Statistical Inference and the Plethora of Probability Paradigms: A Principled Pluralism

Version 1, October 29, 2018
(uploaded to PhilArchives, October 30, 2019)

Mark L. Taper, Gordon Brittan Jr., and Prasanta S. Bandyopadhyay

Montana State University, Bozeman, MT 59717, USA

MarkLTaper@gmail.com
GBrittan17@gmail.com
PSB@montana.edu

Abstract:

The major competing statistical paradigms share a common remarkable but unremarked thread: in many of their inferential applications, different probability interpretations are combined. How this plays out in different theories of inference depends on the type of question asked. We distinguish four question types: confirmation, evidence, decision, and prediction. We show that Bayesian confirmation theory mixes what are intuitively "subjective" and "objective" interpretations of probability, whereas the likelihood-based account of evidence melds three conceptions of what constitutes an "objective" probability.

Keywords: Subjective probability; Objective probability; Propensity; Evidence; Confirmation; Bayesian decision theory, Evidential decision theory; Prediction; Four epistemic questions; Belief question; Evidence question; Decision question; Prediction question; Three types of objective probability; Combining different types of probabilities

Acknowledgments:

These ideas have their origin at a conference funded by a Japanese Data Centric Science Research Commons, Research Organization of Information and Systems Grant to Dr. Yamashita and Japan Society for the Promotion of Science KAKENHI Grant Number 25340110 to Dr. K. Shimatani. We would also like to thank José M. Ponciano, Brian Dennis, and Subhash R. Lele for discussions on these issues.

Statistical Inference and the Plethora of Probability Paradigms: A Principled Pluralism

> "[T]he house of statistics is divided according to the interpretation of probability that is adopted. We need to settle on *an interpretation* of probability in order to settle on a global approach to statistical inference. …Given the approach to probability that we settle on, we shall see that each of these approaches to statistics can find its useful place in our inferential armament." (Henry Kyburg, Jr., and Choh Man Teng, 2001. *Uncertain Inference*. Cambridge: Cambridge University Press. p.196, emphasis added)

1 Introduction

1.1 Notions and notations for subjective and objective probability

Many statisticians and philosophers of science are pragmatic pluralists when it comes to statistical inference. That is, they think that different problems require now one, now another statistical paradigm for their resolution. Some problems lend themselves to a Bayesian approach, others to an error-statistical approach, still others to an evidentialist approach and its likelihoodist special case. Ostensibly, each approach has embedded within it a unique interpretation of the probability operator, some "subjective," others "objective."

In this paper we argue a two-fold thesis: that there are "subjective" and "objective" elements within some individual paradigms and that the "objective" elements themselves often meld distinguishable conceptions of *objectivity.*[1] In order to avoid equivocation, a notational

[1] Others have made something like this general point. Ian Hacking (2001), for example, distinguishes (p. 140) between "frequency dogmatists" and "belief dogmatists" and, rejecting the adoption of one position or the other, urges an "eclectic" approach that selects "what seems best from various styles, doctrines, ideas, and methods." His point differs from ours, however, in that it (a) resembles the pluralistic pragmatism mentioned above, and (b) is not

distinction among probability operators should be made. Since we restrict our attention here to two statistical paradigms, Bayesian and Likehoodist, we distinguish between two such operators, $\mathbf{Pr_c}$, for credal probabilities (which for Bayesians measure or represent degrees of belief for some agent that a proposition is true), and $\mathbf{Pr_i}$, for implicit probabilities which represent a formal "deductive" relation, either analytic or algorithmic, between a model and a potential data set[2]. By a "model," we mean a mathematical entity that specifies a probability distribution over a range of data which might be collected.[3] Models need to be distinguished from hypotheses which, unlike models, are statements that do not necessarily contain a quantitative assertion about how probable data are under them.[4]

For the sake of simplicity, we contrast paradigmatic "credel" and "non-credel" probabilities. It follows from our characterization that credal probabilities are temporally-

---

argued for or developed in any detail. Similarly, Donald Gillies (2003) distinguishes between subjective and objective interpretations and argues that both are needed, the former in the social, the latter in the natural sciences. But his position (a) again resembles "pluralistic pragmatism" in the sense that either subjectivity or objectivity is a feature of entire statistical paradigms and are neither mixed nor melded within them, (b) argues implausibly that there is something inherently "subjective" about the social sciences, "objective" about the natural sciences. Although we won't support the claim here, no such (principled) distinction between sciences is plausible.

[2] The notion of "Implicit probability" is related to but not simply a special instance of what Hájek 2012) calls "Quasi-logical probability," a generalization of the intuitional core of Carnap's (1950) logical theory. Both construe "probability" in formal and non-credal terms.. But "implicit probability" in our sense has a narrower definition because it is restricted to probabilities deriving from models and does not include probabilities that flow logically from hypotheses. More important, Hájek's quasi-logical concept "is meant to measure objective evidential support relations" between data and hypotheses considered singly. In contrast, for us evidential support relations are comparative measures among models/hypotheses. An example is the ratio of likelihoods between hypotheses considered pairwise. Still more important, Hájek's notion (like that of Carnap's before him) fuses "evidence" and "confirmation" (e.g., "the logical theory seeks to encapsulate in full generality the degree of support or confirmation that a piece of evidence confers upon a given hypothesis *H*…"), whereas it is one of our main aims to distinguish sharply between them.

[3] It is unfortunate that this use of the word "model," common among statisticians, has little in common with the same term used by philosophers and logicians to designate a domain of individuals and an assignment function such that all of the sentences in a specified vocabulary and formulated in a canonical first-order way are true. On the latter usage, truth is relative to a model and a variable assignment, on the former it is not. What the two usages share is that the relationship is model-relative. This is the sole respect in which models on either usage are in any sense "subjective." In the case of statistical inference, both data and the models generating them must be *described* in a particular way.

[4] As a didactic example, an economic model might say that the price of apples is linearly related (with specified intercept, slope and error variance) to the supply of apples, while an economic hypothesis might only state that the price of apples decreases with their supply.

indexed; *a* believes that *p* is true to some degree *r* at time *t*.[5] Non-credal probabilities may attach to data and models that contain temporal variables, but the probabilities themselves are not similarly indexed. The relation between models and data is both temporally-invariant and agent-independent.

Nothing in our discussion implies that there are no more than two viable statistical paradigms or interpretations of the probability operator, or that there are not other and more subtle ways in which to characterize "credel" and "non-credel."[6] Our argument rather is that a distinction must be made between probability types even as the solution of certain types of statistical questions requires their joint invocation.[7]

1.2 Fundamental Methodological Questions

We begin with two pre-theoretical and fundamental methodological questions, (i) the confirmation question (CQ), and (ii) the evidence question (EQ).[8] Contending parties in what are sometimes termed "the statistics wars" usually agree on the centrality and significance of these questions, but disagree about the extent to which they are distinct and how best handle to them. More specifically, the questions are:

(i) (CQ) Given the data, what should we *believe*, and to what degree?

(ii) (EQ) Do the data provide *evidence* for model *M1* against an alternative model *M2*, and how much?

To better appreciate the rich tapestry of current statistical methodology, we will raise two more questions, (iii) the decision question (DQ), and (iv) the prediction question (PQ). They are:

(iii) (DQ) Given the data what should we *do*?

---

[5] In his seminal (1931), Ramsey draws attention to the necessity of time-stamping belief-probabilities so as to allow for the consistency of the kinematic changes they undergo as agents accumulate new data. Although we have not complicated the notation to indicate it, credal probability operators in what follows are assumed to have a temporal subscript.

[6] Use of jargon here should help to pare back at least some of the misleading connotations clinging to the more conventional vocabulary of 'subjective," "objective," etc., which is misleading at best. It is one of our main aims to replace this simple-minded way of talking with something more granular and perspicacious.

[7] Kyburg and Meng (2001, p. 196) say rightly that "the house of statistics is divided according to the interpretation of probability that is adopted." But they give no reason for claiming in the next sentence that "We need to settle on an interpretation of probability in order to settle on a global approach to statistical inference." Granted that a house divided cannot stand, but there need be no civil war among its inhabitants if each has a role to play distinct from and in harmony with the others' roles.

[8] Following Royall (1997).

(iv) (PQ) Given the data, what is the evidence for the greater *predictive accuracy* of one model against its alternative?

1.3 Outline

In section 2.1, we discuss both subjective Bayesian (hereafter Bayesian) confirmation theory and the likelihoodist account of evidence to show that subjective Bayesians are well-equipped to address CQ, and likelihoodists like Royall are equally well-equipped to address EQ. In section 2.2, we will discuss how a mixing of probability interpretations occurs in the Bayesian account, a melding of them in the Likelihoodist.[9]

In section 2.3, we show how Bayesians are able to address DQ. In the same section, we discuss how an evidentialist would respond to the DQ. In section 2.4, we broach the evidentialist's exploitation of the Akaike Information Criterion (AIC) to address PQ. We also examine issues concerning the consistency of AIC and its connection to an evidential treatment of PQ. We point out that the term "consistency" applies differently to the EQ and PQ questions. To defend our contention that major applications of statistical inference make use of different types of probabilities, in section 2.5 we show how those two accounts - Bayesian and evidential - also involve fusions of probability interpretations. In section 3, we introduce several corollaries that flow from our theses. One is an unconventional epistemic stance on the interpretational issue: we hold that even with regard to a single type of question, there has to be a mixing or melding of different conceptions of probability. Another corollary is that our pluralism is principled; it has to do with the type of statistical question being asked and not with a particular problem to be solved.

