Standards for *Modest* Bayesian Credences

Jessi Cisewski, Joseph B. Kadane, Mark J. Schervish, Teddy Seidenfeld, and Rafael Stern

Yale U., Carnegie Mellon U. (× 3), and the Universidade Federal de São Carlos.

Abstract

In response to G.Belot's (2013) criticism that Bayesian theory is epistemologically *immodest*, we argue that his analysis is misguided. The topological conditions that we understand underpin his criticisms of familiar results about asymptotic Bayesian conditioning are self-defeating. They require extreme *a priori* credences regarding, e.g., the limiting behavior of observed relative frequencies. Instead, we offer a rival explication of Bayesian *modesty*: Rival scientific opinions should be responsive to new facts as a way to resolve their disputes. Using a result of Blackwell and Dubins (1962), we explain how amenability to new evidence may serve as the basis for resolving conflicts among Bayesian investigators. When the new evidence fails to achieve a resolution, that failure can identify epistemologically *immodest* Bayesian credal states. Also we assess A. Elga's (2016) rebuttal to Belot's analysis. He focuses attention on the role that the assumption of countable additivity plays in Belot's criticism of asymptotic Bayesian learning.

1. *Introduction*. Consider the following compound result about asymptotic statistical inference. A community of Bayesian investigators who begin an investigation with conflicting opinions about a common family of statistical hypotheses use shared evidence to achieve a consensus about which hypothesis is the true one. Specifically, suppose the investigators agree on a partition of statistical hypotheses and share observations of an increasing sequence of random samples with respect to whichever is the true statistical hypothesis from this partition.[1] Then, under various combinations of formal conditions that we review in this essay, *ex-ante* (i.e., prior to accepting the new

---

[1] Let H be a (simple) statistical hypothesis for the random variable *X,* i.e., where the conditional probability distribution P(X | H) is well defined. Given H, a *random sample* of size *n* from this distribution, $\{X_1, \ldots, X_n\}$ has a joint distribution that is identically, independently distributed (iid) with respect to P(X | H): P(X_1, \ldots, X_n | H) = $\prod_{i=1}^{n} P(X_i | H)$.

evidence) it is practically certain that each of the investigators' conditional probabilities approach 1 for the one true hypothesis in the partition.

The result is compound: individual investigators achieve asymptotic certainty about the unknown, true statistical hypothesis. Second, the shared evidence leads to a consensus among the different investigator's individual degrees of belief. The initial disagreements, the disparate initial credences about different hypotheses, are resolved with increasing shared evidence. Stated in more familiar Bayesian terms, it is practically certain that the likelihood function based on the shared statistical evidence swamps differences in initial *prior* credences to produce a consensus among *posterior* credences.

The strategy to use asymptotics of Bayesian inference to defend against charges of excessive subjectivity is highlighted in the seminal work of Savage (1954, Sections 3.6 and 4.6), and Edwards, Lindman, and Savage (1963). Savage's (1954) results apply to a finite set of investigators who hold non-extreme views over a common finite partition of statistical hypotheses.[2] He establishes that – using a (finitely additive) weak-law of large numbers – given increasing statistical evidence from a sequence of random samples, with probability approaching 1, different non-extreme personalists' conditional probabilities become ever more concentrated on the same one true statistical hypothesis from among a finite partition of rival statistical hypotheses.[3,4] To repeat, the result is compound. It

---

[2] Say that an investigator with degree of belief represented by a probability $P(\cdot)$ holds a non-extreme opinion about an event E if $0 < P(E) < 1$.

[3] Savage's (1954) axiomatic theory of preference, based on postulates P1-P7, is about an idealized Bayesian agent's *static* preference relation over pairs of acts – preferences at one time in the idealized agent's life. The theory of personal probability and conditional probability that follows from P1-P7 is about an idealized agent's epistemic state at that one time: her/his degrees of belief and conditional degrees of belief at that one time. More familiar in the Bayesian literature is a *dynamic* Bayesian rule where conditional probability models the idealized agent's changing beliefs over time, when new evidence is accepted. For details on differences between the *static* and *dynamic* use of conditional probability, see Levi (1984, §4.3).

[4] We illustrate the weak and strong laws of large numbers for independent, identically distributed Bernoulli trials. Let X be a Bernoulli variable with possible values $\{0, 1\}$, where $P(X = 1) = p,$ for some $0 \leq p \leq 1$. Let X_i ($i = 1, 2, \ldots$) be a denumerable sequence of Bernoulli variables, with a common parameter $P(X_i = 1) = p$ and where trials are independent. Independence is expressed as follows. For each $n = 1, 2, \ldots$, let $S_n = \sum_{i=1}^{n} X_n$. Then $P(X_1 = x_1, \ldots, X_n = x_n) = p^{S_n} \times (1 - p)^{(n - S_n)}$.

The weak-law of large numbers for *iid* Bernoulli trials asserts that for each $\varepsilon > 0$,

$$\lim_{n \to \infty} P(|S_n/n - p| > \varepsilon) = 0.$$

addresses both issues of *certainty* and *consensus* among finitely many investigators over a finite partition of statistical hypotheses, assuming they share an increasing sequence of observations from random sampling.

Savage offers these findings as a partial defense against the accusation, voiced by frequentist statisticians of the time, that the theory of (Bayesian) personalist statistics is fraught with subjectivism and cannot serve the methodological needs of the Scientific community, where objectivity is required. The central theme in Savage's response is to understand 'objectivity' in terms of shared agreements about the truth, particularly, when the shared agreements arise from shared statistical evidence. In summary, Savage provides sufficient conditions for when Bayesian methodology makes it *ex-ante* almost certain that shared evidence secures this kind of objectivity for a well-defined community of investigators.

Savage (1954, p. 50) notes that his result about asymptotic certainty can be extended in several ways, by adapting the central limit theorem, the strong law of large numbers, and the law of the iterated logarithm to sequences of conditional probabilities generated by an increasing sequence of random samples. The last two of these laws require stronger assumptions than are needed for the finitely additive weak-law convergence result that Savage presents. Specifically, these stronger results require the assumption that (conditional) probabilities are countably additive.

Savage's twin results have been strengthened also to include shared evidence from non-random samples. Consider an uncountably infinite probability space generated by increasing finite sequences of observable random variables, not necessarily forming a random sample with respect to a statistical hypothesis of interest. Rather than requiring that different agents hold non-extreme views about all possible events in the space of observables, which is mathematically impossible with real-valued probabilities once the

---

The strong-law of large numbers asserts that,

$$P(\lim_{n \to \infty} S_n/n = p) = 1.$$

If P is countably additive, the strong-law version entails the weak-law version.

space is uncountable, instead require that they agree with each other about which events in this uncountably infinite space of observables have probability 0. They share in a family of *mutually absolutely continuous* probability distributions. If the agents' personal probabilities over these infinite spaces also are countably additive, then strong-law convergence theorems yield strengthened results about asymptotic consensus (see, e.g. Blackwell and Dubins, 1962) and also about asymptotic certainty for events defined in the space of sequences of increasing shared evidence. We discuss several of these results in Section 4. There we use considerations both of *certainty* and *consensus* to explicate epistemic *modesty* within a Bayesian framework that contrasts with a critical assessment of Bayesian theory offered by G. Belot (2013), whose work we next consider.

2. *Orgulity as identified by comparing meager sets versus null sets.*
In a 2013 paper in this *Journal*, critical about the methodological significance of some of the strengthened versions of Savage's convergence result for asymptotic certainty, G. Belot arrives at a harsh conclusion:

> The truth concerning Bayesian convergence-to-the truth results is significantly worse than has been generally allowed – they constitute a real liability for Bayesianism by forbidding a reasonable epistemological modesty.
> [Belot, 2013] p. 502

Below, we argue that this verdict is misguided. The criteria for reasonable epistemic modesty that we understand underpin Belot's analysis are self-defeating; hence, his argument is not compelling. When the criteria that we attribute to Belot are satisfied they induce unreasonable epistemic *apriorism* regarding, e.g., how sequences of observed relative frequencies behave.

What makes a (coherent) Bayesian credal state *over-confident* and lacking in *epistemological modesty*? Does the Bayesian position generally forbid "a reasonable epistemological modesty," as Belot intimates? These questions are both interesting and imprecise. There is no doubting that the standard of mere Bayesian *coherence* for a credal state, as formalized in de Finetti's (1937) theory, falls short of characterizing the set of *reasonable* credal states. To use an old and tired example, a person who thinks

each morning that it is highly probable that the world ends later that afternoon does not thereby violate the technical norms of *coherence*.

In order to identify a brand of unreasonableness captured in over-confident, epistemologically *immodest* credal states, Belot supplements Bayesian coherence with a topological standard for respecting what he calls a *typical* event: He defines a *typical* event as a topologically *large* event. When a coherent agent assigns probability 0 to a topologically large set, specifically when a probability null set is *comeager*, Belot thinks that is a warning sign of *epistemological immodesty*.[5] Such a Bayesian agent is practically certain that the topologically *typical* event does not occur. And then Bayesian conditioning (almost surely) preserves that certainty in the face of new evidence. So, the Bayesian agent is not *open-minded* because, in dismissing as probabilistically negligible a topologically *typical* event E, also (almost surely) she is aware *ex-ante* that Bayesian conditioning precludes learning that the *typical* event E occurs.

