theory assessment. For us, evidence is evidence, whether or not the subject knows it, and conversely, whether or not the subject knows something does not thereby qualify it as "evidence" for any particular hypothesis.

[22]See Mayo (2014).

[23]We use "LR" rather than "BF" in what follows to underline the fact that our account of evidence is not in any narrow sense "Bayesian.".

Call this the *Evidential* condition. It serves to characterize data which, in context, play an evidential role. But a quantitative measure is immediately possible by taking account of the numerical ratio of the two hypotheses. Note in this connection that if $1 < \mathrm{LR} \le 8$, then D is often said to provide "weak" evidence for H_1 against H_2, while when $\mathrm{LR} > 8$, D provides "strong" evidence for H_1 over H_2 (Royall 1997). This cut-off point is sometimes determined contextually by the relevant scientific communities and may vary depending on the nature of the problem confronting the investigator, but it follows a statistical practice common among investigators.[24] It follows from the Evidential Condition that the range of values for the LR can vary from 0 to ∞ inclusive.[25]

Since Bayesians by definition do not assign the end-point probabilities 0 or 1 to any empirical proposition, it follows that an agent's (subjective) degree of belief in the hypotheses over which she has distributed prior probabilities does not affect whether D is evidence for H_1 as against H_2.[26] This is especially evident if we assign the extreme probability 0 to H_1. In that case, the LR for H_1 against H_2 relative to D becomes $\Pr(D \mid H_1)/\Pr(D \mid H_2) = [\Pr(D \,\&\, H_1)/\Pr(H_1)]/\ [\Pr((D \,\&\, \mathrm{H2})/\Pr(H_2)] = 0/0$, which is undefined.[27]

We have argued that evidence has to do with the degree to which data help distinguish the merits of competing hypotheses. Several measures are available to capture the resulting evidential strength of one hypothesis as against another. We have fixed on ratios of likelihood functions as our exemplar in this monograph for three reasons. First, it is the most *efficient* evidential function in the sense that we can gather strong evidence for the best model with the smallest amount of data.[28] Second, the LR brings out the essentially comparative feature of evidence in a clear and straightforward way. Third, and as we have pointed out, it is a measure of

[24]Royall (1997) points out that the benchmark value = 8, or any value in that neighborhood, is widely shared. In fact, the value 8 is closely related to the Type I error-rate 0.05 in classical statistics and to an information criterion value of 2. See Taper (2004) and Taper and Ponciano (2016) for more on this issue.

[25]One might argue that the posterior-prior ratio measure (PPR) is equal to the LR measure and therefore precludes the necessity of a separate account of evidence. But the objection is misguided. The LR is equal to $\Pr(H \mid D)/\Pr(H)$ *only when* $\Pr(D)/\Pr(D \mid {\sim}H)$ is close to 1. That is, $[\Pr(D \mid H)/\Pr(D \mid {\sim}H)] = \Pr(H \mid D)/\Pr(H) \times [\Pr(D) \times \Pr(D/{\sim}H)] \approx 1$. Otherwise, the two measures, LR and PPR, yield different values, over and above the fact that they measure very different things.

[26]Berger (1985, p. 146). Levi (1967) also emphasizes the objective character of the likelihood function.

[27]We are indebted to Robert Boik for this clarification.

[28]See Lele (2004) for the proof. It is worth noting the comments of a referee on a similar claim in another paper of ours: "…if in the end we hope for an account of evidence in which evidence gives us reason to *believe*, it is totally unclear why efficiency in the sense described would be taken as a sign of a meritorious measure of evidence." One of the principal aims of this monograph is to disabuse its readers of what we call the "true-model" assumption, that evidence gives us reasons to believe that a hypothesis is true. In our view, evidence provides information about a hypothesis on the basis of which we can make reliable inferences and not reasons to believe that it is true. Since evidence is a way of identifying and assessing such information, efficiency in the sense indicated is indeed a meritorious property of it.

evidence embraced by such otherwise diverse approaches to statistical inference as Bayesian and Likelihoodist (Evidentialist) insofar as we confine ourselves to simple statistical hypotheses. We are in good company.