[9] Although we are going to argue that both statistical paradigms combine different types or aspects of probability we will see that the manner in which each paradigm combines them is slightly different from the other. To distinguish the Bayesian mixing of probabilities from the Likelihoodist (or Evidentialist), we use "melding" for the latter, keeping in mind that often ordinary language is not adequately subtle to keep them separate. One purpose of the paper is to make the distinction between them clearer.

2 Distinctions and Concordances

2.1 Bayesian Account of Confirmation and Likelihood Account of Evidence[10]

The account of confirmation we take as paradigm involves a relation, *C*(*D, M, B)* among data *D*, model *M,* and the agent's background information *B*.[11] Here, "data" are taken to be sets of observations collected by a defined procedure and described in a uniform way. This account of confirmation incorporates the basic rules of probability theory, including the rule of conditional probability, an interpretation of the probability operator as a measure of personal belief, and some reasonable constraints on one's *a priori* degree of belief that whatever empirical model is under consideration is true. We assume that 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 model is true to degree 1, i.e., could not possibly be false, or false to degree 0, i.e., that it could not possibly be true. Empirical models are never more than *merely* probable. This said, an agent learns from experience by up-dating her degrees of belief that a model is true by conditioning on the data as she gathers them, i.e., in accord with the principle, derivable from the axioms of probability theory and the definition of conditional probability, that $\Pr_c(M \mid D) = \Pr_c(M)\Pr_c(M \,\&\, D)/\Pr_c(D)$. Assuming that $\Pr_c(D) \neq 0$, her degree of belief in *M* after the data are known is given by $\Pr_c(M \mid D)$, called the posterior probability of the model, and calculated using Bayes Theorem:

$$\Pr_c(M \mid D) = \frac{\Pr_c(M) \times \Pr_i(D \mid M)}{\Pr_c(D)}$$

The prior probability, $\Pr_c(M)$, of a model depends on the agent's degree of belief that the model is true before new or additional data bearing on the model have been gathered. The marginal probability, $\Pr_c(D)$, represents the probability that *D* would obtain, averaged over all the models being considered. That is $\Pr_c(D) = \sum_i \Pr_c(M_i)\Pr_i(D \mid M_i)$. We call the following *the Confirmation Account*:

[10] "Confirmation" and "evidence" are often used interchangeably in discussions of the ways in which data "support" hypotheses (see footnote 21). It is a main aim of our monograph (xxx) to show how and why they must be distinguished. It is not to our present purpose to repeat that demonstration here. We have been asked whether a distinction of this kind reforms statistical practice. It is already beginning to do so. See (Klement, 2017) and, for an application of an element of our argument (Pulgarin, et. al., 2018).

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

D confirm *M if* and only if $\Pr_c(M \mid D) > \Pr_c(M)$.[12]

This is a qualitative comparison of the probabilities of a model before and after data have been collected at any given time. The bi-conditional forms the intuitional basis of the conception of confirmation. The account rests, as do most Bayesian conceptions of confirmation as a relation between probability measures, on the following principle: for any M, D_1, D_2, the confirmation (disconfirmation) of *M* in the light of D_1 is greater (less) than the confirmation(disconfirmation) of *M* in the light of D_2 just in case $\Pr_c(M \mid D_1) > (<) \Pr_c(M \mid D_2)$.[13] The confirmation account can with ease be made quantitative. There are alternative ways to measure the degree to which a model is confirmed, but a common metric is in terms of the difference between the model's prior and posterior probabilities.[14] It is customarily normative as well; one *should* believe that a model is true to the degree that it has been confirmed.

Evidence, construed as an answer to the second question (EQ), involves a comparison of the merits of two models, M_1, and M_2 (possibly, but not necessarily, $\sim M_1$) relative to the data *D* and background information *B*.[15]

D is evidence for M_1 & *B* as against M_2 & *B* if and only if $\Pr_i(D \mid M_1 \& B) > \Pr_i(D \mid M_2 \& B)$

[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, and hence of the indispensability of prior probabilities, in an analysis of confirmation.
[13]Following (Crupi, Chater, and Tenori, 2013) whose article includes a long list of Bayesians who subscribe to this principle.
[14] Eells (1982), Gillies (1986), Earman (1992), Jeffrey (1992), and Rosenkrantz (1994) are among those who defend the "difference measure" of "evidential support" (identified by us as a measure of "confirmation," not of "evidence"). Another standard measure is the posterior/prior ratio. See footnote 18 on the frequent but misguided claim that this PPR measure is equivalent to the LR measure of evidence.
[15] We will only include *B* in our initial formulation and subsume it into the model in our later discussion. It is important to note that "*~M*" is ambiguous as between a specific alternative to the model of interest and a set of models ($M_1, \ldots, M_i, \ldots M_z$). Deborah Mayo (Mayo and Spanos 2006) and other error-statisticians and the subjective Bayesian Colin Howson (2013, 2016) allege that likelihood accounts cannot account for composite hypotheses. According to Howson in particular, an account such as ours, which maintains that the values of likelihoods and hence of their ratios do not depend on a consideration of prior probabilities, is mistaken since we cannot calculate the $\Pr(D \mid M)$ without knowing the priors of each of the component models. The point is easily deflected. For evidential purposes it is effective to define the generalized likelihood $L_g(D; \{M_1, \ldots, M_i, \ldots, M_z\})$ as $\max(L(D; M_i))$. Using ";" rather than "$\mid$" to express the relationship between data and model helps avoid the tempting but mistaken equation of likelihoods to $[\Pr(M \mid D)\Pr(M)]/\Pr(D)]$ and a misreading of $P(M)$ in belief-probability terms. The use of generalized likelihoods to test composite hypotheses goes back to the early stages of modern frequentist statistics (Neyman and Pearson, 1933). More discussion of the use of generalized evidence functions for inference for composite models can be found in Taper and Lele (2011), Bickel (2012), and our Monograph (2016)

Call this the Likelihoodist's *Evidence* Account.[16] It serves to characterize data which, in context, play an evidential role.

The quantity $Pr_i(D \mid M)$ is often referred to in the *philosophical* literature as the "likelihood." While numerically the likelihood of the model given the data is proportional to the probability of the data given the model, likelihood and probability are not the same thing; likelihood is considered a function of the model, whereas the probability is considered a function of the data. We here adopt the common philosophical notation of denoting the likelihood by $Pr_i(D \mid M)$ rather than the common statistical notation of $L(M; D)$,[17] but do not mean to imply that the model M needs to be considered a random variable. Because credal probabilities are not involved in the calculation, the likelihood is objective.

As with the confirmation account, this statement is initially qualitative, but a commonly used quantitative measure of the degree of evidence vis-a-vis one model over another is the numerical ratio of the likelihoods[18] for the two models. Note in this connection that if $1 < LR \leq 8$, then D is often said to provide "weak" evidence for M_1 against M_2, while when $LR > 8$, D provides "strong" evidence for M_1 over M_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

[16] Anthony Edwards (1972) was the first statistician to make explicit the connection between the likelihood ratio and what he calls "degree of support" and we, following Royall (1997), more narrowly call "evidence."The LR is a special case of "evidence functions" useful for specified as opposed to estimated models Lele (2004). For simplicity, we frame our philosophical arguments using the LR as an exemplar of this more general class of statistical functions. It is important to note that we differ sharply from Edwards as concerns the range of what he calls the "Likelihood Axiom." In his words (1972, p. 31, emphasis in original): "Within the framework of a statistical model, *all* the information which the data provide concerning the relative merits of two hypotheses is contained in the likelihood ratio of those hypotheses on the data, and the likelihood ratio is to be interpreted as the degree to which the data support one hypothesis against the other." We disagree that "all" of the information which the data provide is contained in the likelihood ratio (which is to say that the notion of "support" is ambiguous). Rather, we distinguish "confirmation" from "evidence" in the attempt to show that both concepts must be incorporated in an adequate account of statistical inference/hypothesis testing. I.J. Good (1983) uses the likelihood ratio in the way we do to characterize what he calls "the weight of evidence."

[17] 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 $Pr(D \mid M)$ the model is a random variable, while $Pr(D; M)$ indicates that the probability of the data is conditioned on a model that is considered fixed.