We understand Belot's criticism (2013, Section 4) to be that Bayesian convergence-to-the-truth results about hypotheses that are formulated in terms of sets of observable sequences fail this concern about *typical* events. The strengthened convergence results allow the Bayesian agent to dismiss (*ex-ante*) a probabilistically negligible set of sequences of observations where the convergence-to-the-truth fails. This set has "prior" probability 0. Except, Belot complains, that failure set may be comeager in the usual topology for the sequences of observables. Hence, the failure set may be a *typical* event in the space of observables, about which a *modest* investigator should keep an open mind. But, instead, Bayes updating (almost surely) ignores these *typical* events by continuing to assign them probability 0, even as the evidence grows. Thus, the strengthened asymptotic certainty results that Belot criticizes do not conform to the topological standards of epistemic *modesty* in the sense of *modesty* that we understand he advocates.

---

[5] A topologically *meager* set is one that is a denumerable union of nowhere dense sets. A *comeager* set, or a *residual* set, is the complement of a meager set.

Though he does not explicitly formulate criteria for *immodesty*, based on the examples and analysis he offers, we understand Belot's primary requirements to be these two[6]:

- ***Topological Condition #1***: Do not assign probability 1 to a meager set of observables. Also, we find that Belot argues for a more demanding standard,[7]

- ***Topological Condition #2***: Assign probability 0 to each hypothesis that is a meager set in the space of sequences of observables.

Ordinary statistical models violate Topological Condition #1 by their unconditional probabilities, independent of whether learning is by Bayesian updating. Already, Condition #1 is inconsistent with the strong laws of large numbers, including the ergodic theorem, which are asymptotic results for unconditional probabilities. (See Oxtoby [1980, p. 85].)

Here we show that Topological Condition #2 entails a radical probabilistic *apriorism* towards observed relative frequencies that has little to do with questions about Bayesian over-confidence. In particular, this topological standard requires that with probability 1, relative frequencies for an arbitrary sequence of (logically independent) events oscillate maximally. From a Bayesian point-of-view, almost surely new evidence leaves this extreme epistemic attitude wholly unmodified. A Bayesian agent whose credal state conforms to Condition #2 knows *ex-ante* that she is practically certain never to change her mind that the relative frequencies for a sequence of events oscillate maximally. In this sense, we find that Conditions #1 and #2 are self-defeating through a lack of

---

[6] A helpful referee suggests that Belot might instead restrict Topological Condition #1 to the space of *chance* hypotheses rather than extending it to the space of observables as we do. However, Belot's criticism of Bayesian methodology in Section 4 of his (2013) depends upon applying topological standards for a "typical event" to probabilities over hypotheses defined in terms of sequences of observables. There are no *chance* hypotheses in those cases.

[7] Belot (2013, p. 488), **Remark 2**, notes that when R is a meager element of a measurable space $<\Omega, \mathcal{B}>$ then the set of probabilities that assign R probability 0 is comeager in the space of probability distributions on the same measurable space $<\Omega, \mathcal{B}>$. This points to what we label as ***Topological Condition #2***. It is a higher-order application of Belot's idea that a topologically typical set (i.e., a comeager set) should be reflected with probability 1. For Condition #2, that reasoning is applied to typical sets of probabilities.

6

humility. They promote excessive *apriorism* with respect to ordinary properties of limiting frequencies.

The Bayesian convergence-to-the-truth results that are the subject of Belot's complaints are formulated as probability strong-laws that hold *almost surely* or *almost everywhere*. In order to make clear why we think Belot's verdict is mistaken thinking these results about convergence to the truth are a liability for Bayesian theory, revisit the familiar instance of the strong law of large numbers, as reported in *fn*. 4.

Let $<\Omega, \mathcal{B}, P>$ be the countably additive measure space generated by all finite sequences of repeated, probabilistically independent [*iid*] flips of a "fair" coin. Let 1 denote a "Heads" outcome and 0 a "Tails" outcome for each flip. Then a point x of Ω is a denumerable sequence of 0s and 1s, $x = <x_1, x_2, \ldots >$, with each $x_n \in \{0, 1\}$ for $n = 1, 2,$ $\ldots$. Let $X_n(x) = x_n$ designate the random variable corresponding to the outcome of the nth flip of the fair coin. $\mathcal{B}$ is the Borel σ-algebra generated by *rectangular* events, those determined by specifying values for finitely many coordinates in Ω. P is the countably additive *iid* product *fair-coin* probability that is determined by

$$P(X_n = 1) = 1/2 \ \ (n = 1, 2, \ldots)$$

and where each finite sequence of length n is equally probable,

$$P(X_1 = x_1, \ldots, X_n = x_n) = 2^{-n}.$$

Let $L^{\frac{1}{2}}$ be the set of infinite sequences of 0*s* and 1*s* with limiting frequency ½ for each of the two digits: a set belonging to $\mathcal{B}$. Specifically, let $S_n = \sum_{i=1}^{n} X_n$. Then $L^{\frac{1}{2}} = \{x:$ $\lim_{n \to \infty} S_n/n = 1/2\}$. The strong-law of large numbers asserts that $P(L^{\frac{1}{2}}) = 1$. What is excused with the strong law, what is assigned probability 0, is the null set $N (= [L^{\frac{1}{2}}]^c)$ consisting of the complement to $L^{\frac{1}{2}}$ among all denumerable sequences of 0*s* and 1*s*.

The null set N is large, both in cardinality, and in category under the product topology for 2^{ω}. It is a set with cardinality equal to the cardinality of its complement, the cardinality

of the continuum.[8] When 2^{ω} is equipped with the infinite product of the discrete topology on $\{0, 1\}$,[9] then the null set N is topologically large. N is a comeager set (Oxtoby, 1980, p. 85).[10] That is, the set $L^{\frac{1}{2}}$ is meager and so is judged topologically "small," or *atypical*. By Condition #1, a Bayesian who adopts the *fair coin* model for her/his credences is epistemologically *immodest* with respect to denumerable sequences of possible coin flips: the space of sequences of observations that drive the asymptotic certainty result.

This strong-law counterexample to Condition #1 should come as no surprise in the light of the following result, reported by Oxtoby (1980, T.1.6):

(*)   Each non-empty interval on the real line may be partitioned into two sets, $\{N, M\}$ where N is a Lebesgue measure null set and its complement $M = N^{\mathbf{c}}$ is a meager set.

Oxtoby generalizes (*) with his Theorem T.16.5.[11] In his illustration of T.16.5 using the strong-law of large numbers, the binary partition $\{N, L^{\frac{1}{2}}\}$ displays the direct conflict between the measure theoretic and topological senses of *small*. N has probability 0 under the fair-coin model, and $L^{\frac{1}{2}}$ is a meager set in the product topology of the discrete topology on $\{0,1\}$. The tension between the two senses of *small* is not over some

---

[8] For each $0 \leq y \leq 1$, with $y \neq \frac{1}{2}$, N contains at least one sequence with limiting relative frequency y, and these are pairwise different sequences for different values of y.

[9] This product topology is homeomorphic to the Cantor Space.

[10] Oxtoby sketches the proof of this claim in the *Supplementary Notes* [1980, p. 99]. The claim follows from an elegant application of the Banach-Mazur Game. Belot's (2013, p. 498) Remark 5, fn. 41, adapts Oxtoby's argument to show the following.

Consider a point x in Cantor Space. A prior P is "open minded" with respect to the hypothesis x provided that, given any finite initial segment of x, $(x_1, \ldots, x_m)$, there is a finite continuation $(x_1, \ldots, x_m, x_{m+1}, \ldots, x_n)$ where $P(\{x\} \mid (x_1, \ldots, x_n)) > .50$, and there exists some other finite continuation of $(x_1, \ldots, x_m)$, $(x_{m+1}, \ldots, x_{n'})$ where $P(\{x\} \mid (x_1, \ldots, x_{n'})) < .50$. Say that hypothesis x *flummoxes* prior P provided that, for infinitely many values of n, $P(\{x\} \mid (x_1, \ldots, x_n)) > .50$ and for infinitely many values of n, $P(\{x\} \mid (x_1, \ldots, x_n)) < .50$. Then, the set of sequences in Cantor Space (i.e., the set of hypotheses) that *flummoxes* an open minded P is comeager in the infinite product topology on Cantor Space. What Belot observes is a special case of Proposition (**), which we introduce and discuss on the next page.

[11] Oxtoby's [1980, p. 64] Theorem 16.5 establishes that if the measure space $< X, \mathbfcal{B}, P>$, satisfies

- P is nonatomic,
- X has a metrizable topology $\mathcal{T}$ with a base whose cardinality is less than the first weakly inaccessible,
- and, the σ-field $\mathbfcal{B}$ includes the Borel sets of $\mathcal{T}$,

then X can be partitioned into a set of P-measure 0 and a meager set.

esoteric binary partition of the space of binary sequences, but applies to the event that the sequence of observed outcomes has a limiting frequency 1/2.

We exemplify the general conflict encapsulated in Oxtoby's T. 16.5 with the following claim, which we use to criticize Condition #2.