We wish to be clear, however, that the likelihood ratio is a special case of a rich family of evidence measures or functions.[29] In principle, an evidence function, Ev (D, M_1, M_2), is the difference between the statistical distance[30] between the probability distributions generated by each of the models, M_1 and M_2, and an estimate of the "true" underlying distribution based on D. That is to say,

$$Ev(D, M_1, M_2) = n(SD(\hat{\tau}_D, M_1) - SD(\hat{\tau}_D, M_2)).$$

where n is the sample size, SD(•,•) is a statistical divergence (distance) between two probability distributions indicated by place-holders, $\hat{\tau}_D$ is the estimate of the "true" distribution based on the data D. For discrete distributions, this could be simply the proportion of observations in each discrete data category. For continuous distributions, it will be some smoothed estimate such as a kernel density estimate (Silverman 1986). The inclusion of n in the formula for evidence conforms to the intuitive expectation that the strength of evidence should increase with increased amounts of data.[31]

The most famous statistical distance is the Kullback-Leibler Distance (KLD), the expected log likelihood ratio between two distributions. For discrete distributions, P and Q with categories indexed by I, this is given as: $KLD(P, Q) = \sum_i P_i \cdot [\log(P_i) - \log(Q_i)]$

It is easy to show that the log (LR) is the KLD-based evidence function for comparing simple hypotheses. Similarly, the differences of information criterion values are KLD-based evidence functions for cases where the models compared differ in the number of parameters estimated.[32]

---

[29]What follows is drawn from Lele (2004). We go into the details, technical as some of them are, simply because evidence functions are so much less familiar than confirmation functions.

[30]In fact, evidence functions only use the weaker criterion of divergences (statistical distances are divergences with some additional constraints). A divergence quantifies the dissimilarity between two probability distributions. The statistical literature is quite loose in its use of the terms "divergence," "disparity," "discrepancy," and "distance." We use "distance" rather than the more general term "divergence" because "distance" is more intuitively evocative. Good discussions of statistical distances can be found in Lindsay (2004) and Basu et al. (2011).

[31]Although it does not do so in a linear fashion.

[32]The technical definition of evidence functions (Lele 2004) includes a consistency requirement (i.e., the probability of correctly selecting the best approximating model must go to 1 as sample size goes to infinity). Thus only "order consistent" information criteria such as Schwarz's criterion (variously denoted as the SIC or BIC) can be used to construct evidence functions.

A large number of potential statistical distances could be used to construct evidence functions in response to different types of evidence question. For example, one alternative to the KLD is the Hellinger distance.[33] Evidence functions based on different distances will have different statistical properties. For instance, while KLD-based evidence functions will be maximally efficient, Hellinger-based evidence functions will be more resistant to the presence of outliers. But none of the following discussion turns on these technical details.

Absolute and Incremental Confirmation

An important objection that has been made to our account is that its main theme, that confirmation and evidence should be distinguished, has already been developed in the literature in terms of confirmation alone.[34] The objection is that both (Carnap 1950) and (Salmon 1983) long ago made a parallel distinction between "absolute" and "incremental" confirmation which does all of the work that ours does without recourse to any distinct notion of evidence. On the absolute concept, sometimes also called "the high probability requirement," the data confirm H if and only if the probability of H given D exceeds some suitably high threshold, say, 0.9 (or minimally >0.5). It thus picks up on an ambiguity in the word "confirmed" already introduced. In certain contexts, the word connotes something like "well confirmed" or "put to rest", as in "Well, we certainly managed to confirm that guess." In this respect it is to be contrasted with the incremental way in which we confirm hypotheses or for that matter hunches, gathering evidence as we go along, becoming more and more sure that our hunch was right or wrong.

On the incremental conception of confirmation favored by most Bayesians, the data D confirm H if and only if the data raise the probability of H relative to its prior probability. It is how we understand confirmation here. Although there are important differences between the incremental and absolute conceptions, neither can be used to explicate, still less is tantamount to, anything like the notion of evidence that we and many others draw on. First, and as we will see in more detail in the next chapter, strong evidence does not entail a high degree of confirmation, either absolute or incremental. Second, confirmation of both varieties is sensitive to an agent's prior probability distribution and endorses a rule for up-dating degrees of belief, whereas evidence is insensitive to prior probabilities, characterizes the relation between data and hypothesis regardless of whether the agent knows or believes that either data or hypothesis is probable, let alone certain or true, and indicates no way in which to up-date degrees of belief. Third, and unlike our

[33](HD): $HD(P,Q) = {}^{1}\!/\!\sqrt{2}\sqrt{\sum_i \left(\sqrt{P_i} - \sqrt{Q_i}\right)^2}$ for discrete distributions.