[18] Or the logarithm of the likelihood ratio. See Eells and Fitelson (2002, p. 9, n. 10) for a reason to use logged versions of the quantitative measures of what we term "confirmation" and "evidence." Nothing in our discussion turns on it. The respective ordinal structures remain the same.

among investigators.[19] It follows from the Likelihoodist account of evidence that the range of values for the LR can vary from 0 to ∞ inclusive

On this account of evidence, a model is defined as an idealized data-generating mechanism. Observed data support $model_1$ over $model_2$ if the potential data generated by $model_1$ are by some measure closer to the observed data than the data generated by $model_2$. Thus on this account, data provide evidence for one model against its alternative *independent of what* the agent *knows or believes* about either the available data or the models being tested.

Based on our discussion so far, there are two ways to distinguish "confirmation" from "evidence."[20] First, confirmation is *psychological, involving as it does changes in an agent's personal degree of belief that a model is true, and evidence is "logical," an agent-independent relation between models and data.* It follows that confirmation is also "kinematic," beliefs re-adjusted over time as data are accumulated, evidence is "static," a timeless relation between models and data (although the evidence will change if different data are considered). Second, confirmation on the standard Bayesian account is normative in that an agent's belief adjustments should cohere with the axioms of probability theory and their implications, evidence on the likelihoodist account is merely descriptive of relations obtaining between models and data. One

---

[19] 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.

[20] Ellery Eells and Brandon Fitelson (2002) distinguish these and other similar measures, but assume that they all measure the same thing, viz., the degree of "evidential support" which data confer on hypotheses. From our point of view, three features of their discussion are note-worthy. First, our "confirmation" and "evidence" measures score equally well with respect to the criteria of adequacy they propose and better than any of the other traditional measures, e.g. Keynes' (1921), Carnap's (1962, #67) and Christensen's (1999), considered. Second, although Eells and Fitelson describe all of the measures evaluated as "non-equivalent," they don't specify in what respects. Some have argued that Keynes' posterior-prior ratio measure (PPR) is equal to the LR measure and therefore precludes the necessity of a separate account of evidence (among the most recent, Evans, 2016). But the claim is misguided. The LR is equal to ($Pr_c(M_1 \mid D)/Pr(M)$ *only when* $Pr(D)/Pr(D \mid {\sim}M)$ is close to 1. That is, $[Pr(D \mid M)/Pr(D \mid {\sim}M)] = Pr(M \mid D)/Pr(M) \times [Pr(D) \times Pr(D / {\sim}M)] \approx 1$. Otherwise, the two measures, LR and PPR, yield different values, over and above the fact that in our view they differ conceptually. Others (e.g., Howson, 2016) note that the two measures are equivalent in the sense that whenever data confirm a hypothesis on our account they at the same time provide evidence for it. It is indeed easy to show that $Pr(H \mid D) > Pr(H)$ iff $Pr(D \mid H)/Pr(D \mid H') > 1$, but this equivalence holds only of the qualitative measures of confirmation and evidence, is true only on the condition that *H* and *H'* are mutually exclusive and jointly exhaustive, and, most important, is not linear, i.e., in individual cases the difference between posterior and prior probabilities may be small and the likelihood ratio large and conversely (one reason for distinguishing them). Third, Eells and Fitelson note in passing (2001, pp. 130-131) that "we define and label the pertinent measures of evidential support … where 'Pr' denotes probability, on some appropriate interpretation, usually a 'subjective' or 'logical' interpretation." I.e., they apparently assume that whatever the measure at stake, the probability operators within it are to be given a univocal interpretation. On our argument, this is not the case, and the various interpretations within particular measures should, be distinguished conceptually and identified notationally to avoid confusion and to make the meanings of "subjective" and "objective" more nuanced as the statistics wars grind on,.

of the consequences of the distinction between confirmation and evidence is that in the case of confirmation, there is room for rational disagreement between two agents to the extent that their priors differ, whereas since the likelihood function is based on the mathematical calculation of a model, there is no place for any such rational disagreement in the case of evidence.[21]

2.2 Two Accounts and Probability Interpretation Complexity

We will argue that adequate accounts of confirmation and evidence presuppose different concepts of "probability" and therefore different readings of the probability operator in our various conditions and equations. Here probability operators apply to "events," "models," and "propositions." Our view is that one needs to *mix* or *meld* probabilities of different kinds (where the different usages "mix" or "meld" single out which account we address), and not rely on one to the exclusion of the others, even within statistical paradigms. As we will now show, to do otherwise would be to seek an overly monistic methodology.

2.2.1 Subjective confirmation

Consider first the Bayesian account of confirmation. It is informed by a subjective or psychological understanding of what a "probability" measures. *D* confirm *M* just in case an agent's prior degree of *belief* that *M* is true is raised. The degree to which *D* confirm *M* is measured by the extent to which the degree of belief has been raised. The probabilities involved have to do with beliefs, and are in this sense "subjective." But determining an agent's degree of belief that a model is true given the data, $\Pr_c(M \mid D,)$ requires determining the likelihood of *M* on *D*, L(*M;D)* is proportional to $\Pr_i(D|M)$, which is independent of an agent's belief and therefore objective.

One possible response to this claim is that the likelihood function is mistakenly construed as objective; it needs to be construed as a conditional belief-probability in which the probability operator is still to be understood in *credal* terms. It is of course possible to do so.

---

[21] Although our argument for making the distinction has to do with the clarification it introduces into discussions of statistical inference and not with the ways in which the terms "confirmation" and "evidence" are commonly used, one can detect an intuitional bias for it in the literature. Thus John Vickers (1988, p. 81, our emphasis): "As far as *evidence* is concerned, for logicism the support of probability judgments is logical and definitional. Frequentism is explicit that probability statements are *confirmed* by statements of finite relative frequencies." Although at other places (e.g., p. 93) Vickers blurs the confirmation/evidence distinction, and the contrast he draws here between "logicism" and "frequentism" is not quite ours, "evidence" in this passage is a formal or logical relation, "confirmation" not. See Eells (1985) and Fitelson (2000) for discussions of the "confirmation"/"evidence" terminology as it is used in the literature. Like almost all philosophers and Bayesian statisticians, Eells and Fitelson themselves use the terms interchangeably (2002, p. 129, n. 1).

David Lewis has proposed a well-known way. He suggests that when the chance probability of a proposition, A, is externally and reliably available, one needs to set an agent's subjective probabilities, $Pr_c(A|M) = Pr_i(A|M)$, where *M* is a statistical model (Lewis,1980). Lewis thinks that this alignment of subjective probability with objective chance is possible if we adopt the *Principal Principle* (PP). The idea behind this principle is that when we have the likelihood of a model[22] given the data available we should set an agent's subjective probability of the data given the model equal to its objective likelihood. This alignment of the subjective probability with the objective chance[23] allows us to treat both probabilities - likelihood functions and prior probabilities - in the application of the Bayes theorem as credal (for a justification of the PP, see Pettigrew, 2016).[24]

In practice but often implicitly, Bayesians typically presuppose the Principal Principle and align objective likelihoods and subjective probabilities in this way.[25] The apparent aim is to avoid equivocation, i.e., to interpret the probability operators in Bayes Theorem in a consistently personal-belief way, in particular to treat the posterior probability in it as credal. But no equivocation results if (a) credal and non-credal probability operators are clearly distinguished, (b) it is assumed that subjectivity spreads through a compound probability statement like falsehood through conjunctions and truth through disjunctions. Thus we write Bayes Theorem as $Pr_c(M \mid D) = Pr_c(M)Pr_i(D \mid M)/Pr_c(D)$ and understand that although the calculation of the posterior probability incorporates an objective likelihood, nevertheless it results in the adjustment of a personal-belief probability. Moreover, this sort of adjustment of our personal beliefs, quite apart from any commitment to a statistical paradigm, seems to proceed in just this way, by mixing

[22] Our use of the term "model" aligns with Lewis' use of the term "proposition."

[23] Which is possible only by way of the stipulation that "credence [be] conditional on the truth," i.e., on all of the "evidence" that has been collected. This stipulation, characteristic of what might be called "strong objective Bayesianism," is critiqued in our paper (2010). From the perspective of our present argument, the strong objective Bayesian conflates agent-invariance (everyone with the same prior) and agent-independence (likelihoods are belief-free). In the penultimate section of this paper, we take up questions concerning agent-invariance and "Bayesian convergence" in more detail.

[24] There is a long tradition in statistics and philosophy in which even those who are considered non-Bayesians like Fisher, Efron, Kyburg, and Sober *always* recommend the use of Bayes theorem when prior probabilities are frequency-based. However, they are critical of the use of the theorem when prior probabilities are subjectively grounded. So, when both probabilities (prior probabilities and the likelihood function) are objectively given, then, according to the *Principal Principle*, the agent should adjust her subjective probabilities according to the available objective probabilities.

[25] Blurring the distinction between confirmation and evidence in the process. See Lewis (1980, 285): "As new evidence arrives, our credence in [particular or general propositions about chances] should wax and wane in accordance with Bayesian confirmation theory."

together data, models, and logical relations between them. At least, the adjustment of beliefs required by Bayesians would seem to be "rational" if we do so.