Consider the space 2^{ω}, with points $x = <x_1, x_2, \ldots >$ of denumerable sequences of 0s and 1s, equipped with the infinite product of the discrete topology on $\{0,1\}$. Define the set of sequences $L^{<0,1>}$ consisting of those points x whose relative frequency does not oscillate maximally, i.e., where,

$$lim.inf. \; \textstyle\sum_{j=1}^{n} x_j/n \; > 0 \;\; \text{or} \;\; lim \; sup. \; \textstyle\sum_{j=1}^{n} x_j/n \; < \; 1.$$

The complement to $L^{<0,1>}$, $\mathbf{OM} = [L^{<0,1>}]^{\mathbf{c}}$, is the set of binary sequences whose observed relative frequencies *Oscillate Maximally*.

(**) *Proposition*: $L^{<0,1>}$ is a meager set, i.e., $\mathbf{OM}$ is a comeager set.[12]

Theorem A1 of the *Appendix* establishes that sequences of logically independent random variables that *oscillate maximally* are comeager with respect to infinite product topologies on the sequence of random variables. *Proposition* (**) is a corollary to *Theorem A1* applied to binary sequences, i.e. where there are only two categories for observables.

What *Proposition* (**) establishes is that only extreme probability models of relative frequencies satisfy Topological Condition #2. That is, consider a measure space $<2^{\omega}, \mathcal{B},$ P> where $\mathcal{B}$ includes the Borel sets from 2^{ω}, and where 2^{ω} is equipped with the infinite product of the discrete topology as above. Each probability with $P(L^{<0,1>}) > 0$ produces a non-null set that is meager.

---

[12] This Proposition is established by Calude and Zamfirescu (1999) using an application of Oxtoby's (1957) theorem for the Banach-Mazur Game. In Appendix A, we establish the more general Theorem A1 with a direct argument, which extends beyond Oxtoby's (1957) theorem for the Banach-Mazur Game, and which has the Proposition (**) as a Corollary.

Unless a probability model P for a sequence of relative frequencies assigns probability 1 to the set of sequences of observed frequencies that oscillate maximally, then P assigns positive probability to a meager set of sequences, in violation of Condition #2. Evidently, the standard for epistemological *modesty* formalized in Topological Condition #2, which requires meager sets of relevant events be assigned probability 0, itself leads to probabilistic orgulity because it requires an unreasonable *a priori* opinion about how observed relative frequencies behave. Let P satisfy Condition #2. Given evidence of a P-non-null observation *o* of observed relative frequencies, the resulting conditional probability leaves this extreme *a priori* opinion unchanged: P(**OM** | observation *o*) = 1.

Familiar Bayesian models also violate the weaker Topological Condition #1. Consider an exchangeable probability model over 2^{ω}. Then, by de Finetti's Theorem [1937] each exchangeable probability assigns probability 1 to the set L of sequences with well defined limiting frequencies for 0s and 1s. That is, then P$\{x$: *lim.inf.* $\sum_{j=1}^{n} x_j/n = $ *lim. sup.* $\sum_{j=1}^{n} x_j/n\} = 1$. But L is a subset of $L^{<0,1>}$; hence, L is a meager set.

In summary, our understanding is that Belot applies Topological Conditions #1 and #2 in order to identify an epistemically *immodest* coherent credal state. We find that each of these two conditions is excessively restrictive and is self-defeating as a criterion for epistemic *immodesty*. The credences that satisfy these conditions with respect to the sets of sequences of observables that ground the almost-sure Bayesian convergence results embed extreme a priorism about, e.g., the limiting behavior of observed relative frequencies.

In Section 4, below, we argue that a better account of Bayesian epistemological *modesty/immodesty* uses interpersonal standards for asymptotic *consensus* within a community of investigators about the set of *certainties* that arise from an idealized sequence of observations. Belot's approach for identifying epistemic *immodesty* applies topological conditions of adequacy to a stand-alone credence function and avoids issues

of *consensus*. In contrast, we supplement coherence with criteria involving asymptotic consensus among a community of investigators about which certainties they might acquire based on a sequence of shared evidence.

3. *But what if probability is merely finitely additive and not countably additive?*
Elga (2016) responds to Belot's criticism by focusing on the premise of countable additivity for probability, which is needed for the strong-law versions of Savage's convergence result. The subjective theory of probability, especially as promoted by de Finetti (1974), Dubins and Savage (1965) and Savage (1954), does not mandate countable additivity for credences. This added generality is of importance for contemporary Bayesian practice, as argued in Kadane et al. (1986).

As we understand Elga's response to Belot's criticism, it is based on an example. The example purportedly shows how, using a finitely but not countably additive probability P, Belot's standard for being an *open-minded* Bayesian credal state may be satisfied without also being burdened with the *immodesty* of treating a comeager failure set as a P-null set, as follows when probability is countably additive. Elga argues that, in his example, the associated set of data-sequences where the convergence to the truth result fails with the credal state P has positive P-probability, contrary to what happens in the countably additive case. Elga asserts that in his example, the agent's finitely additive conditional probabilities do not (almost surely) converge to the true statistical hypothesis about limiting relative frequencies; hence, such a Bayesian agent escapes Belot's criticism as this agent is epistemologically *humble* about becoming certain of the true limiting relative frequency in the observed sequence.

First and foremost, we dispute Elga's analysis of the specific example he offers. We argue that, contrary to Elga's assessment, his merely finitely additive probability model P satisfies a finitely additive convergence-to-the-truth theorem that is needed to defend Bayesian learning. The Bayesian agent of Elga's example is *not* humble about whether (with increasing probability) she will achieve asymptotic certainty for the limiting

frequency hypothesis in question: At each stage of her investigation, looking forward, she remains practically certain that her posterior probability will converge to the true limiting frequency hypothesis.

Second, the credal state P in Elga's example fails what we call Belot's Condition #1. P assigns probability 1 to a meager set of sequences of observations. Hence, though Elga argues that P is modest with respect to one limiting frequency hypothesis, according to Condition #1 P is immodest for a different but related hypothesis about existence of well-defined limiting frequencies.

Nonetheless, we agree with Elga (and with others who have argued the same point previously) that finitely but not countably additive probability models allow failures of the strengthened convergence results. We illustrate this point using a finitely additive probability, P', that is a simple variant of Elga's model P. But in our judgment, this phenomenon – where a finitely additive model P' *fails* the strengthened convergence-to-the-truth result – does not provide a satisfactory rebuttal to Belot's criticism. Belot's criticism, which is directed at countably additive credences, is that they display Bayesian orgulity. To argue that, on the contrary, the finitely additive probability P' assigns positive probability to a set of sequences where convergence to the truth fails, does not show that such a merely finitely additive probability is *reasonable*.

According to the rival standards for epistemic modesty that we offer in Section 4, such a finitely additive probability P' is *unreasonable* on two counts simultaneously: The Bayesian agent with credence P' knows in advance that *each* data sequence that might be observed will fail to induce certainty, both in the short term and in the limit. Also, P' fails the test for reasonableness based on consensus. That is, the agent with credences fixed by P' does not reach consensus with other members of a community of investigators who use countably additive credences and agree with P' about which (finite) sequences of observables are probability-0 events. But the others reach consensus among themselves.

For a detailed discussion of Elga's example, begin with a review of some relevant mathematical considerations. When probability P is defined for a measurable space, the principle of countable additivity has an equivalent form as a principle of *Continuity*. Let A_i ($i = 1, \dots$) be a monotone sequence of (measurable) events, where $\lim_i A_i = A$, also is a (measurable) event.

- *Continuity* $P(A) = \lim_i P(A_i)$.

When probabilities satisfy *Continuity*, the probabilities for a class $\mathcal{C}$ of events that form a field also determine uniquely the probabilities for the smallest σ-field generated by $\mathcal{C}$. (See Halmos [1950], Theorem 13A.) And if an event H belongs to that σ-field, then H can be approximated in probability by events from the field $\mathcal{C}$.   Specifically, for each $\varepsilon >$ 0 there exists a $C_\varepsilon \in \mathcal{C}$ such that $P([H-C_\varepsilon] \cup [C_\varepsilon-H]) < \varepsilon$. (See Halmos [1950], Theorem 13D.) This result has important consequences when H is a tail-field event in 2^ω.[13]

Consider the countably additive probability P for *iid* flips of a fair coin and, for example, the tail-field event $L^{\frac{1}{2}}$ in 2^ω. Then, $L^{\frac{1}{2}}$ can be approximated ever more precisely in probability by a sequence of finite-dimensional events $\{E_n: n = 1, \dots \}$, each of which is determined by a finite number of coordinates from the set of denumerable binary sequences, 2^ω. Choose a sequence, $\{\varepsilon_n > 0: n =1, \dots \}$, with $lim_n \varepsilon_n = 0$. That is, for each $n = 1, \dots$, $P([L^{\frac{1}{2}}- E_n] \cup [E_n - L^{\frac{1}{2}}]) < \varepsilon_n$ and each E_n depends upon only finitely many coordinates from 2^ω. With P the product measure for *iid* fair-coin flips and $L^{\frac{1}{2}}$ the tail-field event that is to be approximated, then the finite dimensional events E_n may be chosen as the set of sequences with relative frequency of 1's sufficiently close to ½ through the first n trials. However, when Continuity fails, and P is merely finitely

---

[13] An event T belongs to the *tail-field* of 2^ω provided that, when a point **x** belongs to T, so too does each point **x'** that differs in only finitely many coordinates from **x**. It is straightforward to verify that the set of tail-field events of 2^ω form a field.

additive but not countably additive, then the probabilities over $\mathcal{C}$ may fail to define the probabilities over the smallest σ-field generated by $\mathcal{C}$.