[34]Two previous readers raised this objection. We are assuming the standard account of absolute confirmation. Please see the following footnote for more on this point.

characterization of evidence, absolute confirmation is, as its name suggests, absolute; if one restricts what counts as evidence to data that confirm a hypothesis absolutely, there is no way in which to determine whether some evidence is stronger than others, still less a way to quantify its strength.[35] Given a prior $\Pr(H) = 0.2$ and a posterior $\Pr(H \mid D) = 0.901$, H is intuitively much more strongly confirmed than in a parallel situation in which $\Pr(H) = 0.899$ and $\Pr(H \mid D) = 0.90$, but the notion of absolute confirmation is unable to capture this intuition. Still worse, it undermines it.

Finally, and perhaps most importantly, the concept of absolute confirmation, unlike the more standard conception of evidence, applies to hypotheses considered singly and not pair-wise, i.e., it does not necessitate a separation and comparison between rival hypotheses which is, we contend, at the center of what we should expect a measure of evidence to provide; again as in courts of law, evidence is what distinguishes the guilty from the blameless and indicates which data determine who is guilty and who not, and to what degree.

Quite apart from its failure to capture the intuitive concept of evidence, the notion of absolute confirmation has its own difficulties. We mention two because each throws more light both on the incremental conception and on the distinction with evidence that we use it to make.

One difficulty with absolute confirmation is that it runs directly into the lottery paradox.[36] Suppose a fair lottery with a thousand tickets. Exactly one ticket will win and, since the lottery is fair, each stands an equal chance of doing so. Consider the hypothesis, "ticket #1 will not win." This hypothesis has a probability of 0.999. Therefore we have a conclusive reason, on the absolute conception, to believe that it is true. But the same line of reasoning applies to all of the other tickets. In which case, we should never believe the hypothesis that any one of them will win. But we know, given our initial supposition, one of them will win. The paradoxical result can be avoided by denying that any hypothesis is ever absolutely confirmed.[37]

The lottery paradox also exposes one way in which our distinction between (incremental) confirmation and evidence is of real use. Sober uses the lottery paradox to argue for a wholesale rejection of the notion of "acceptance." First, to accept a hypothesis is to have a good reason for believing that it is true (or empirically adequate, or whatever). But the converse does not hold. However good our reason for believing it to be true, however well-confirmed, we might still not accept the hypothesis. We might not accept it because the otherwise confirming

[35]There is a non-standard account of "absolute confirmation" in the literature on which it does admit of degrees; on this account a hypothesis is "absolutely confirmed" if it is "confirmed strongly," where "confirmed strongly" can have different degrees. See Eells (2006, p. 144). Our argument depends on the standard account, on which D confirm H "absolutely" just in case $\Pr(H \mid D) = r$, where r is identified as a particular number (or, occasionally, any number greater than 0.5).

[36]First propounded by Kyburg (1961). We follow Elliott Sober's informal statement of it (1993).

[37]Kyburg himself avoids the paradox by denying the uncritical "conjunction principle" on which it rests, that if each of a set of hypotheses is accepted, then their conjunction must be as well.

data are not evidentially significant.[38] They are not evidentially significant, on our characterization, when they fail to distinguish between competing hypotheses. In the lottery case, the likelihoods of all of the competing hypotheses, that is, the likelihood of cashing a winning ticket on the hypothesis that it is not a winning ticket, are the same. In which case, the data fail to distinguish between the competing hypotheses, in which case they are, in context, evidence for none of them. If acceptance requires evidential significance, as we will in more detail in Chap. 6, then we should not accept any of the hypotheses.

The other difficulty with the absolute conception of confirmation is that it entails what is sometimes called the inconsistency condition: the data can never confirm incompatible hypotheses. But when more than two hypotheses are at stake, the data can and do incrementally confirm more than one, however incompatible they might be.[39]

Our Two Accounts and Interpretations of Probability[40]

Before proceeding further, we need to make explicit what has been implicit from the outset, that adequate accounts of confirmation and evidence presuppose different interpretations of the concept of "probability", and therefore different readings of the probability operator in our various conditions and equations as it is applied to events, hypotheses, and propositions. What is important to our fundamental distinction is not so much the details of these readings, still less the widespread and often intense controversies to which they have given rise, but that some

[38]This is our view, not Sober's.