Of course, it is possible to take the model itself to be about credences. However, since the values of the probabilities are specified in the model, there will be no disagreement about the likelihoods when people have different beliefs, because although the prior and posterior probability operators embedded in Bayes Theorem are given a consistently credal reading, the likelihood component of that Theorem will be in a straightforward sense "agent-independent"[26].

In order to make the case that credal and non-credal interpretations are mixed in the Bayesian account of confirmation we have taken as paradigm, we must expand our characterization of objectivity as "agent-independent". Doing so requires taking a closer look at the likelihood account of evidence and has as a corollary a further distinction between senses of "objectivity." In the process it will allow us to clarify different but equally indispensable conceptions of "non-credal" probability.

We have already argued that the likelihood account of evidence makes no use of belief-probabilities and in this sense is thoroughly objective. But more must be said to indicate why this claim is something more than an *obitur dictum.* The likelihood of a model given the data is an essential component of the account. For simplicity's sake, we identify it as proportional to the conditional probability, $Pr_i(D|M)$, where "D" stands for a random variable that can take on any value that can potentially be generated by model M. What data could be realized *a priori* depends on what sorts of models we have proposed. The relationship between a model and its unrealized data is deductive in the sense that M entails the distribution of "D" for a given specification of "M", and "D". 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 a probability of 0.7 to land heads up. Assume further that M_2 says that it has probability 0.4, M_3 says that it has probability

[26] We are hardly the first to point out that the likelihoods in Bayes Theorem constitute an objective component in it. Isaac Levi (1967), for example makes just this point, that likelihoods are fixed and agent-invariant, and in this sense "objective," and unlike variable conditional probabilities that are agent-dependent. But for him and other Bayesians, likelihood probabilities are nonetheless credal. For another example of trying to have it both ways, see James Hawthorne (1994). In his term of art, likelihoods are "*fairly* objective," (our emphasis) but as he goes on to make clear (p. 241), "they are the means by which evidence affects inductive support." In other words, they provide the inferential tie between data and model in just the same way that *logic* generally considered provides the inferential tie between premises and conclusions quite independent of what anyone knows or believes. On our view, the Bayesian attempt to "have it both ways," i.e., to claim that likelihoods are "objective" but nonetheless credal depends on conflating agent-independence with agent-invariance. We return to this mistaken conflation in the section on "Further Corollaries" below.

0.6, M_4 says that it has probability 0.5, and so on. Given these models, before any actual 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 is completely deductive. So, too, are the relationships between other competing models and D.

Now assume that a real coin has been flipped. Further assume that out of 100 flips, 58 of them have landed with heads and the rest are tails. Let's call the data set d_1 58 heads and 42 tails. Even though d_1 is a real world data set, the relationships expressed via $Pr_i(D=d_1|M_1)$, …, $Pr_i(D=d_1|M_k)$ continue to be deductive. The expressions $Pr_i(D=d_1|M_j)$ simply ask what the probability deductively calculated from model M_j would be if D had the value d_1. 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. For the four models we have explicitly listed, we have $Pr_i(D=d_1|M_1)=0.0032$, $Pr_i(D=d_1|M_2)=0.00011$, $Pr_i(D=d_1|M_3)=0.074$, and $Pr_i(D=d_1|M_4)=0.022$. M_3 has the highest likelihood and is thus best supported by the data gathered. Yet the LR of M_3 to M_4 of 3.32 only indicates weak evidence for M_3 over M_4 after the coin has been flipped. But, regardless of any observed realization, the relationships between each possible model and possible outcomes are themselves deductive. Given any model, in this framework there is always a derivable countable frequency-based probability, Pr_i, (Hájek, 2012) connecting the possible outcomes to the model. That is to say, the probability of outcomes are the proportion of times each possible outcome would occur if the random processes of the model were to be repeated an infinite number of times.

2.2.2 Propensity as an Objective Probability

In the real world, there is a tendency on the part of objects in standard conditions to behave in routine ways (Peirce, 1878; Popper, 1957, 1959; Hájek, 2012). This natural tendency of objects and conditions is called their *propensity*, and has long been characterized as a probability. Notationally, we designate this form of probability as Pr_n. The propensity of a coin to land heads or tails in a particular flipping experiment is exhibited in a sequence of flips; out of 100 flips, 58 come up heads, for example, and 42 tails. We have no way to know for sure the true nature of the coin, but we can estimate it via a finite empirical frequency (Hàjek, 2012). This is yet another conception of probability, which we designate Pr_f. Empirical finite frequencies can be thought of as estimates of natural propensities. For our purposes in this paper, propensity and frequency conceptions are, like the conception of implicit probability, brought together under the heading of "non-credal probability."

It is natural to assume that the tendency of a model to generate a particular sort or set of data *represents* a causal tendency on the part of natural objects represented in the model to have particular properties or behavioral patterns. This tendency or "causal power" can be both represented and sometimes explained by a corresponding model. In our view, a fully objective account of evidence requires that we must make this realist assumption and thus take model probabilities as modeled propensities.

Very briefly, then, evidence is "fully objective" not only insofar as its characterization is not in any way a function of prior (personal) probabilities, for as noted the likelihoods are already fixed by the model and thus agent-independent, but also to the extent that it provides us with information about objects in the world. We might put this by way of a formula: objectivity requires reality as well as independence. Reality, in this case, amounts to propensities on the part of natural objects to behave in particular ways in certain circumstances, quite apart from what we might believe about them. The aim of science is to identify and where possible to explain them. Again very briefly, fully objective probabilities are thus modeled propensities.[27]

According to the propensity interpretation of probability,[28] the probability of an event is a dispositional property of the process that generates that event.[29] In this case, the probability of a coin's landing heads estimated from a sequence of repeated flips is 0.58. It is then reasonable to assume that the true propensity is something near 0.58. The only way we have of knowing what the propensity (or tendency) of the coin is to estimate it through finite empirical frequencies.

[27] A more extended version of this argument can be found in (xxx, 1994).

[28] See (Peirce, 1878), (Popper, 1957, 1959), (Suppes, 1987), and (Gillies, 2000). 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.

[29] It is important to realize that while the propensity is a property of the generating conditions, the generating conditions are not a property of the propensity. Different generating conditions may induce the same propensity. As we pointed out in footnote 17, there are two kinds of conditioning that occur in statistics and stochastic processes. One can condition on a random variable, or one can condition on a fixed effect. Conditioning on a random variable (B) restricts to a subset the sample space that the probability of the event A is being calculated over. Likewise conditioning on A restricts to a subset the sample space that the probability of B is being calculated over. However, if one conditions on a fixed effect B, the value of random variable A has no effect on the fixed effect because it is, well, fixed. The rules of probability that Humphreys (1985) found violated involve conditioning on random variables. Considering propensity as a random distribution conditioned on fixed effects (i.e. generating conditions) avoids the Humphreys paradox.

2.2.3 Finite Empirical Frequencies

The empirical interpretations of probability championed by von Mises[30] and Reichenbach[31] define probability in terms of the limits of relative frequencies in hypothetical infinite sequences of trials; 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 issues 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 implausible *tout court*. 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. Rather, finite frequency is no more than a way of estimating propensities in particular cases. Troublesome cases, such as assigning probabilities to single events, do not undermine an interpretation of "probability" so much as they raise difficulties for approximating values under certain conditions.

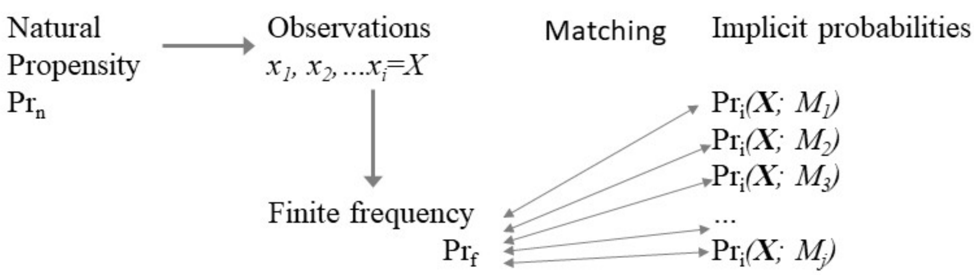

Figure 1: Schematic representation for the relationships among kinds of probabilities in a statistical inference. In subjective Bayesian statistics, matching is measured by degrees of belief. In likelihood or evidential statistics, matching is measured through evidence.

2.2.4 Combining Probabilities

In other words, and in our view, it is not misleading to say that there are three distinct conceptions of objective probability – implicit, propensity, and finite frequency – although it would be more to our purposes to say that these are "melded" aspects of a single interpretation of objective or non-credal probability. *Likelihoods allow us to identify models whose implicit*

[30] See (von Mises, 1957)
[31] See (Reichenbach, 1949).

probabilities best match natural propensities as approximated by finite frequencies. Figure 1 shows a schematic of the interplay of multiple features of probability in using statistics to make a scientific inference. We estimate the natural propensity expressed under specified conditions via the relative frequency of observations of various type under those conditions. We choose the model (theory) that best matches that finite frequency. This is the *only* way to make sense of the likelihood-based inference if a likelihoodist wants to connect her paradigm to empirical and real-world applications.