For example, pick two values $0 \leq p \neq q \leq 1$. A coherent, merely finitely additive probability $P^{p,q}$ on 2^ω may assign values to each finite-dimensional event according to *iid* trials with constant Bernoulli probability p, but assign probabilities to the tail-field events according to *iid* trials with constant Bernoulli probability q. Then, the strong law of large numbers does not entail the weak-law of large numbers with the same values. While *finite* sequences of 0s and 1s follow an *iid* Bernoulli-*p* product law, with $P^{p,q}$ probability 1, the tail-event of the limiting relative frequency for 1s is q. This phenomenon is at the heart of Elga's example.

Let P be a merely finitely additive probability on the Borel σ-algebra of 2^ω where $P(\cdot) = [P^{p,q}(\cdot) + P^{q,p}(\cdot)] / 2$. Elga considers the case with $p = 1/10$ and $q = 9/10$. This finitely additive probability assigns probability ½ to the tail-field event $L^{1/10}$ (the set of sequences with limiting frequency 1/10) and probability ½ to the tail-field event $L^{9/10}$ (the set of sequences with limiting frequency .9). For $x \in 2^\omega$, let $\mathbf{I}_L^{1/10}(x)$ be the indicator function for the event $L^{1/10}$ and $\mathbf{I}_L^{9/10}(x)$ the indicator function for the event $L^{9/10}$. So, P{x: $I_L^{1/10}(x) + I_L^{9/10}(x) = 1$} = 1. Thus, we see from the Proposition of Section 2, Elga's example stands in violation of *Topological Condition #1*, since with P-probability 1 the sequence of coin flips has a convergent limiting relative frequency. This forms a meager set of among the set of all binary sequences.

Elga asserts that the conditional probabilities associated with the (merely) finitely additive P-distribution fail the almost-sure strong-law convergence result. Here is the argument his offers for that conclusion. Let x be an element of the set $L^{1/10}$, a sequence with limiting relative frequency 1/10, which is practically certain to occur according to the P-distribution on sequences *if and only if* the $P^{9/10, \, 1/10}$ coin is flipped. (Otherwise, with P-probability 1, a sequence x almost surely has a limiting relative frequency 9/10,

since it is then following a $P^{1/10, \, 9/10}$ law.) Then, for each $\varepsilon > 0$ there exists integer

n_ε, such that for each $n > n_\varepsilon$, the observed sequence $\{X_1, \ldots, X_n\}$ has a relative frequency

of 1s close enough to 1/10 so that the posterior probability satisfies, $P(L^{9/10} \mid X_1, \ldots, X_n)$

$> 1\text{-}\varepsilon$.

This conditional probability assigns high probability to the event $L^{9/10}$ that the limiting

frequency of the sequence is .9 even though the sequence that generates the observations

in fact has limiting frequency 1/10. In this sense, the sequence of conditional

probabilities generated by x (an element of the set $L^{1/10}$) converge to the wrong tail-field

event, $L^{9/10}$, even though the sequence that generates the observations has limiting

frequency 1/10. Likewise, the convergence is to the wrong tail-field event, $L^{1/10}$, when

the sequence is generated by an element of the set $L^{9/10}$. Elga concludes that conditional

probabilities from this merely finitely additive P-model do not satisfy the (almost-sure)

strong-law convergence-to-the-truth results. Then, regarding either tail-field event $L^{1/10}$

or $L^{9/10}$, the agent with conditional credences fixed by probability P is both *open-minded*

and *modest*.[14] But this analysis is misleading regarding convergence-to-the-truth because

it conditions on P-null events, as we now explain.

Define the denumerable set of countably additive probabilities $\{P_n\}$ on 2^ω so that P_n is

the *iid* product of a Bernouilli-p probability for the first n-coordinates and is the *iid*

product of a Bernouilli-q probability for all coordinates beginning with the $n+1^{st}$ position.

Each P_n is a countably additive probability on the measurable space $\langle 2^\omega, \mathcal{B} \rangle$.

Distribution P_n has a **change point** after the nth trial. Let the change point, $N = n$, be

chosen according to a purely finitely additive probability, with $P(N = n) = 0$, $n = 1, 2, \ldots$.

Finally, let P be the induced (marginal) unconditional probability on the Borel σ-algebra

of sequences of coin-flips, $\langle 2^\omega, \mathcal{B} \rangle$.

---

[14] Regarding each of the two tail-field hypotheses, $L^{1/10}$ and $L^{9/10}$, the P-credences are *open-minded*, since $P(L^{1/10} \mid X_1, \ldots, X_n)$ may cross the 0.5 threshold, in either direction, as a function of finitely many future observations, $\{X_{n+1}, \ldots, X_{n+k}\}$. The P-credences are *modest*, since *ex ante*, given that the infinite sequence x belongs to $L^{1/10}$ (or to $L^{9/10}$) the agent assigns probability greater than 0 to the respective failure set of sequences.

As required for Elga's construction, this finitely additive probability P behaves as $P^{p,q}$. Its distribution is the *iid* product of a Bernoulli-*p* distribution on finite dimensional sets, and is the *iid* product of a Bernoulli-*q* distribution on the tail-field events.[15] P satisfies the weak-law of large numbers over finite sequences with Bernoulli parameter *p* and satisfies the strong-law of large numbers on the tail-field with Bernoulli parameter *q*. Hence, the strong-law does not entail the weak-law with the same parameter value.

Given an observed history, $h_j = \{X_1 = x_1, X_2 = x_2, \ldots, X_j = x_j\}$, the Bayesian agent in Elga's example assigns a purely finitely additive conditional probability to the distribution of the change point (N) so that, with conditional probability 1, the change point is arbitrarily far off in the future. For each finite history h_j and for each $k = 1, 2, \ldots$, $P(N > k \mid h_j) = 1$. An agent who uses Elga's finitely additive P-model precludes learning about the change point variable, N. That agent is *closed-minded* in the relevant sense that, no matter what she observes, she is certain that the change point lies in the yet-to-be-observed future.

So, whenever the agent observes a finite history of coin flips with observed relative frequency of heads near to 9/10, she has high posterior probability for the tail-field event $L^{1/10}$. Likewise, whenever the agent sees a finite history of coin flips with observed relative frequency of heads near to 1/10, she/he has high posterior probability for the tail-field event $L^{9/10}$. And since this agent is always sure, given each finite history h_j, that the change point (N) is in the distant future of the sequence of coin flips, she always assigns arbitrarily high posterior probability to **correctly** identifying the tail-field event between $L^{1/10}$ and $L^{9/10}$.

---

[15] Elga follows Rao and Rao (1983, pp. 39-40) using the technique of Banach limits to establish the existence of a finitely additive probability corresponding to the $P^{p,q}$ distribution on repeated flips of the coin, based on the set of countably additive probabilities $\{P_n\}$. The method we use here, where the change point N is incorporated explicitly as a random variable in the finitely additive joint probability model, generates all the $P^{p,q}$ distributions over sequences of repeated flips of a coin as may be obtained with Banach limits. However, in addition it provides the added machinery needed to assess the agent's conditional credences $P(N \mid X_1, \ldots, X_n)$, which reflects also the agent's opinion about whether the sequence of coin flips passed the change point. Elga's reasoning ignores the fact that, for each $n = 1, \ldots$, $P(N > n \mid X_1, \ldots, X_n) = 1$, which is what the alternative analysis makes salient.

For example, this agent assigns probability near 1 to observing indefinitely long finite histories that have observed relative frequencies that linger near 9/10 exactly when the sequence *x* has a limiting relative frequency of 1/10. This finitely additive credal state satisfies the conclusion of the finitely additive almost-sure convergence-to-the-truth result: Almost surely, given the ***observed*** histories from a sequence *x*, the conditional probabilities converge to the correct indicator for the tail-behavior of the relative frequencies in *x*.

Elga's analysis to the contrary is based on having the agent consider conditional probabilities, $P(L^{1/10} \mid h_n)$ at histories h_n that run *beyond* the change point. But with Elga's finitely additive probability P-model, the agent's credence is 0 of ever witnessing such a history. That is, Elga's argument, whose conclusion is that the agent's conditional probabilities converge to the wrong indicator function, requires the agent to condition on an event of P-probability 0, i.e., that she has made finitely many observations that go past the change point in the sequence. But, at *each* finite stage in the history of observations, this event is part of a P-null-event where a failure of the (finitely additive) *almost sure* convergence to the truth is excused. Where this case differs from the countably additive one is that with the merely finitely additive probability P, the countable union of all these infinitely many P-null events (namely, that that change point has been reached by the kth observation, k = 1, 2, …), is a certain event – since the change point is certain to arrive, eventually.