[39]A well-known example was provided by Popper (1959, p. 390). "Consider the next throw with a homogeneous die. Let x be the statement 'six will turn up'; let y be its negation…; and let z be the information 'an even number will turn up'. We have the following absolute probabilities: $p(x) = 1/6$; $p(y) = 5/6$; $p(z) = 1/2$. Moreover, we have the following relative probabilities: $p(x, z) = 1/3$; $p(y, z) = 2/3$. We see that x is supported by z, for z raises the probability of x from 1/6 to 2/6 = 1/3. We also see that y is undermined by z, for z lowers the probability of y by the same amount from 5/6 to 4/6 = 2/3. Nevertheless, we have $p(x,z) < p(y, z)$." Popper mistakenly drew the conclusion that there was a logical inconsistency in Carnap's confirmation theory. But the inconsistency follows only if we were to take confirmation in its "absolute" sense, i.e., just in case the data raise the probability of a hypothesis beyond a high threshold. There is no inconsistency if confirmation is taken, as we do, in its incremental sense. See also Salmon (1983, pp. 102−03).

[40]John G. Bennett has drawn our attention to the need to address the varying interpretations of probability involved in our accounts. In a very helpful e-mail communication, Marshall Abrams has underlined some of the difficulties in calculating probabilities from probabilities defined by two different interpretations and in taking as paradigm some of the ways in which the likelihood relationship between data and hypothesis has been construed. As for the first, we offer reasons in what follows for taking posterior probabilities as subjective despite their embedding objective components. As for the second, our account of objective probabilities uses empirical frequencies to estimate and not to "define" probabilities.

of them are generally "subjective" and others "objective", and that there are "objective" ingredients in our account of confirmation as well as in our account of evidence. In our view, one needs to mix and match probabilities of different kinds and not rely on one to the exclusion of the others. As we will now show, to do otherwise would be to seek an overly monistic methodology.

Consider first our account of confirmation. It is informed by a subjective Bayesianism. *D* confirm *H* just in case an agent's prior degree of belief in *H* is raised. The degree to which *D* confirm *H* is measured by the extent to which the degree of belief has been raised. The probabilities involved have to do with belief, and are in this sense subjective. But determining an agent's degree of belief in a hypothesis, $\Pr(H \mid D)$ requires determining the probability of *D* on *H*, $\Pr(D \mid H)$, which is independent of an agent's belief and is therefore objective.

One possible response to this claim is that the probability function is mistakenly construed as objective; it needs to be construed as a conditional belief-probability in which the probability operator is understood in *subjective* terms. But even otherwise staunch Bayesians are not in full agreement to do so, witness L.J. Savage's admission in his famous 1963 paper with W. Edwards and H. Lindman[41] that likelihoods are "public," implying that they are objective. As Isaac Levi points out in this connection,[42] likelihoods are agent-invariant and fixed, and in this sense objective, and unlike variable conditional probabilities which are agent-dependent.[43] At the same time, subjectivity spreads through a compound probability like falsehood through conjunctions and truth through disjunctions; since the posterior probability, $\Pr(H \mid D)$, is calculated in part on an agent's prior degree of belief it is subjective, as should be expected on a Bayesian account of confirmation.

There is a possible justification from the subjective Bayesian standpoint regarding how one could mix and match those two types of probabilities together. David Lewis argues that when the chance probability of a proposition, A, is available, one needs to set an agent's subjective probabilities, $\Pr(A|M) = \Pr(M)$ where M is a statistical model (Lewis 1980.)[44] Lewis thinks that this alignment of subjective probability with objective chance is possible because of the *Principal Principle*. The idea behind this principle is that when we have the likelihood of a

[41]Edwards et al. (1963).

[42]Levi (1967).

[43]This is perhaps clearest when *D* are entailed by *H*, for in this case $\Pr(D \mid H) = 1$, regardless of what anyone happens to believe.

[44]We are indebted to both John. G. Bennett and Colin Howson for some clarifications about the need to introduce the *Principal Principle*. For more on the justification of the Principal Principle, see the forthcoming book by Robert Pettigrew, *Accuracy and the Laws of Credence*, Part II of which is devoted to it. We are indebted to Jason Konek for calling our attention to this important discussion.

model given the data (what we have just called "propositions") available we should treat an agent's subjective probability which is a conditional probability of the data given a model to be equal to its objective likelihood. This alignment of the subjective probability with the objective chance helps treat both probabilities, the likelihood functions, and prior probabilities, in the application of the Bayes theorem as subjective.[45]