Bayesians select models with the highest posterior probability. Likelihoodists choose models by matching their implicit distributions with realized distributions; the model with the best match, i.e., maximum likelihood, is chosen. As the posterior probability is proportional to the product of the likelihood and the prior probability, one could say that the likelihood is being biased towards models of high prior probability.[32] Alternatively, the posterior probability can be seen as a weighted distribution with the likelihood the weight function for prior distribution (Patil et al. 1986). In this conceptualization, the prior is being biased towards models with high likelihood support. Thus, under either conceptualization of the Bayesian approach, models are chosen on the basis of matching both to empirical frequencies and to prior probabilities. In contrast, in the likelihoodist/evidentialist approach models are chosen only on the basis of matching to empirical frequencies.

We think that the melded probabilities invoked by likelihoodists and other evidentialists provide a case for their measure(s) being agent-independent in a realist as well as an invariant way, and thus "fully objective" (insofar as they are also taken as modeled propensities). Hence they are distinguished from subjective or credal belief-probabilities. Confirmation is in a fundamental sense "in the head," evidence is in that same sense "in the world." The fundamentally multiple interpretations of "probability" in credal and non-credal cases make this sense clearer.

[32] Since the Bayesian will always choose the model with the higher posterior, if two models have the same likelihood, the Bayesian will choose the one with the higher prior. The primary implication would seem to be that model choice is no longer evidential, but belief-dependent. Often the difference will be small, but see Howson (2013, 2016) for an extended argument in behalf of the claim that in the case of certain "gerrymandered" hypotheses (e.g., that Santa Claus wills a sequence of 100 coin-flips that in every case come up heads), it will and should be very large and our reply to same (2016).

2.3 Bayesian and Evidential Accounts' of the Decision Question

For subjective Bayesians the answer to the decision question, "given the data what should we do?" flows naturally from their answer to the confirmation question. Subjective Bayesians answer the question by invoking the idea of maximizing expected utility which accommodates both the posterior probability of states of nature given by Pr_{ci} (i = 1,2) and the utility of having some "wealth" in each state W_i (i = 1,2) given by $U_j = U_j(W)$ of doing an action j. Assuming that there are only two mutually exclusive actions, A1, and A2, we apply the expected utility formula $E(Uj) = (1-Pr_{c1})\, U_j(W1) + Pr_{c2}\, U_j(W2)$ for each action and then choose the action which has the maximum expected utility. For a Bayesian, the expected utility of an act depends on both her subjective posterior probability for that act and the utility of opting for that act. In short, the Bayesian response to the decision question is that for any set of options one should always choose the act that maximizes her expected utility, when the expectation is taken over the posterior distribution of her belief after all data have been taken into account.

An evidential account of the decision question can be developed paralleling the Bayesian approach. The distinction between the two concerns the probability distribution over which the expectation of utility is calculated. An evidential calculation identifies which state of potential nature is best supported by the data. Associated with each evidential assessment is a probability (M_L) that the evidence has been misleading.[33] M_L is called the probability of misleading evidence because it is evidence for the false hypothesis. It is not a credal probability, but is a quantity derived from the data-generating properties of the models involved[34]. The utility expectation is taken over the distribution of probabilities that the states are correctly identified.

Revisiting the two state-of-nature cases mentioned above in the discussion of Bayesian decisions, let us index the states so that state best supported by data has index 1. The evidential expected utility of an action j is given as: $E(Uj) = (1- M_L)\, U_j(W1) + M_L\, U_j(W2)$. To reiterate,

[33] Note that M_L is the local or post-data probability of misleading evidence discussed by Taper and Lele (2011), not the global or pre-data probability of misleading evidence of Royall (1997).

[34] M_L might also be estimated as a finite frequency through a statistical non-parametric bootstrap of the model identification process. The usefulness of the bootstrap for M_L estimation, particularly in the presence of model misspecification, is developed in another publication currently under review.

the distinction here is that this evidential decision weights utility with a non-credal probability as opposed to the Bayesian decision where the weighting is credal.[35]

2.4 An Evidential Account of the Prediction Question

It has been claimed (e.g., Taper & Ponciano 2016) that the PQ is not an evidential question *sensu stricto*. Here we argue that the PQ can be seen as a species of the evidence question. To continue the discussion, we need to be precise as to what "predictive accuracy" means.[36] The model with the lowest mean-squared error of prediction (MSPE), say as measured by cross-validation, is commonly considered the most predictively accurate. For models with normally distributed errors, 1/MSPE can be readily seen as a biased estimate of the expected log-likelihood (ELL) of a predicted observation. Thus the expected log-likelihood can be viewed as a generalization of the MSPE as a measure of predictive accuracy. The model with the greatest expected log-likelihood is considered the most accurate. The observed log-likelihood is a biased estimate of n*ELL, with the amount of bias depending on the sample size (n), the number of parameters estimated (k), and the true process generating the data.

Akaike (1973) showed that if the AIC value for the ith model is defined as $2(-LL_i+k_i)$ then the differences of AIC values Δ_i is an approximately unbiased estimate of the difference in the ELL of the two models. This approximation is very good if both models are fairly close to the true process.[37] Another and equally valid description of the AIC is that the model selected minimizes the expected difference, as measured by the Kullback-Leibler (1951) divergence (KL), between the true distribution and the model estimated distribution. These two descriptions of the AIC indicate that (1) comparison of models by the AIC directly answers the PQ, and (2) the comparison is evidential in nature in that it is made by contrasting data-based estimates of a divergence of models from truth.

[35] It was noted in footnote 32, and should be repeated, that one might take a Bayesian stance with respect to model choice and yet maintain as we do here that the probability of the data on a given model is non-credal.
[36] Bayesians are also capable of addressing the PQ.
[37] The true process is of course and unfortunately unknown. Akaike (1973) brilliantly showed that despite this the relative distance (divergence) between two models and truth can still be estimated. Mathematically, this is because the effect of the true process, being common to all models, can be eliminated though the contrasts of information criterion values (see Taper and Ponciano 2016b for a detailed exposition).

For clarity, let us revisit our coin flipping example. The model set is now reduced to two members: M_1 in which we assume the coin is fair, (i.e., the parameter is set to 0.05 without estimation), and M_2 where we estimate the probability of heads on a coin toss from the data. The maximum likelihood estimate of this parameter is 0.58. The AIC value for M_1 is 3.80 while the value for M_2 is 4.52. Thus for this data an AIC comparison indicates that although the difference is slight for predictive accuracy, you are better off continuing to assume a fair coin.

It has been claimed (e.g., Taper and Ponciano 2016a) that, in apparent contrast to this analysis, differences of AIC values are not "evidence functions." The claim is both correct and incorrect. The technical definition of evidence functions (see Lele 2004) has a consistency requirement that the model selected by an evidence function should converge (in probability) to the true model (or its best approximation in the model set) as sample size increases. AIC model selection does not behave in this fashion. It is well known that the AIC makes many over-fitting errors even at large sample size (Taper 2004). AIC model selection does not converge to the true model (or its best approximation in the model set) as sample size increases. So, the correct part of Taper and Ponciano's (2016a) claim is that AIC model selection is not a consistent estimator of causal processes because it is unable to find the true model required for estimating causal processes as the sample size increases.

However, AIC was never designed to be an estimator of causal processes. Therefore, the incorrect or misleading part of their claim is that AIC estimates true causal processes. It was designed to answer what in our view is a different question, which model will have the highest predictive accuracy? Despite its tendency to over-fit in terms of the number of parameters, AIC model selection is consistent for the quantity of interest in the PQ question. The predictive accuracy of the model selected by AIC converges to the predictive accuracy of the true model (or its best approximation in the model set) as sample size increases towards infinity (Stone 1977, 1979). So, AIC model selection is a consistent estimator of the model with the best predictive accuracy.

Clearly, the AIC is operating in a very evidential fashion in its comparison of models through estimated distances. However, it steps just outside the technical definition of an evidence function by seeking to answer the prediction question rather than the evidence question. Here the philosophical analysis has outstripped the statistics and remains a desideratum for further research regarding an extension of the technical definition of evidence to include evidence for features, functions and functionals of models, and not just for models themselves.

2.5 The Decision and Prediction Accounts and Probability Interpretation Complexity

A combination of different probability interpretations is going to occur in applications of Bayesian and AIC frameworks to DQ and PQ that is similar to what we came across in section 2 in the accounts of confirmation and evidence. Recall the Bayesian framework is designed to address the decision question and AIC framework to address the prediction question.