Apart from this peculiar merely finitely additive credal attitude that precludes learning about the change point N, there is something else unsettling about this Bayesian agent's finitely additive model for coin flips. Perhaps the following makes clearer what that problem is. Modify Elga's model to the finitely additive probability P' so that
$$P'(\cdot) = [P^{5/10,\ 1/10}(\cdot) + P^{5/10,\ 9/10}(\cdot)] / 2,$$
with the change point N chosen, just as before, by a purely finitely additive probability, $P(N = n) = 0$ for *n* = 1, 2, … . Then the strong-law result applies to tail-field events and, P'-almost surely, the limiting frequency for heads is either 1/10 or 9/10; also, just as in Elga's P-model. However, the two finitely additive coins, $P^{5/10,\ 1/10}$ and $P^{5/10,\ 9/10}$, assign

the same probability to each finite history of coin flips. Letting h_n denote a specific history of length n,

$$P^{5/10,\ 1/10}\ (h_n) = P^{5/10,\ 9/10}\ (h_n) = 2^{-n}.$$

But then $$P'(L^{1/10} \mid h_n) = P'(L^{9/10} \mid h_n) = \tfrac{1}{2} = P'(L^{1/10}) = P'(L^{9/10}),$$

for each possible history.   That is, contrary to the strengthened convergence-to-the-truth result, in this modified P'-model, the agent is completely certain that her posterior probability for either of the two tail-field hypotheses, $L^{1/10}$ or $L^{9/10}$, is stationary at the prior value 1/2. Under the growing finite histories from each infinite sequence of coin flips, the posterior probability moves neither towards 0 nor toward 1. Within the P'-model, surely there is no convergence to the truth about these two tail-field events given increasing evidence from coin-flipping.[16]

Evidently, one aspect of what is unsettling about these finitely additive coin models is that the observed sequence of flips is entirely uninformative about the change point variable, N.   No matter what the observed sequence, the agent's posterior distribution for N is her/his prior distribution for N, which is a purely finitely additive distribution assigning 0 probability to each possible integer value for N. It is not merely that this Bayesian agent cannot learn about the value of N from finite histories. Also, two such agents who have finitely additive coin models that disagree only on the tail-field parameter cannot use the shared evidence of the finite histories to induce a consensus about the tail-field events, since they are both certain that their shared evidence has yet to cross the change point. In the next section, we use these themes about certainty and consensus based on shared evidence to provide a different answer to Belot's question about what distinguishes *modest* from *immodest* credal states.

4. *On standards for epistemic modesty using asymptotic merging and consensus.*
In his [1877] paper, *The Fixation of Belief*, C.S.Peirce argues that sound methodology needs to defend a proposal for how to resolve interpersonal differences of scientific

---

[16] The P'-model does not contradict Savage's (1954) finitely additive weak-law result, which we reported in Section 1.   That is, the P'-model does not satisfy Savage's requirement that the rival statistical hypotheses, $P^{5/10,\ 1/10}(\cdot)$ and $P^{5/10,\ 9/10}(\cdot)$ have different likelihood functions given some P'-non-null data.

opinion. Peirce asserts that the scientific method for resolving such disputes wins over other rivals (e.g., *apriorism*, or the *method of tenacity*) by having the *Truth* (aka observable *Reality*) win out – by settling debates through an increasing sequence of observations from well designed experiments. With due irony, much of Peirce's proposal for letting *Reality* settle the intellectual dispute is embodied within personalist Bayesian methodology.[17] Here, we review some of those Bayesian resources regarding three aspects of *immodesty*.

One kind of epistemic *immodesty* is captured in a dogmatic credal state that is immune to revision from the pressures of new observations. Such a credal state is closed-minded. And a closely related, second kind of *immodesty* is that two rival dogmatic positions cannot find a resolution to their epistemic conflicts through shared observations. They are persistent in their closed-mindedness. These two suggest that a credal state can be assessed for epistemic *immodesty* according to three considerations:

      (i)   how large is the set of conjectures, and

      (ii)   how large is the community of rival opinions, and

      (iii)   for which sets of sequences of shared observations

does Bayesian conditionalization offer resolution to interpersonal credal conflicts by bringing the different opinions into a consensus regarding the truth. In other words, qualitative degrees of epistemic immodesty are revealed with these three considerations, which synthesize criteria of asymptotic consensus and certainty. We discuss this sense of "immodesty" in the remainder of this Section.

We use as our starting point an important result due to Blackwell and Dubins (1962) about countable additive probabilities. Let $<X, \mathcal{E}>$ be a measurable Borel product-space with the following structure. Consider a denumerable sequence of sets X_i (i = 1, …) each with an associated σ-field $\mathcal{E}_i$. Form the infinite Cartesian product $X = X_1 \times \dots$ of denumerable sequences $(x_1, \dots) = x \in X$, where $x_i \in X_i$. That is, each x_i is an atom of its

---

[17] The irony is, of course, that Peirce objected to *conceptualism* (aka personalist probabilities) because he thought that it inappropriately combined subjective and objective senses of "probability." See his [1878] *The Probability of Induction*.

algebra $\mathcal{B}_i$. In the usual fashion, let the measurable sets in $\mathcal{B}$ be the σ-field generated by the measurable rectangles.

> *Definition*: A *measurable rectangle* $(A_1 \text{ x } \dots) = A \in \mathcal{B}$ is one where $A_i \in \mathcal{B}_i$ and $A_i = X_i$ for all but finitely many i.

Blackwell and Dubins (1962) consider the idealized setting where two Bayesian agents have this same measurable space of possibilities, each with her/his own countably additive personal probability, creating the two measure spaces $<X, \mathcal{B}, P_1>$ and $<X, \mathcal{B}, P_2>$. Suppose that P_1 and P_2 agree on which measurable events have probability 0, and admit (countably additive) predictive distributions, $P_i(\cdot \mid X_1, \dots, X_n)$ $(i = 1, 2)$, for each finite history of possible observations.[18] In order to index how much these two are in probabilistic disagreement, Blackwell and Dubins adopt the total-variation distance.

Define $\rho(P_1(\cdot \mid X_1{=}x_1, \dots, X_n{=}x_n), P_2(\cdot \mid X_1{=}x_1, \dots, X_n{=}x_n)) =$

$$\sup_{E \in \mathcal{B}} \mid P_1(E \mid X_1{=}x_1, \dots, X_n{=}x_n) - P_2(E \mid X_1{=}x_1, \dots, X_n{=}x_n) \mid$$

The index ρ is one way to quantify the degree of consensus between the two agents who share the same history of observations, $(x_1, \dots, x_n)$. This index focuses on the greatest differences between the two agents' conditional probabilities.

Here is the related strong-law result about asymptotic consensus:
(***) [Blackwell and Dubins, 1962, Theorem 2] For $i = 1, 2$,

P_i-almost surely, $lim_{n \to \infty} \rho(P_1(\cdot \mid X_1{=}x_1, \dots, X_n{=}x_n), P_2(\cdot \mid X_1{=}x_1, \dots, X_n{=}x_n)) = 0.$

In words, the two agents are practically certain that with increasing shared evidence their conditional probabilities will merge, in the very strong sense that the greatest differences

---

[18] Blackwell and Dubins use the concept of predictive distributions to mean those that admit *regular conditional distributions* with respect to the subalgebra of rectangular events. (See Breiman, 1968, p. 77 for a definition of a regular conditional distribution.) For discussion of countably additive probabilities that do not admit regular conditional distributions see Seidenfeld, Schervish, and Kadane (2001, p. 1614, Corollary 1).

in their conditional opinions over all measurable events in $\mathcal{B}$ will diminish to 0. And they remain practically certain of this future development for each (non-null) observed history. Thus, this result supports a conclusion about idealized asymptotic consensus from idealized application of the Scientific Method that Peirce asserted he could not prove but only defend as having no equal.[19]

Since, for each event in the space $\mathcal{B}$, the familiar strong-law convergence-to-the truth result applies, separately, to each investigator's opinion, the added feature of merging allows a defense against the charge of individual "immodesty" by showing that two rival opinions come into agreement about the truth, almost surely, in the strong sense provided by the ρ-index. In the setting of the Blackwell-Dubins (1962) result, almost surely two such investigators agree that they can resolve all conflicts in their credal states over all

---

[19] In Section V of (1877) Peirce writes,

> To satisfy our doubts, therefore, it is necessary that a method should be found by which our beliefs may be determined by nothing human, but by some external permanency -- by something upon which our thinking has no effect. Some mystics imagine that they have such a method in a private inspiration from on high. But that is only a form of the method of tenacity, in which the conception of truth as something public is not yet developed. Our external permanency would not be external, in our sense, if it was restricted in its influence to one individual. It must be something which affects, or might affect, every man. And, though these affections are necessarily as various as are individual conditions, yet the method must be such that the ultimate conclusion of every man shall be the same. Such is the method of science. Its fundamental hypothesis, restated in more familiar language, is this: There are Real things, whose characters are entirely independent of our opinions about them; those Reals affect our senses according to regular laws, and, though our sensations are as different as are our relations to the objects, yet, by taking advantage of the laws of perception, we can ascertain by reasoning how things really and truly are; and any man, if he have sufficient experience and he reason enough about it, will be led to the one True conclusion. The new conception here involved is that of Reality. It may be asked how I know that there are any Reals. If this hypothesis is the sole support of my method of inquiry, my method of inquiry must not be used to support my hypothesis. The reply is this: 1. If investigation cannot be regarded as proving that there are Real things, it at least does not lead to a contrary conclusion; but the method and the conception on which it is based remain ever in harmony. No doubts of the method, therefore, necessarily arise from its practice, as is the case with all the others. 2. The feeling which gives rise to any method of fixing belief is a dissatisfaction at two repugnant propositions. But here already is a vague concession that there is some *one* thing which a proposition should represent. Nobody, therefore, can really doubt that there are Reals, for, if he did, doubt would not be a source of dissatisfaction. The hypothesis, therefore, is one which every mind admits. So that the social impulse does not cause men to doubt it. 3. Everybody uses the scientific method about a great many things, and only ceases to use it when he does not know how to apply it. 4. Experience of the method has not led us to doubt it, but, on the contrary, scientific investigation has had the most wonderful triumphs in the way of settling opinion. These afford the explanation of my not doubting the method or the hypothesis which it supposes; and not having any doubt, nor believing that anybody else whom I could influence has, it would be the merest babble for me to say more about it. If there be anybody with a living doubt upon the subject, let him consider it.

elements of $\mathcal{B}$, and have their posterior probabilities almost surely concentrate on true hypotheses, by sharing increasing finite histories of observations from a sequence $\boldsymbol{x}$. Thus, in fine Peircean style, they are *not* open-minded about the efficacy of the Scientific Method for creating consensus and certainty.