Consider, second, our evidentialist account. It makes no use of belief-probabilities and in this sense is thoroughly objective. But in a way generally similar to the account of confirmation, it too mixes and matches interpretations of the probability operator in a way that without explication might invite misunderstanding but that, though nuanced, is necessary to get our inference-engine running. The likelihood of a model given the data is an essential component of the account. For simplicity's sake, we identify it with a conditional probability, $\Pr(D_0|M)$, where "D_0" stands for a yet to be realized model-generated datum and M is the generating model. What data will be realized a priori depends on what sorts of models we have proposed. At the initial stage, when no data have yet been realized, the relationship between a model[46] and its unrealized data is deductive. One could come up with an open-ended number of distinct models, M_1, M_2, M_3,…M_k. Let's assume M_1 says that the coin has 0.9 probability to land with heads. Assume further that M_2 says that it has 0.3 probability, M_3 says that it has probability 0.5, and so on. Given these models, before any real data have been observed, each will tell us how probable any set of observations would be under the model. So, the relationship between M_1 and D_0 is completely deductive. So, too, are the relationships between other competing models and $_0$. Now assume that a real coin has been flipped. Further assume that out of 100 flips, 60 of them have landed with heads and the rest are tails. Let's call the data D_1 60 heads and 40 tails. The relationships expressed via $\Pr(D_1|M_1)$, $\Pr(D_1|H_2)$ …..$\Pr(D_1|H_k)$ continue to be deductive. We find that each model tells us how probable those data are under that model, although the probability values will vary from one model to the next.

It is natural to assume that the "propensity" of a model to generate a particular sort or set of data represents a causal tendency on the part of natural objects being modeled to have particular properties or behavioral patterns and this tendency or

[45]Given the limited scope of this Monograph, we are not going to evaluate rigorously whether this consideration is well-grounded.

[46]As we noted earlier and for the reason given, we use "model" (less often) and "hypothesis" (more often) interchangeably. But there are crucial differences between them that in the present context might give rise to misunderstandings. A hypothesis is a verbal statement about a state of nature or some natural process. A model is a mathematical abstraction that captures some dimensions of the hypothesis. When we say that the likelihood relationship between data and hypothesis is "logical," this is not in the precise sense in which Keynes or Carnap use the term; it has to do rather with the evaluation of a single hypothesis based on the statistical data. On the other hand, the calculation of the probabilities of the data given two or more models is in our sense of the term "deductive." The central point is that the relationships between data and hypothesis and data and models must be kept distinct; our use of the words "logical" and "deductive" is intended to do so, whatever other connotations those words might carry.

"causal power" is both represented and explained by a corresponding hypothesis. According to the propensity interpretation of probability,[47] the probability of an event is a dispositional property of it.[48] In our case, the probability of a coin's landing heads, if it is flipped, is estimated at 0.6. The theme behind this interpretation is that the probabilistic statement is true because the object in the nature of the case possesses a special sort of dispositional property, called a *propensity*. If a sugar cube would dissolve when immersed, the sugar cube has the dispositional property of solubility. In the same vein, if a coin has a 0.6 probability of landing heads when flipped, the coin is to have a propensity of a certain strength to land heads when flipped. However, we have no way to know for sure what this tendency is except to estimate it through finite empirical frequencies.

The empirical interpretations of probability championed by von Mises[49] and Reichenbach[50] define probability in terms of the limits of relative frequencies; but the empirical sequences we encounter in the world are always finite, and we have often good reason to suppose that they cannot be infinite. Although there are significant mathematical considerations for that kind of idealization in their frequency interpretation, that is, that sequences are infinite, insofar as scientific experiments are concerned this kind of idealization is not implausible. We cannot assume that in a physical experiment we have an infinite number of helium nuclei in various stages of excitation. Nor should we assume in any case that the probability operator is to be "defined" in frequency terms.[51] Rather, finite frequency is no more than a way of estimating propensities in particular cases. Troublesome cases do not undermine an interpretation of "probability" so much as they raise difficulties for approximating values under certain conditions.

In other words, and in our view, it is misleading to say that there are three distinct interpretations of objective probability—deductive, propensity, and finite frequency. Rather, likelihoods allow us to choose models whose deductive probabilities best match natural propensities as approximated by finite frequencies. It is in this way, we think, that they are best to be understood, a case for their agent-independence and objectivity to be made, and a distinction with subjective belief-probabilities to be drawn. As we indicated at the outset, confirmation is in a

[47]See Popper (1957). See also Berkovitz (2015) for a defense of the propensity interpretation of probability against the traditional twin criticisms that the explication of propensities is circular and therefore non-informative and that it is metaphysical and therefore non-scientific.