To address DQ, Bayesians apply the idea of maximization of expected utility, which combines an agent's posterior probability for an action with her expected utility from choosing it. Hence, the agent's posterior probability has been used in a particularly significant fashion. We argued that the Bayesian calculation of the posterior probability mixes both credal probability (prior probability) and objective likelihood functions. Therefore, mixing two types of probability interpretations also becomes vital for Bayesians addressing the DQ.

Recall how AIC addressed the PQ. Insofar as melding of different probability interpretations is concerned, since the AIC framework is a penalty function plus a log-likelihood function and some constants, the AIC framework will behave exactly like the likelihood function for the likelihoodist's account as discussed in section 2.

To summarize, the Bayesian response to the Decision Question mixes subjective or credal and objective or non-credal probabilities, as the evidentialist's response to the Prediction Question melds different objective features together in a single non-credal account in the same way in which the two parallel responses to the Confirmation and Evidence Questions similarly mix and meld.

3 Further Corollaries

Towards the end of his foundational essay, *Foresight: Its Logical Laws, Its Subjective Sources*[38], De Finetti has this to say:

> Two completely opposed points of view [on the meaning and the value of the notion of probability]: the first, the most commonly accepted, considers the subjective element of

[38] English translation from the original French (1937) in Kyburg and Smokler (1964).

the naïve notion of probability which is found in our everyday life as a dangerous element which ought to be eliminated in order that the notion of probability be able to attain a truly scientific status; the opposite point of view considers…that the subjective elements are essential, and cannot be eliminated without depriving the notion and theory of probability of all reason for existing.…Both of these two points of view seek to give a well-defined meaning to probability statements, but the domains in which these concepts should receive a meaning are completely different" (p. 149).

We agree that the two conceptions of "probability" differ in essential ways, but contend at the same time that they find, indeed must find, conjoined application to perhaps their most time-honored and well-defined domain, that of statistical inference. For as we have argued in this paper, statistical inference correctly understood in light of distinctions between four fundamental questions that such inference is designed to answer must employ both of these conceptions. Indeed, and in sharp contrast to De Finetti and his many followers, the Bayesian paradigm mixes intuitively "objective" with similarly "subjective" elements, and the likelihood paradigm melds rather different conceptions of how "objectivity" is itself to be understood.[39]

We are not the first to touch upon the complexity of probability interpretation in statistical inference. In a recent book, Elliott Sober (2015) writes, "the idea of each school [i.e., Bayesians and Frequentists being] wedded to a single interpretation of probability is *too simple*." (p.62, emphasis added.) However, we have extended this discussion concerning complexity of probability interpretation in two ways. First, we showed *how* different statistical schools must necessarily incorporate more than one probability interpretation to undertake statistical inference. Second, we extended the complexity of probability interpretation to different questions concerning statistical inference and decision-making. What sort of mixed or melded interpretation one needs to use for practical purposes is contingent on what general type of question one is asking. To repeat: our pluralism is principled. If we are asking the confirmation question, then the account of confirmation combines both subjective and objective interpretations

[39] De Finetti identifies the objective conception with the limiting frequency interpretation of probability, contrasts it point by point with his own subjectivist view, and goes on to underline several of its difficulties. Interestingly, however, the title of his essay suggests an identification of "objective" with "logical," and thus is closer to our own view, on which probability is a logical relationship between data and models, the empirical content of which is estimated in terms of relative (if not also limiting) frequencies and whose explanatory value is to be understood in terms of the causal powers or propensities of physical objects to behave in certain ways in particular kinds of experimental set-ups.

of the probability operators within it. When the evidence question is posed, the Likelihoodist melds three conceptions of "objective probability." If the question is about a decision to make, then the Bayesian decision-theoretic account which includes the posterior probability needs to invoke both subjective and objective probabilities as in the case of the account of confirmation. When the decision question is approached from an evidential perspective, only a melding of different types of objective probabilities is required. When the Evidentialist is asking the prediction question, since the AIC framework is a combination of a likelihood function plus a bias correction in the form of a penalty function, the AIC framework is reducible to the Likelihoodist's, which melds three types or, perhaps less misleadingly, aspects of objective probability. The strongest claim we made is that even with regard to a single question, one has to deploy a mixture or melding of different types or aspects of probability. Failure to recognize this fact is perhaps the main source of the unending and unnecessary 'statistics wars" that have plagued discussions of important methodological questions having to do with model testing[40]. Charges of "subjectivity" and claims of "objectivity" are tossed around without an adequately nuanced understanding of how they are best to be understood. Perhaps the chief implication of our discussion is that these terms, at least on their traditional and vague readings, have outlived their usefulness. Our time is better served in focusing on such distinctions as that between evidence and confirmation.[41]

That said, on our view there are two further and rather closely-related corollaries to draw from our discussion.

The first corollary has to do with so-called "Bayesian convergence." It is worth quoting Hawthorne at some length in this connection.

> If Bayesian induction is to yield either an objective assessment or intersubjective agreement among the agents regarding the inductive support for hypotheses, then the evidence must somehow produce a convergence to agreement among the posterior probabilities of different Bayesian agents in spite of their initial disagreements about the possibilities of these

[40] Our Monograph (2016) discuss extensively the conflation of confirmation and evidence in the philosophical literature and the untoward consequences to which it leads.

[41] In the same way that the distinctions we have drawn are beginning to reform applied statistical inference (see footnote 10 above), we hope that these same distinctions will reform philosophical understanding of same. Our aim here has been to move in a decisive way beyond the discussions of "objectivity"/"subjectivity" in the work of Hacking and Gilles cited in footnote 1.

> hypotheses. Hence, advocates of Bayesian induction have investigated the circumstances under which evidence might "washout" the effects of subjective priors and bring agents into agreement on posterior probabilities. Bayesian convergence theorems establish conditions under which accumulating evidence can induce agents to converge to agreement (1994, p. 241).

Various conditions, event-exchangeability (order-insensitivity and value-replacement of experimental outcomes) and certain lesser-known others have been invoked, and the point at which convergence is reached depends on certain crucial assumptions, in particular on the size of the data set,[42] the identifiability of the models considered, and the suppression of what Kyburg calls "pig-headed priors."[43] From our point of view it does not really matter, for the idea that "Bayesian convergence" guarantees "objectivity" in the needed sense of the word is misleading.

On the one hand, objectivity is here equated with "intersubjective agreement," which is to say that for all of the "consensus" involved, the probability is not agent-independent. However much the prior probabilities of individual agents might "wash out" numerically as data accumulate, the calculation of posterior probabilities via Bayes Theorem still must include reference to them in principle. Priors converge only by way of successive up-datings.[44] "Agreement" is never more than that, a concurrence among beliefs, a shared disposition to operate with the same posteriors.

On the other hand, and more importantly, there is something unacceptably vague as well as powerfully suggestive about the asymptotic intuition embedded in the notion of Bayesian convergence. For it leads naturally if not also logically to the conclusion that as confirming data are accumulated, the hypothesis is a closer and closer approximation to the way things stand in the world. But this Whiggish suggestion is hostage to the history of science; Newton's theory of motion was highly confirmed by data accumulated over almost 200 years and well-nigh universally accepted as true, but although a good approximation for practical purposes to the way things stand, it is not nearly so good an approximation as Einstein's to the way things "really" stand. Even when the point is made in a more Bayesian spirit, by saying that as confirming data for a particular hypothesis or model are accumulated, agents have stronger and stronger reasons

[42] Hawthorne uses "evidence" synonymously with "data," as we do not.
[43] (1983, p. 67).

to believe that it is true, it continues to face the fact that the hypothesis is nonetheless strictly false. Confidence may be raised, even to the point of near-certainty, when it is in hindsight unwarranted. Bayesians sometimes try to avoid this unwelcome conclusion by stretching out "in the long run." 200 years is no more than a drop in the bucket of time. Agreed, but Keynes' celebrated remark to the effect that in the long run we are all dead applies as effectively to the history of science as it does to economic forecasting. Bayesian model identification only converges to truth if the "true" model is in your model set. Otherwise, convergence is just to the best model in the model set, with best model being defined as the model with the smallest Kullback-Leibler divergence to the generating process or "true model."

The second corollary of our earlier discussion of confirmation and evidence helps explain the first. Bayesian confirmation considers hypotheses one at a time; they are believed-true to the degree to which subsequent data-collection raises or lowers their posterior probabilities. Its calculations do not take into account the possibility of alternative hypotheses that could match the data to an even greater degree, where "match the data" entails that the data are more likely on the alternative, that is, provide better *evidence* for it[45]. It's important to note in this connection that "better evidence" does not mean "closer to the truth," still less "should be believed." For the context is always truth- and belief-free. *D* provides better evidence for *H* than for *H'* if and only if *D* are more likely on *H* than on *H*' from which it does not follow that *H* is more likely to be true.. There is no commitment to the belief that a particular model is true.

The underlying problem was characterized succinctly by the statistician G.A. Barnard (1949):

> To speak of the probability of a hypothesis implies the possibility of an exhaustive enumeration of all possible hypotheses, which implies a degree of rigidity foreign to the true scientific spirit. We should always admit the possibility that our experimental results may be best accounted for by a hypothesis which never entered our own heads.