Schervish and Seidenfeld (1990, section 3) explore several variations on this theme by enlarging the set of rival credal states in order to consider larger communities than two investigators, and by relaxing the sense of merging (or consensus) that is induced by shared evidence from a common measurable space $\langle X, \mathcal{B} \rangle$. They show that, depending upon how large a set of different mutually absolutely continuous probabilities is considered, the character of the asymptotic merging varies. This is where topology plays a useful role in formalizing "immodesty."

Here, we summarize three of those results. Let $\mathcal{R}$ be the set of rival credences that conform, pairwise, to the Blackwell-Dubins conditions, above. Consider three increasing classes of such communities.

(1) If $\mathcal{R}$ is a subset of a convex set of rival credences whose extreme points are compact in the discrete topology, then all of $\mathcal{R}$ uniformly satisfies the Blackwell-Dubins merging result. That is, then merging in the sense of ρ occurs simultaneously over all of $\mathcal{R}$.

(2) If $\mathcal{R}$ is a subset of a convex set of rival credences whose extreme points are compact in the topology induced by ρ, then all that is assured is a weak-law merging. That is, if $\{P_n, Q_n\}$ is an arbitrary sequence of pairs from $\mathcal{R}$, and $R \in \mathcal{R}$ is an arbitrary credence from the set of rivals, then

$$\rho(\, P_n(\cdot | X_1, \ \ldots, X_n), \, Q_n(\cdot \ | X_1, \ \ldots, X_n)\,) \xrightarrow{R} 0.$$

(3) And if $\mathcal{R}$ is a subset of a convex set of rival credences whose extreme points are compact in the weak-star topology induced by ρ, then not even a weak-law merging of the kind reported in (2) is assured.

Not surprising then, as the community $\mathcal{R}$ increases its membership, the kind of consensus that is assured – the version of community-wide probabilistic merging that results from shared evidence – becomes weaker. So, one way to assess the epistemological "immodesty" of a credal state formulated with respect to a measurable space $<X, \mathcal{B}>$ is to identify the breadth of the community $\mathcal{R}$ of rival credal states that admits merging through increasing shared evidence from $\mathcal{B}$. For example, the agent who thinks each morning that it is highly probable that the world ends later that afternoon has an *immodest* attitude because there is only the isolated community of like-minded pessimists who can reconcile their views with commonplace evidence that is shared with the rest of us.

When the different opinions do not satisfy the requirement of mutual absolute continuity, the previous results do not apply directly. Instead, we modify an idea from Levi [1980, §13.5] so that different members of a community of investigators modify their individual credences (using convex combinations of rival credal states) in order *to give other views a hearing* and, in Peircean fashion, in order to allow increasing shared evidence to resolve those differences.

Let $I = \{i_1, \dots \}$ serve as a finite or countably infinite index set, and let $\mathcal{R} = \{P_i: i \in I\}$ represent a community of investigators, each with her/his own countably additive credence function P_i on a common measurable space $<X, \mathcal{B}>$. It may be that, pairwise, the elements of $\mathcal{R}$ are not even mutually absolutely continuous. In order to allow new evidence to resolve differences among the investigators' credences for elements of $\mathcal{B}$ (rather than trying, e.g., to preserve common judgments of conditional credal independence between pairs of elements of $\mathcal{B}$) each member of $\mathcal{R}$ shifts to a credal state by taking a mixture of each of the investigators' credal states: a "linear pooling" of those states. Specifically, for each $i \in I$, let $\tilde{\alpha} = \{\alpha_{ij}: \alpha_{ij} > 0, \sum_{j=1}^{\infty} \alpha_{ij} = 1\}$ serve as a set of weights that investigator$_i$ uses to create the credal state $Q_i = \sum_{j=1}^{\infty} \alpha_{ij} P_i$ to replace P_i. It might be that for each $i \in I$, each Q_i is *self-centered* in the following sense. Let $\varepsilon > 0$. The Q_i might be *self-centered* in that $\alpha_{ii} \geq 1-\varepsilon$. Then, pairwise, the Q_i satisfy the

assumptions for the Blackwell-Dubins' result (***) despite being self-centered. Depending upon the size of the community $\mathcal{R}$, using the replacement credal states $\{Q_i\}$ results (1), (2), and (3) obtain.

We conclude this discussion of probabilistic merging with a reminder that merely finitely additive probability models open the door to reasoning to a foregone conclusion, Kadane Schervish, and Seidenfeld [1996], in a different sharp contrast with the P' model above to the almost sure asymptotic merging and convergence-to-the-truth results associated with countably additive probability models. Key to these asymptotic results for countably additive probabilities is the *Law of Iterated Expectations*.

Let *X* and *Y* be (bounded) random variables measurable with respect to a countably additive measure space $<\Omega, \mathcal{B}, P>$. With $\mathcal{E}[X]$ and $\mathcal{E}[X \mid Y = y]$ denoting, respectively, the expectation of *X* and the conditional expectation of *X*, given the event $Y = y$, then

Law of Iterated Expectations $\qquad \mathcal{E}[X] = \mathcal{E}[\ \mathcal{E}[X \mid Y]\ ].$

As Schervish et al. established [1984], each merely finitely (and not countably) additive probability defined on a σ-field of sets fails this law even when the variable *X* is an indicator variable. That is, each merely finitely additive probability fails to be *conglomerable* in some denumerable partition, here associated with the random quantity *Y*. Specifically, with a merely finitely additive probability P, there exists a measurable hypothesis *H* and denumerable partition of measurable events $\pi = \{E_i : i = 1, \ldots\}$ where

$$P(H) \ < \ infimum_{E_i \in \pi} P(H \mid E_i).$$

Then, contrary to the *Law of Iterated Expectations*, with expectations $\mathcal{E}$ over all $E_i \in \pi$,

$$P(H) < \mathcal{E}[\ P[H \mid E_i \in \pi]\ ].$$

Let the random variable Y have a denumerable sample space, $Y = \{y_1, y_2, \ldots\}$. Associate the event E_i with the outcome $Y = y_i$. Then if P is non-conglomerable for H in the partition generated by *Y*, in the partition π_Y, then P fails the Law of Iterated Expectations in π_Y.

A set of such merely finitely additive probabilities, each of which is non-conglomerable in the same partition of the shared evidence, can display *reasoning to contrary foregone conclusions* with both in the short run in asymptotically with increasing shared evidence. Because the investigators' conditional probabilities for a pair of contrary hypotheses {H_1, H_2} are non-conglomerable in the partitions of their increasing shared evidence, each investigator may become increasingly certain of a different hypothesis as a function solely of the sample *size* of their shared evidence, regardless what those samples reveal. Moreover, this assured increasing divergence in their updated opinions is a fact they are aware of *ex ante*.

The lesson we draw is this. Bayesian agents who use merely finitely additive probabilities face a trade-off between:

- the added flexibility in modeling that comes with relaxing the constraint of countable additivity

versus

- the added restrictions on the kinds of shared evidence necessary to achieve the desirable methodological laws about asymptotic consensus and certainty illustrated in the countably additive strong-laws.

5. *Summary*. Savage (1954) and Blackwell-Dubins (1962) offer important results showing that Bayesian methodology uses increasing shared evidence in order to temper and to resolve interpersonal disagreements about personal probabilities. We contrast interpersonal standards of asymptotic consensus about certainties that arise from a sequence of shared evidence with Belot's (2013) proposal to use a topological standard of "meagerness" in order to determine when a credal state is immodest, based on a stand-alone assessment of that credal state.

We understand Belot to endorse Topological Condition #1, which requires that comeager sets are assigned positive probability. Where a probability model treats a comeager set as null, that shows the model is *immodest* because it dismisses a topological large set as probabilistically negligble. But, in the light of the fact that the set of sequences whose

frequencies oscillate maximally is comeager, we see that all the familiar probability models violate Condition #1. We believe that, also, Belot endorses Condition #2, which requires that a *typical* set of sequences should receive a *typical* probability, i.e., a meager set should be assigned probability 0. This topological standard entails extreme *a priori* credences about the behavior of observed relative frequencies. Condition #2 mandates that, with probability 1, observed frequencies oscillate maximally in order to avoid being contained in a meager set. This creates its own kind of dogmatism since (almost surely) the conditional probability from this model persists in assigning conditional probability 1 to the hypothesis that observed frequencies oscillate maximally.