[48]A doctor in Molière's *Le Malade Imaginaire* attributes a "dormitive power" to opium, becoming the object of ridicule then and for the next 300 years. But to attribute a dispositional property is not to provide a causal explanation but to aggregate natural processes which must then be explained at a deeper, in this case pharmaceutical, level of analysis. Sugar has a propensity to dissolve in water, but it has almost no propensity to dissolve in gasoline. This difference can be understood through the chemical structure of sugar, water, and gasoline.

[49]See von Mises (1957).

[50]See Reichenbach (1949).

[51]See Hàjek (1997).

fundamental sense "in the head", evidence in that same sense "in the world." The fundamentally two-fold interpretation of "probability" makes this sense clearer.

The Limits of Philosophical Taxonomy

Philosophical taxonomies play a useful role in locating positions on the conceptual map, drawing attention to their general features, and instructing students. But they can mislead. Although there is a functional, but not conceptual, connection between our characterizations of evidence and confirmation, they are, as just indicated, otherwise quite distinct, and it would be a mistake to bring them under some such common rubric as "Partial-Subjective Bayesianism" or "Probabilistic Evidentialism." Confirmation is to some degree "subjective", involving personal probabilities and the up-dating of beliefs. Evidence is not. Following Kant's guidance, our task here in any case is not to suggest a new statistical or hypothesis-testing paradigm, but to use two already well-established paradigms to draw a distinction which, once drawn and understood, will help chart a new direction in our thinking about uncertain epistemic inference, mitigate, to some extent, the overly adversarial nature of its discussion, and dissolve some traditional epistemological puzzles.

Appendix

A Note on the Likelihoodist Treatment of Simple and Composite Hypotheses

Bayesians use the Bayes Factor (BF) to compare[52] hypotheses (Kass and Rafferty 1995), while others use the likelihood ration (LR) to measure evidence. For simple hypotheses, as in the tuberculosis example discussed near the outset of the next chapter, the Bayes Factor and the Likelihood Ratio are identical; both capture the essential core of our analysis of the concept of evidence. Throughout the monograph we assume that the hypotheses being tested are simple statistical hypotheses, which specifies a single value to a parameter, in contrast to a compound or

[52]Glymour (1980), p. 102, rightly draws attention to what he calls "misplaced rigor." But rigor is otherwise indispensable. Error-statisticians have focused their criticisms of the evidential position on an alleged failure to deal with composite hypotheses; see Mayo (2014). This is our reply (following Taper and Lele 2011, Sects. 6 and 7). It is rather technical in character, and does not affect the main line of our argument. We have similarly added technical Appendices to Chap. 6 to reply to another error-statistical criticism of our evidential account, that it needlessly employs multiple models in its account of hypothesis testing, and to Chap. 11, to illustrate some of our main themes in more mathematical detail.

composite hypothesis which restricts a parameter θ only to a range of values.[53] Since some philosophers have claimed recently that the likelihood account of evidence cannot deal with composite hypotheses, it is worth our while to argue why they are mistaken.

Here is a test case:[54]

> [M]edical researchers are interested in the success probability, θ, associated with a new treatment. They are particularly interested in how θ relates to the old treatment's success probability, believed to be about 0.2. They have reason to hope θ is considerably greater, perhaps 0.8 or even greater. To obtain evidence about θ, they carry out a study in which the new treatment is given to 17 subjects, and find that it is successful in nine.

How would an evidentialist test the composite hypothesis that the true proportion (θ) is greater than 0.2?

The maximum likelihood or ML 0.5294. The $k = 32$ support interval for quite strong evidence is [0.233, 0.811]. Royall would say that for any value θ' outside of this interval, there is quite strong evidence for the maximum likelihood estimate or MLE as opposed to θ'. For an evidentialist, this is sufficient to infer with strong evidence that $(\theta) > 0.2$, even though the likelihood of the MLE is not the likelihood of the composite. The following considerations support this claim.

(1) There is quite strong evidence against any value outside the interval relative to a value inside the interval (i.e. the maximum likelihood estimate).
(2) No two values inside the interval can be quite strongly differentiated.
(3) (1) and (2) together imply that there is quite strong evidence that the true proportion θ is in the support interval [0.233, 0.811].
(4) Since 0.2 is entirely below the support interval, there is therefore quite strong evidence that the 0.2 is less than the true proportion.