In the present context, his point might be rephrased in this way: Bayesian convergence presupposes an initial probability distribution with respect to which subsequent up-dating takes place, but there is and can be no guarantee that the initial distribution has among its members the

[45] See pages 69-70 of Bandyophadyay, Brittan, and Taper (2016) for more extended discussion.

model that best accounts for the experimental results, which may be outside the model set.[46] It is important to note that Barnard does not say that the unconsidered model must be "closer to the truth."

In discussing Bayesian convergence we have, perhaps inevitably and in reference to Peirce, rather casually introduced the notion of "truth" and committed Bayesian confirmation theory to what is sometimes called the "true-model in model set" assumption. Our critique of this assumption might give the impression that we are also criticizing the idea that the goal of scientific experimentation, model-building, and inference is the discovery of the truth. But this is misleading for at least two reasons. First, although we are skeptical about all *claims* that one or another hypothesis or model is (really) true, we have no reason to be skeptical about the notion of truth itself, still less any reason to reduce or replace it.[47]

As necessary as objective evidence is for the communication and interpretation of scientific results, the discovery of scientific results generally begins in the subjective. Often the word "intuition" is introduced at this point. Such intuition gives rise to thought experiments – famously in the cases of Galileo, Newton, and Einstein – which in turn suggest hypotheses or models, which in a further turn are put to experimental test – rolling balls down inclined planes or measuring tidal motions or placing the Michelson-Morley experiment in a new context. Science, as for that matter does everyday experience, begins with prior probability distributions over outcomes of possible courses of actions. A scientist begins with a plausible idea, refine it as a hypothesis/model, and proceed to see what data are probable on it and to what extent they confirm it[48]. A wise scientist, guided both by an understanding of her own fallibility and by a

---

[46] In the personal decision context, the Barnard objection is less serious than in the confirmation or evidence context. Action is being taken based on a choice among a finite set of models. The belief represented in the prior probability is not so much a belief in the truth of the model, but a belief that a model is the best in a limited set. Problems arise for Bayesian decision theory when public or group decisions are considered. As we have mentioned above, the posterior distribution is biased towards the prior. Whoever controls the prior has an untoward influence on the decision. An evidence of maximum utility decision procedure along the lines we have discussed would avoid this problem, and is worth further research.

[47] The concept of belief is incomprehensible without it.

[48] Kyburg and Meng (2001, pp. 201-202) ask an appropriate question, "what do we need probability for?" and then, echoing Bishop Butler (1692-1752), answer it: "We want probability as a guide in life. For it to function as a *guide*, we require it to be objective rather than subjective. Two agents sharing the same evidence – that is, agreeing on the same objective *facts* about the world – should assign the same probabilities." From our point of view, the answer is much too sweeping, wrongly assimilates *data* and *evidence*, and somewhat ironically makes a case for an exclusively "objective" conception of probability. For if probability is to be the very guide in life it has to

desire for rhetorical power to convince others, then asks what other models could explain the phenomena and what data could discriminate among her models.

This emphasis on the necessity of a jointly subjective-objective account of statistical inference mirrors Kant's account of experience in the *Critique of Pure Reason.* What could be taken as his main theme is that a purely subjective - empiricist or idealist - account is fundamentally incoherent. The concept of the self as a subject of experience requires as a condition of its application an objective world in which that subject can be located and on which it has a temporally-extended perspective. There is no "I" without an "it," no subject of experience without an object of experience, no adequate account of statistical inference without both confirmation and evidence. The subject and the subjective come into play for Kant when human action is at stake. Which is to say that the Bayesian paradigm is uniquely well-suited as an answer to the personal decision question, "what should I do?"[49]

The subjective probabilities can also apply to the question of model choice. The expression "model choice," deeply entrenched in the literature, is ambiguous and depends on the reason for which a choice is desired. It is almost always used to mean something like belief in or acceptance of a model on the basis of its confirmatory or evidential superiority, something after-the-fact. But we think that it is also appropriately applied to the act of choosing which models to put to experimental test, before any data have been accumulated in a systematic way. This is where prior probabilities play an indispensable role. De Finetti gets at this, or at least can be taken as getting at something like this.[50]

> Rather than by seeking to bring everything back to the objective, one can attain clarity by reducing any such concept systematically to the subjective; the value of a concept would then result from the analysis of the deep and essential reasons which have made us,

---

be connected to action by way of belief and normative in character. That is, probability is a "guide" insofar as it offers an account of what should be believed on the basis of what we have experienced. But of course this is precisely what the Bayesian account, eschewed by Kyburg and Meng, does. The limitations with the Bayesian account of probability lie in a different direction. To put it as briefly as possible, it doesn't capture or quantify the uncertainty embedded in our statistical inferences, intrinsic elements of our scientific world-view, and uneliminable dimensions of our lived experience. To say that probabilities can never have 0 or 1 as their values doesn't begin to scratch these just-mentioned surfaces.

[49] The fact that the Bayesian probabilifier is routinely identified as an "agent" hints at this; the Bayesian agent *collects* data, *up-dates* posteriors*, makes* decisions concerning the initial distribution of probabilities over models. So far as evidence is concerned, there is no "doing," no temporal-intervals, no agent.

[50] The word "sources" in the title of his essay connotes the genesis of hypotheses, not their assessment.

perhaps unconsciously, introduce it, and which furnish us with the explanation of its usefulness (Kyburg and Smokler, p.149).

.

4 Conclusion

We have not argued that the Subjective Bayesian position is wrong-headed. Our view is that its account of confirmation is an indispensable component of an adequate account of personal epistemology and decision. All we claim is that a thoroughly objective account of evidence must be added to it for the public epistemology of scientific inference. At least two statistical paradigms and multiple interpretations of probability are necessary. Their respective tasks can, as we have tried to show, be set out in a principled way.

This is, of course, not the end of the matter. It is one thing to distinguish between types or aspects of interpretations of probability, and to note the distinction syntactically, and another to make each adequately clear. There are various axiomatizations of belief and propensity probability concepts,[51] but not yet general agreement on which axioms should be taken as standard. Indeed, there is no longer general agreement that the Kolmogorov axioms constitute the essential core of *any* adequate explication of these concepts.[52]

So far as we can see, our argument in this paper does not depend on settling any of these questions here. It is enough to show here that within the Bayesian paradigm there is a mixing of credal and non-credal concepts of probability, and that within the evidentialist paradigm a melding of different non-credal components, and that the two paradigms are indispensable both in making clear the distinction between evidence and confirmation and in indicating the ways in which what we call the decision and prediction questions are to be addressed. That noted, a further sharpening of the argument is necessary. But there is enough here on the basis of which to refine statistical practice and reform philosophical understanding, on the one hand in drawing distinctions, e.g., between confirmation and evidence, the neglect of which leads to confusion, and on the other in being persuaded that the labeling and subsequent criticism of one or another statistical paradigm as "subjective" or "objective" is to paint with too broad and misleading a

[51] As an example of the former see (Crupi, et. al., 2013), of the latter see (Suppes. 1973).
[52] See (Hájek, 2012).

brush. One must proceed in a more principled and particular way. Or so we have tried to demonstrate.

5 References

Abrams, Marshall. (2012): “Mechanistic Probability.” *Synthèse*, **187**(2), July, pp. 343-375.

Achinstein, Peter. ed. (1983): *The Concept of Evidence*. (Oxford University Press).

Akaike, Hirotuga. (1983): “Information Theory as an Extension of the Maximum Likelihood Principle.” Pages, 267-281 in *Second International Symposium on Information Theory Akademiai Kiado*. Budapest.

Bandyopadhyay, Prasanta. and Gordon. Brittan (2010): “Two Dogmas of Strong Objective Bayesianism.” *International Studies in the Philosophy of Science* 24(1), pp. 45-65.

Bandyopadhyay, Prasanta, Gordan Brittan Jr, and Mark. L. Taper. 2016. *Belief, Evidence, and Uncertainty: Problems of Epistemic Inference*. (New York: Springer Verlag).

Barnard, George A. 1949. Statistical Inference. *Journal of the Royal Statistical Society Series B-Statistical Methodology* 11:115-149.

Berkovitz, Joseph. 2015. “The Propensity Interpretation of Probability: A Re-evaluation.” *Erkenntnis*, prepublication version available at http://link.springer.com/article/10.1007%2Fs10670-014-9716-8.

Berger, James O. 1985. *Statistical Decision Theory and Bayesian Analysis*, 2nd edition. (New York: Springer Verlag).

Bickel, David R. 2012. The strength of statistical evidence for composite hypotheses: inference to the best explanation. *Statistica Sinica* **22**, pp. 1147-1198.

Brittan, Gordon. 1994. "Kant and the Quantum Theory," in Parrini, ed. *Kant and Contempoary Epistemology*. (Dordrecht: Kluwer), pp. 131-55.