In contrast with Belot's approach, in Section 4 we outline a different strategy for assessing epistemic *modesty/immodesty*, based on considerations of both asymptotic certainty and consensus among investigators who share evidence. Belot's strategy is to impose additional requirements that, in the spirit of coherence, apply to a stand-alone credence function. We follow, e.g., Peirce in requiring that sound Scientific methodology provides investigators with the resources to resolve interpersonal disagreements through shared evidence. This consideration allows for results about conditions for asymptotic consensus among a set of investigators to serve also as a standard for their epistemic modesty regarding interpersonal disagreements.

As a separate issue, in Section 3 we discuss Elga's (2016) reply to Belot's analysis. Elga focuses on the assumption of countably additivity in the strengthened convergence results. His rebuttal to Belot's analysis uses a merely finitely additive P probability to illustrate that merely finitely additive conditional probabilities need not satisfy the countably additive asymptotic (strong-law) convergence results. These are the results that Belot argues reveals an *immodesty* in the countably additive Bayesian methodology.

We agree with Elga (and as has been argued before) that the asymptotics of merely finitely additive conditional probabilities are different in kind from those of countably additive conditional probabilities. But we do not agree with Elga about which are the relevant asymptotic results in his P-model for assessing Bayesian learning of limiting

frequencies. In addition, the P-model fails Condition #1, which we understand is one of Belot's standards for *modesty*.

As we illustrate in Section 3, the conditional probabilities arising from a different (but related) merely finitely additive probability P' fail the asymptotic certainty and consensus results that follow when either Savage's or Blackwell-Dubins' analysis applies. We argue that the added generality afforded by merely finitely additive probabilities over countably additive probabilities carry an extra price if they are to be used reasonably. They require more restrictive conditions than do countably additive probabilities, if the sequence of conditional probabilities that arise from an increasing sequence of shared evidence is to resolve interpersonal credal disagreements.

Appendix

In his classic discussion of measure and category, Oxtoby [1980: Theorem 1.6 (p. 4) and Theorem 16.5 (p. 64)] establishes that, quite generally, a topological space that also carries a Borel measure can be partitioned into two sets: One is a measure 0 set and the other, which is its complement, is a meager set. Here we show (*Theorem A1*) that this tension between probabilistic and topological senses of being a "small" set generalizes to sequences of random variables relative to a large class of infinite product topologies. We follow that result with a corollary, namely, *Proposition* (**) in the main text is an instance of *Theorem A1* for binary sequences.

Let χ be a set with topology $\Im$ and Borel σ-field, $\mathcal{B}$, i.e., the σ-field generated by the open sets in $\Im$. Let χ^{∞} be the countable product set with the product topology $\Im^{\infty}$ and product σ-field, $\mathcal{B}^{\infty}$, which is also the Borel σ-field for the product topology (because it is a countable product). Let $<\Omega, \mathcal{A}, \mathrm{P}>$ be a probability space, and let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random quantities such that, for each n, $X_n: \Omega \to \chi$ is $\mathcal{A}$ and $\mathcal{B}$ measurable. Define X: $\Omega \to \chi^{\infty}$ by $X(\omega) = <X_1(\omega), X_2(\omega), \ldots>$. Let $\mathcal{S}_X = X(\Omega)$ be the image of X, i.e., the set of

sample paths of X. We denote elements of $\mathcal{S}_X$ as $y = <y_1, y_2, \ldots>$. $\mathcal{S}_X$ is a subset of χ^∞.

Therefore, we endow $\mathcal{S}_X$ with the subspace topology. In the remainder of this presentation, we identify certain subsets of $\mathcal{S}_X$ as being either meager or comeager. These results depend solely on the topology for the measurable space $<\Omega, \mathcal{A}>$, and not on the probability P. However, the probability P is needed in order to display the tension between the two, rival senses of being a "small" set.

In what follows we require a degree of "logical independence" between the X_n's. In particular, we need the sequence $\{X_n\}_{n=1}^\infty$ to be capable of moving to various places in χ^∞ regardless of where it has been so far.

***Condition* A**: Specifically, for each j, let $B_j \in \mathcal{B}$ be a set such that B_j has nonempty interior B_j^o. Assume that for each n, for each $x = <x_1, \ldots, x_n> \in <X_1, \ldots, X_n>(\Omega)$, and for each j, there exists a positive integer $c(n, j, x)$ such that $<X_1, \ldots, X_n, X_{n+c(n, j, x)}>^{-1}(\{x\} \times B_j^o) \neq \varnothing$.

Condition **A** asserts that, no matter where the sequence of random variables has been up to time n, there is a finite time, $c(n, j, x)$, after which it is possible that the sequence reaches the set B_j^o. For example, suppose that each X_n is the average of the first n in a sequence of Bernoulli random variables and that $\{\varepsilon_j\}_{j=1}^\infty$ is a sequence of positive real numbers whose limit is 0. If $B_j = (0, \varepsilon_j)$ for even j and $B_j = (1-\varepsilon_j, 1)$ for odd j then, independent of the particular sequence x, the longest we would have to wait to reach B_j is

$$c_{n,j} \;=\; \frac{n(1 - \varepsilon_j)}{\varepsilon_j} \;+\; 1$$

in order to be sure that there is a sample path that takes us from an arbitrary initial sample path of length n to B_j by time $n+c_{n,j}$. Thus, $c_{n,j}$ is a *worst case* bound for waiting. For some $x = <x_1, \ldots, x_n>$, the minimum $c(n, j, x)$ might be much smaller than this $c_{n,j}$. For instance, with jointly continuous random variables with strictly positive joint density in which $<X_1, \ldots, X_n>(\Omega) = \chi^n$ for all n, then $c(n, j, x) = 1$ for all n, j, and x.

For each $y \in \mathbf{S}_X$, define $\tau_0(y) = 0$, and for $j > 0$, define

$$\tau_j(y) = \begin{cases} min \, \{n > \tau_{j-1}(y): y_n \in B_j\}, & \text{if the minimum is finite,} \\ \\ \infty & \text{if not.} \end{cases}$$

Let $B = \{y \in \mathbf{S}_X : \tau_j(y) < \infty$ for all j$\}$, and let $A = \mathbf{S}_X \setminus B = B^c \cap \mathbf{S}_X$.

Note that A is the set of sample paths each of which fails to visit at least one of the B_j sets in the order specified. Because we do not require that the sets B_j are nested, it is possible that the sequence reaches B_k for all $k > j$ without ever reaching B_j. Or the sequence could reach B_j before reaching B_{j-1} but not after.

Theorem A1: A is a meager set.

Proof: Write $A = \cup_j C_j$, where $C_j = \{y \in \mathbf{S}_X: \tau_j(y) = \infty\}$. Then A is meager if and only if C_j is meager for every j. We prove that C_j is meager for every j by induction.

Start with $j = 1$. We have $\tau_j(y) = \infty$ *if and only if* $y \in C_1 = \cap_{n=1}^{\infty} \{y \in \mathbf{S}_X: y_n \in B_1^c\}$. To see that C_1 is meager, notice that $C_1^c = \cup_{n=1}^{\infty} D_n$, where

$$D_1 = \mathbf{S}_X \cap (B_1 \times \chi^{\infty}),$$

and for $n > 1$, $\qquad D_n = \mathbf{S}_X \cap (\chi^{n-1} \times B_1 \times \chi^{\infty}).$

Each D_n contains a nonempty sub-basic open set O_n obtained by replacing B_1 in the definition of each D_n by its interior B_1^o. So C_1^c contains the nonempty open set $O = \cup_{n=1}^{\infty} O_n$.

Next, we show that O is dense; hence, C_1 is meager as it is nowhere dense. We verify that $O \cap E \neq \varnothing$ for every nonempty basic open set E. If E is a nonempty basic open set, then there exists an integer k and there exist nonempty open subsets $E_1, \ldots, E_k$ of χ such that

$$E = \mathbf{S}_X \cap (E_1 \times \ldots \times E_k \times \chi^{\infty}).$$

Let $y \in E$, and let x_k be the first k coordinates of y. Then there exist points in $\boldsymbol{S}_X$ whose first k coordinates are x_k and whose $k + c(k, 1, x_k)$ coordinate lies in B_1^o. Hence,

$$O \cap E \supseteq \boldsymbol{S}_X \cap (E_1 \times \ldots \times E_k \times \chi^{c(k,1,x_k)-1} \times B_1^o \times \chi^\infty) \neq \varnothing.$$

Next, for $j > 1$, assume that C_r is meager for all $r < j$. To complete the induction, we show that C_j is meager, which follows the same reasoning as in the base case. Write

$$C_j = C_{j-1} \cup_{r=j-1}^\infty F_r$$

Where $\qquad F_r = \{y \in \boldsymbol{S}_X : \tau_{j-1}(y) = r \text{ and } y_n \in B_j^{\boldsymbol{c}} \text{ for all } n > r\}.$

It suffices to show that each F_r is meager.

Notice that F_r is a subset of

$$G_r = \{y \in \boldsymbol{S}_X : y_r \in B_j\} \cap \{y : \text{for all } n > r, y_n \in B_j^{\boldsymbol{c}}\}$$

It suffices to show that G_r is meager.