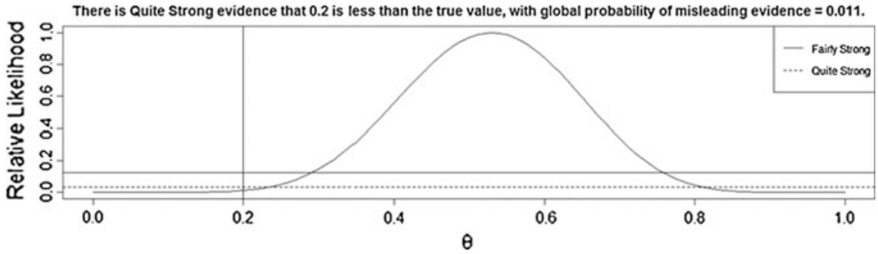
It does make explicit that there are limits on how high the true proportion is likely to be.

We will use a notion called "the probability of misleading evidence" which will be discussed in much detail in Chap. 8. The probability for the presence of evidence for a hypothesis is called misleading because although there is probability for the presence of the evidence for the hypothesis, the latter is in fact false. If one had set up $k = 32$ (quite strong evidence) then the probability of misleading evidence for this statement is $M_G < 1/32 = 0.031$. The M_L represents the probability of misleading *local* evidence *after* the data have been gathered. The M_G represents the probability of misleading *global* evidence *before* the data have been gathered. Both M_L and M_G represent a

[53] This assumption is common to different schools of statistics. Both Royall (1997), who is committed to a likelihoodist (and not, as already noted, Bayesian) approach, and Mayo (1996) who operates within the error-statistical framework, also take the assumption for granted (at least Mayo did in 1996, although she does so no longer; again see Mayo 2014).

[54] Royall (1997, pp. 19–20).

bound on the probability of misleading evidence for a hypothesis. The post hoc probability of misleading evidence is a little lower MG = 0.011.



Taper and Lele (2011) suggest that since there is an estimated parameter in finding the MLE, the M_G is biased high, and that composite intervals should use a biased corrected estimate of the likelihood. We use Akaike's bias correction (for simplicity). With the bias correction, the quite strong evidence support interval is a little bit wider at [0.201, 0.840]. The inference is still the same. There is still quite strong evidence that the true proportion is greater than 0.2, but now the post hoc probability of misleading evidence is slightly greater at 0.030. Using the more severe Swartz bias correction, we find that there is only fairly strong evidence for 0.2 being less than the true value, with a M of 0.045 (see also Taper and Ponciano 2016).

References

Achinstein, P. (Ed.). (1983). *The concept of evidence*. Oxford: Oxford University Press.

Achinstein, P. (2001). *The book of evidence*. Oxford: Oxford University Press.

Bandyopadhyay, P., & Brittan, G. (2010). Two Dogmas of Strong Objective Bayesianism. *International Studies in the Philosophy of Science, 24*, 45–65.

Bandyopadhyay, P., & Forster, M. (Eds.). (2011). *Handbook of the philosophy of statistics*. Amsterdam, Elsevier: North-Holland.

Bandyopadhyay, P. (2007). Why Bayesianism? A Primer on a Probabilistic Theory of Science. In Upandhyay, U. Sing & D. Dey, (Eds.). *Bayesian statistics and its applications*. New Delhi: Anamaya Publishers.

Basu, A., Shioya, H., & Park, C. (2011). *Statistical inference: the minimum distance principle approach*. New York: Chapman and Hall (CRC Press).

Berger, J. (1985). *Statistical analysis and Bayesian analysis* (2nd ed.). New York: Springer.

Berkovitz, J. (2015). The Propensity Interpretation of Probability: A Re-evaluation. *Erkenntnis*, pre-publication version available at http://link.Springer.com/article/10.1007%2Fs10670-014-9716-8.

Carnap, R. (1950). *Logical foundations of probability*. Chicago: University of Chicago Press.

Crupi, V., Chator, N., Tenori. K. (2013). New Axioms for Probability and Likelihood Measures. *British Journal for the Philosophy of Science 64*, 189–204.

Earman, J. (1992) *Bayes or bust?* Cambridge, MA: MIT Press.

Edwards, M., Lindman, H., & Savage, L. (1963). Bayesian Statistical Inferences for Psychological Research. *Psychological Review, 70*(3), 193–242.