Burnham, Kenneth, and David Anderson. 2002. *Model Selection and Multi-Model Inference*. (SpringerVerlag, New York).

Christensen, David. 1999. "Measuring Confirmation." *Journal of Philosophy*, **xcvi,** pp. 437-461.

Vincenzo, Crupi, Nick, Chater, and Katya Tentori 2013. "New Axioms for Probability and Likelihood Measures." *British Journal for the Philosophy of Science* 64: 189-204.

Carnap, Rudolf. 1950. *Logical Foundations of Probability*. (Chicago: University of Chicago Press).

De Finetti, Bruno. 1937. "La prévision: ses lois logiques, ses sources subjectives." *Annales de l'Institut Henri Poincaré, Vol. 7.*

Earman, John. 1992. *Bayes or Bust: A Critical Examination of Bayesian Confirmation Theory.* (Cambridge, MA: MIT Press).

Edwards, Ward, Harrold Lindman, and Leonard J. Savage. 1963. "Bayesian Statistical Inferences for Psychological Research". *Psychological Review* 70.3: pp. 193-242.

Edwards, Anthony W. F. 1972. *Likelihood.(*Cambridge: Cambridge University Press).

Eells, Ellery. 1985. "Problems of Old Evidence." *Pacific Philosophical Quarterly,* **66**, pp. 283-302.

Eells, Ellery and Branden Fitelson. 2000. "Measuring Confirmation and Evidence." *Journal of Philosophy,* **XVII** (12), pp. 663-672.

Eells, Ellery and Branden Fitelson. 2002. "Symmetries and Asymmetries in Evidential Support." *Philosophical Studies,* **107**, pp. 129-142.

Evans, Michael. 2016. ""Measuring Statistical Evidence Using Relative Belief." *Computational and Structural Biotechnology Journal*, **14**, 91-96.

Gillies, Donald. 1986. "In Defense of the Popper-Miller Argument." *Philosophy of Science*, **53**, pp. 110-113.

Gillies, Donald. 2000. "Varieties of Propensity," *British Journal for the Philosophy of Science,* **51**: pp. 807-35.

Gillies, Donald. 2003. *Philosophical Theories of Probability.* London: Routledge.

Good, Irving J. 1984. "The Best Explicatum for Weight of Evidence." *Journal of Statistical Computation and Simulation*, **19**, pp. 294-299.

Goodman, Nelson. 1965. *Fact, Fiction, and Forecast*. Cambridge, MA. Harvard University Press.

Hacking, Ian. 2001. *An Introduction to Probability and Inductive Logic*. London: Cambridige.

Hàjek,, Alan. 2012. "Interpretations of Probability." *The Stanford Encyclopedia of Philosophy* (Winter, 2012) edition), E.N. Zalta, ed. (http://plato.stanford.edu/archives/win2012/entires/probability-interpret/>.

Hawthorne, James H. "On the Nature of Bayesian Convergence," *PSA 94*, Vol. 1, pp. 241-249.

Howson, Colin 2013. "Exhuming the No-miracles Argument." *Analysis*, **73**, pp. 205-211.

Humphreys, Paul. 1985. "Why Propensities Cannot Be Probabilities," *Philosophical Review*, **94**: pp. 357-70.

Keynes, John M. 1921. *A Treatise on Probability*. London: Macmillan.

Klement, Rainer J. (2017). "Beneficial Effects of Ketogenic Diets for Cancer Patients – A Realist View with Focus on Evidence and Confirmation." www.ncbi.nlm.nih.gov/pubmed/28653283.

Kyburg, Henry E. and Howard E. Smokler (1964). *Studies in Subjective Probability* (New York: Wiley).

Kyburg, Henry E. (1983). *Epistemology and Inference.(* St. Paul, MN: University Press).

Kyburg, Henry E. and Choh M. Teng (2001). *Uncertain Inference*. Cambridge: Cambridge University Press.

Lele, Subhash R. 2004. "Evidence Functions and the Optimality of the Law of Likelihood." In (Taper and Lele, 2004).

Levi, Isaac. 1967. "Probability Kinematics". *British Journal for the Philosophy of Science* 18: pp. 200-205.

Lewis, David. 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)

Mayo, Deborah G., and Ari Spanos. 2006. "Severe testing as a basic concept in a Neyman-Pearson philosophy of induction." *British Journal for the Philosophy of Science* **57**: pp. 323-357.

Neyman, Jersey., and Egon. S. Pearson. 1933. "On the Problem of the Most Efficient Tests of Statistical Hypotheses." *Philosophical Transactions of the Royal Society of London. Series A*, **231**: pp. 289-337.

Pulgarin, Paulo C., Juan P. Gomez, Scott Robinson, Robert E. Ricklefs, and Carlos D. Cadena. 2018. "Host Species, and Not Environment, Predicts Variation in Blood Parasite Prevalence, Distribution, and Diversity Along a Humidity Gradient in Northern South America." *Ecology and Evolution*, **13**; 8(8), pp. 3800-3814.

Patil, Ganapati P., Calyampudi R. Rao, and Makarand. V. Ratnaparkhi. 1986. "On discrete weighted distributions and their use in model choice for observed data." *Communications in Statistics-Theory and Methods* 15:907-918.

Peirce, Charles. S. 1878. "Illustrations of the Logic of Science III —The doctrine of chances." *Popular Science Monthly* 12:604-615.

Pettigrew, Richard. (2016). *Accuracy and the Laws of Credence*. Oxford. Oxford University Press.

Popper, Karl R. 1957. "The Propensity Interpretation of the Calculus of Probability and the Quantum Theory". In Korner, S., ed. *Observation and Interpretation: Proceedings of the Ninth Symposium of the Colston Research Society*, University of Bristol, pp.65-70 and 88-89.

Popper, Karl. R. 1959. "The Propensity Interpretation of Probability." *British Journal for the Philosophy of Science* 10:25-42.

Ramsey, Frank P. (1931): "Truth and Probability." In R.B. Braithwaite, ed., *The Foundations of Mathematics and Other Logical Essays.* London: Routledge and Kegan Paul.

Reichenbach, Hans. 1949. *The Theory of Probability*. Berkeley: University of California Press.

Rosenkrantz, Rodger D. 1994. "Bayesian Confirmation: Paradise Regained." *British Journal for the Philosophy of Science*, **45**, pp. 467-476.

Royall, Richard. 1977. *Statistical Evidence: A Likelihood Paradigm*. New York: Chapman Press.

Salmon, Wesley C. 1983: "Confirmation and Evidence." In P. Achinstein, ed., *The Concept of Evidence*, Oxford: Oxford University Press, 1983.

Sober, Eliot. 2015. *Ockham's Razor: A User's Manual*. Cambridge: Cambridge University Press.

Stich, Stephen P. and Richard E. Nisbett. 1980. "Judgment and the Psychology of Human Reasoning." In *Philosophy of Science*, 47, pp.188-202.

Stone, Mervyn. 1977. "Asymptotic Equivalence of Choice of Models by Cross-Validation and Akaike's Criterion". *Journal of the Royall Society Series B-Methodological* 39: pp.44-47.

Stone, Mervyn. 1979. "Model Selection Criteria of Akaike and Schwartz." *Journal of the Royal Statistical Society Series Methodological* 41: pp. 276-278.

Suppes, Patrick C. 1973. "New Foundations of Objective Probability: Axioms of Propensities," in P. Suppes, et. al., eds, *Logic, Methodology, and Philosophy of Science IV: Proceedings of the Fourth International Conference for Logic, Methodology, and Philosophy of Science* (Amsterdam: North-Holland).

Suppes, Patrick C. 1987. "Propensity Interpretations of Probability." *Erkenntnis* **26**: pp. 335-58.

Taper, Mark .L., and José.M. Ponciano, 2016a. “Evidential Statistics as a Statistical Modern Synthesis to Support 21st Century”. *Population Ecology*. 58. Pp.9-29.

Taper, Mark .L., and José.M. Ponciano. 2016b. “Projections in model space: Multimodel inference beyond model averaging “in (Bandyopadhyay, Prasanta, Gordan Brittan Jr, and Mark. L. Taper. 2016. *Belief, Evidence, and Uncertainty: Problems of Epistemic Inference*. New York: Springer Verlag)

Taper, Mark L. 2004. “Model Identification from Many Candidates.” Taper and Lele (eds.) 2004. Pp. 488-501.

Taper, Mark L. and Subhash R. Lele, eds. 2004. *The Nature of Scientific Evidence*: Chicago: University of Chicago Press.

Taper, Mark. L., and Subhash. R. Lele. 2011. “Evidence, Evidence Functions, and Error Probabilities” in P. S. Bandyopadhyay and M. R. Forster, editors. *Philosophy of Statistics*. (Oxford: Elsevier).

Vickers, John M. (1988). *Chance and Structure: An Essay on the Logical Foundations of Probability*. (Oxford: Clarendon Press).

Von Mises, Richard. 1957. *Probability, Statistics, and Truth*, revised English edition. New York: Macmillan.