As in the case $j = 1$, write $\qquad G_r^{\boldsymbol{c}} = \{y \in \boldsymbol{S}_X : y_r \in B_r^{\boldsymbol{c}}\} \cup \cup_{n=r+1}^\infty D_n$

where, $\qquad D_n = \boldsymbol{S}_X \cap (\chi^{n-1} \times B_j \times \chi^\infty).$

Each D_n contains a nonempty sub-basic open set O_n obtained by replacing each B_j in the definition of each D_n by its interior B_j^o. So $G_r^{\boldsymbol{c}}$ contains a nonempty open set

$O = \cup_{n=r+1}^\infty O_n.$

Last, we establish that O is dense; hence, G_r is meager. Reason as in the base case $j = 1$. We verify that $O \cap E \neq \varnothing$ for every nonempty basic open set E. If E is a nonempty basic open set, then there exists an integer k and there exist nonempty open subsets $E_1, \ldots, E_k$ of χ such that $E = \boldsymbol{S}_X \cap (E_1 \times \ldots \times E_k \times \chi^\infty)$. Let $y \in E$, and let x_k be the first k coordinates of y. Then there exist points in $\boldsymbol{S}_X$ whose first k coordinates are x_k and whose $k + c(k, j, x_k)$ coordinate lies in B_j^o. Hence,

$$O \cap E \supseteq \boldsymbol{S}_X \cap (E_1 \times \ldots \times E_k \times \chi^{c(k,j,x_k)-1} \times B_j^o \times \chi^\infty) \neq \varnothing,$$

which completes the induction. $\diamond_{T.\,A1}$

Next, return to consider the sequence $\{X_n\}_{n=1}^{\infty}$ of random variables described earlier. Suppose that each X_n is the sample average of some other sequence of random variables. That is, $X_n = \frac{1}{n}\sum_{k=1}^{n} Y_k$, where each Y_k is finite. Assume that *Condition* **A** obtains. Namely, assume that the dependence between the Y_k is small enough so that $c(n, j, x_k) < \infty$, for all n, j, x_k. For example, assume that there exist $c < d$ with c either finite or $c = -\infty$, and either with d finite or $d = \infty$, such that for each $j > 1$ and each $y \in <Y_1, \ldots, Y_{j-1}>(\Omega)$,

$$sup_{\omega \in A_y} Y_j(\omega) = d \text{ and } inf_{\omega \in A_y} Y_j(\omega) = c,$$

where $\qquad\qquad A_y = \{\omega: <Y_1(\omega), \ldots, Y_{j-1}(\omega)> = y\}.$

Then *Condition* **A** obtains for a sequence of *iid* random variables. It also obtains for a sequence of random variables such that $\{Y_1, \ldots, Y_n\}$ has strictly positive joint density over $(c, d)^n$ for all n. In such a case, we could let

$$B_j \;=\; \begin{cases} [c, a] & \text{if c is finite,} \\ [-\infty, a] & \text{if } c = -\infty \end{cases}$$

where $c < a < d$. Then A contains all sample paths for which *lim inf*$_n$ $X_n > a$ along with some sample paths for which *lim inf*$_n$ $X_n = a$. If we repeat the construction of A for a countable collection of a_n with $a_n \downarrow c$, then the union of all of the A sets is meager. Then, the set of sample paths for which the *lim inf*$_n$ $X_n > c$ is meager. A similar construction shows that the set of sample paths for which *lim sup*$_n$ $X_n < d$ is meager. Hence the union of these last two sets is meager, and the sequence of sample paths along which X_n oscillates maximally is a comeager set.

Theorem A1 applies directly to the sequence $\{X_n\}_{n=1}^{\infty}$. It shows that certain sets of sample paths of this sequence are meager or comeager. If, as in the case of sample averages, each X_n is a function of $\{Y_1, \ldots, Y_n\}$, we can evaluate the category of a set of sample paths of the $\{Y_n\}_{n=1}^{\infty}$ sequence. If $<X_1, X_2, \ldots>$ is a bicontinuous function of $<Y_1, Y_2, \ldots>$, then the two sets of sample paths are homeomorphic. In particular, this implies that the category of a set of sample paths of one sequence will be the same as the category of the

corresponding set of sample paths of the other sequence: the one is meager if and only if the other is.

In the case of sample averages, we can exhibit the bicontinuous function explicitly. To be specific, let $\chi = \Re$, and for each n, define $X_n = \frac{1}{n}\sum_{k=1}^{n} Y_k$. Let $X = <X_1, X_2, \ldots>$ as above, with $Y = <Y_1, Y_2, \ldots>$. Let $\boldsymbol{S}_X$ and $\boldsymbol{S}_Y$ be the sets of sample paths of X and Y, respectively. That is, $\boldsymbol{S}_X = X(\Omega)$ and $\boldsymbol{S}_Y = Y(\Omega)$. For each $y \in \boldsymbol{S}_Y$, define

$$\phi(y) = (\ldots, \frac{1}{n}\sum_{k=1}^{n} y_k, \ldots).$$

For each $x \in \boldsymbol{S}_X$, define

$$\varphi(x) = (x_1, 2x_2 - x_1, \ldots, nx_n - (n\text{-}1)x_{n-1}, \ldots).$$

Then, by construction, $\phi(Y) = X$ and $\varphi(X) = Y$. That is $\varphi: \boldsymbol{S}_X \to \boldsymbol{S}_Y$ is the inverse of $\phi: \boldsymbol{S}_Y \to \boldsymbol{S}_X$. In order to have the category of the two sample paths to be the same, it is sufficient that both ϕ and φ are continuous. If they are continuous as functions both from and to $\Re^{\infty}$, then they will be continuous in their subspace topologies. It suffices to show that $\phi^{-1}(B)$ and $\varphi^{-1}(B)$ are open for each sub-basic open set B. Every sub-basic open set is of the form $B = \prod_{n=1}^{\infty} B_n$ where each $B_n = \Re$ except for one value of $n = n_0$ for which B_{n_0} is open as a subset of $\Re$. Then each of $\phi^{-1}(B)$ and $\varphi^{-1}(B)$ has the form $C \times \Re^{\infty}$ where C is a n_0-dimensional open subset of $\Re^{n_0}$; hence both sets are open, and we have that $\boldsymbol{S}_X$ is homeomorphic to $\boldsymbol{S}_Y$.

The *Proposition* (**) in the main text is an instance of *Theorem A1* for binary sequences.

References

Belot, G. (2013) *Bayesian Orgulity*. Phil. Sci. 80: 483-503.

Billingsley, P (1986) *Probability and Measure*, 2nd ed. New York: John Wiley.

Blackwell, D. and Dubins, L. (1962) *Merging of opinions with increasing information*. Ann. Math. Stat. **33**: 882-887.

Breiman, L. (1968) P*robability*. Reading, MA: Addison-Wesley,

Calude, C. and Zamfirescu, T. (1999) *Most numbers obey no probability laws. Publ. Math. Debrecen* 54: 619-623.

Dubins, L, and Savage, L.J. (1965) *How to Gamble if You Must: Inequalities for Stochastic Processes.* New York: McGraw-Hill.

Edwards, W., Lindman, H., and Savage, L.J. (1963) *Bayesian statistical inference for psychological research.* Psych. Rev. 70: 193-242.

de Finetti, B. (1937) *Foresight: Its logical laws, its subjective sources*. In H.E.Kyburg, Jr. and H.Smokler (eds.) *Studies in Subjective Probability,* 1964. New York: John Wiley, pp. 93-158. (Translated from the French version by H.E.Kyburg, Jr.)

Dubins, L.E. (1975) *Finitely additive conditional probabilities, conglomerability, and disintegrations*. Ann. Probability **3**: 89-99.

Elga, A. (2016) *Bayesian Humility*. Phil. Sci. 83: 305-323.

Kadane, J.B., Schervish, M.J. and Seidenfeld, T. (1986) *Statistical Implications of Finitely Additive Probability*. In P.K.Goel and A. Zellner (eds.) *Bayesian Inference and Decision Techniques*. Elsevier: Amsterdam, pp. 59-76.

Kadane, J.B., Schervish, M.J., and Seidenfeld, T. (1996) *Reasoning to a Foregone Conclusion*. JASA **91**: 1228-1235.

Levi, I. (1980) *The Enterprise of Knowledge*. Cambridge: MIT Press.

Oxtoby, J.C. (1957) *The Banach-Mazur Game and Banach Category Theorem*, in Dresher, M., Tucker, A. W., & Wolfe, P. *Contributions to the Theory of Games* (Vol. 3). Princeton University Press: pp. 159-163

Oxtoby, J.C. (1980) *Measure and Category* (2nd edition). Springer-Verlag.

Peirce, C.S (1877) *The Fixation of Belief*. Popular Science Monthly.

Peirce, C.S. (1878) *The Probability of Induction*. Popular Science Monthly.

Savage, L.J. (1954) *The Foundations of Statistics*. New York: J. Wiley Pub.

Schervish, M.J., Seidenfeld, T., and Kadane, J.B. (1984) *The Extent of Non-conglomerability of finitely additive probabililties*. Z.Wahr. 66: 205-226.

Schervish, M.J. and Seidenfeld, T. (1990) *An approach to certainty and consensus with increasing evidence*. J. Statistical Planning and Inference **25**: 401-414.

Seidenfeld, T., Schervish, M.J., and Kadane, J.B. (2001) *Improper Regular Conditional Distributions*. Ann. Probability **29**: 1612-1624.