Eells, E. (2006). Confirmation Theory. In Sarkar and Pfeiffer, (Eds.). *The philosophy of science: A-M*. London: Taylor and Francis.
Eells, E., & Fitelson, B. (2000). Measuring Confirmation and Evidence. *Journal of Philosophy, 17*, 663–672.
Fitelson, B. (1999). The Plurality of Baeysian Measures of Confirmation and the Problem of Measure Sensitivity. *Philosophy of Science, PSA, 66*(1), 362–378.
Forster, M. & E. Sober. (2004). Why Likelihood. In (Taper and Lele, 2004).
Glymour, C. (1980). *Theory and evidence*. Princeton: Princeton University Press.
Hàjek, A. (1997). 'Mises Redux'—Redux: Fifteen Arguments Against Finite Frequentism. *Erkenntnis, 45*, 209–227.
Hitchcock, C., & Sober, E. (2004). Prediction versus Accommodation and the Risk of Over-fitting. *British Journal for the Philosophy of Science, 55*, 1–34.
Hawthorne, J. (2011) Confirmation. In (Bandyopadhyay and Forster, 2011).
Kaplan, M. (1996). *Decision theory as philosophy*. Cambridge: Cambridge University Press.
Kass, R., & Rafetry, A. (1995). Bayes Factors. *Journal of the American Statistical Association, 90* (430), 773–795.
Keynes, J. M. (1921). *A treatise on probability*. London: Macmillan.
Kotzen, M. (2012). Dragging and Confirming. *Philosophical Review* 121.1, 55–93.
Kyburg, H. (1961). *Probability and the logic of rational belief*. Wesleyan, CT: Wesleyan University Press.
Lele, S. (2004). Evidence Function and the Optimality of the Law of Likelihood. In (Taper and Lele, 2004).
Levi, I. (1967). Probability Kinematics. *British Journal for the Philosophy of Science, 18*, 200–205.
Lewis, D. (1980). A Subjectivist's Guide to Objective Chance. In R. C. Jeffrey (Ed.), *Studies in inductive logic and probability* (Vol. 11). Berkeley: University of California Press.
Lindsay, B. (2004). Statistical Distances as Loss Functions in Assessing Model Adequacy. In (Taper and Lele, 2004).
Mayo, D. (1996). *Error and the growth of experimental knowledge*. Chicago: University of Chicago Press.
Mayo, D. (2014). How likelihoodists exaggerate evidence from statistical tests. https://errorstatistics.files.wordpress.com/2014/11/images2.jpeg.
Popper, K. (1959) (first published in 1938). *The logic of scientific discovery*. New York: Basic Books.
Popper, K. (1957). The Propensity Interpretation of the Calculus of Probability and the Quantum Theory. In: Körner, S., (Ed.) *Observation and interpretation: proceedings of the ninth symposium of the colston research society, university of bristol*, pp. 65–70 and 88–89.
Reichenbach, H. (1949). *The theory of probability*. Berkeley: University of California Press.
Royall, R. (1997). *Statistical evidence: a likelihood paradigm*. New York: Chapman Hall.
Royall, R. (2004). "The Likelihood Paradigm for Statistical Evidence." In (Taper and Lele, 2004).
Salmon, W. (1983). Confirmation and Evidence. In (Achinstein, 1983).
Shafer, G. (1976). *A mathematical theory of evidence*. Princeton: Princeton University Press.
Silverman, B. (1986). *Density estimation for statistics and data analysis*. London: Chapman and Hall/CRC Press.
Skyrms, B. (1986). *Choice and chance* (3rd ed.). Belmont, CA: Wadsworth Publishing.
Sober, E. (1993). Epistemology for Empiricists. In French, Uehling, & Weinstein (Eds.), *Midwest studies in philosophy* XVIII: 39–61.
Taper, M., & Lele, S. (Eds.). (2004). *The nature of scientific evidence*. Chicago: University of Chicago Press.
Taper, M., & S. Lele. (2011). Evidence, Evidence Functions, and Error-Probabilities. In (Bandyopadhyay and Forster, 2011).

Taper, M., & J. Ponciano. (2016). Evidential statistics as a statistical modern synthesis to support 21st century science. *Population Ecology, 58*, 9–29.

Taper, M., Staples, D., & Shephard, B. (2008). Model Structure Adequacy Analysis: Selecting Models on the Basis of Their Ability to Answer Scientific Questions. *Synthèse, 163*, 357–370.

Von Mises, R. (1957). *Probability, statistics, and truth, revised* (English ed.). New York: Macmillan.

Williamson, T. (2000). *Knowledge and its limits*. Oxford: Oxford University Press.

http://www.springer.com/978-3-319-27770-7

Belief, Evidence, and Uncertainty
Problems of Epistemic Inference
Bandyopadhyay, P.S.; Brittan Jr., G.; Taper, M.L.
2016, XIII, 178 p. 9 illus. in color., Softcover
ISBN: 978-3-319-27770-7