Misspecified Bayesianism

Pooya Molavi

Northwestern University

August 2025

Abstract

An agent is a misspecified Bayesian if she updates her belief using Bayes' rule given a subjective, possibly misspecified model of her signals. This paper shows that a belief sequence is consistent with misspecified Bayesianism if the prior *contains a grain* of the average posterior, i.e., is a mixture of the average posterior and another distribution. A partition-based variant of the grain condition is both necessary and sufficient. Under correct specification, the grain condition reduces to the usual Bayes plausibility. The condition imposes essentially no restriction on the posterior given a full-support prior over a finite or compact state space. However, it rules out posteriors that have heavier tails than the prior on unbounded state spaces. The results cast doubt on the feasibility of testing Bayesian updating in many environments. They also suggest that many seemingly non-Bayesian updating rules are observationally equivalent to Bayesian updating under misspecified beliefs.

Acknowledgments: I am grateful to Alexander Jakobsen, Peter Klibanoff, Alessandro Pavan, Alvaro Sandroni, Eran Shmaya, Alireza Tahbaz-Salehi, and participants at the 2025 ASSA meeting for their valuable comments. An earlier version of this paper was distributed under the title "The Empirical Content of Bayesianism."

Email: pmolavi@kellogg.northwestern.edu

1 Introduction

Following the treatise of Savage (1972), the Bayesian theory of probability has become the dominant paradigm in the modeling of decision making under uncertainty. This paradigm's dominance in economics is not unwarranted. It allows one to assign probabilities to unique or rare events; it has an elegant foundation in the study of rational choice under uncertainty; and it is appealing from a normative perspective—as Epstein and Le Breton (1993) proclaim, "dynamically consistent beliefs must be Bayesian."

Whether Bayesian updating is an accurate positive model of how individuals actually revise their views is a different matter. A large body of evidence documents violations of Bayesian updating (e.g., Coibion and Gorodnichenko 2015; Bordalo, Gennaioli, Ma, and Shleifer 2020). However, standard tests of Bayes' rule are joint tests of Bayesian updating and the assumption that agents have correctly specified models of the data-generating process. Therefore, any rejection by those tests could be due to non-Bayesian updating, misspecification, or a combination thereof. A natural question is then what restrictions (if any) Bayesian updating imposes on the dynamics of beliefs in isolation.

This paper provides an answer. An agent is a *misspecified Bayesian* if she updates her belief using Bayes' rule given an internally consistent but possibly misspecified model of the data she observes. The paper shows that a belief sequence is consistent with misspecified Bayesianism if the prior contains a "grain" of the average posterior. This condition is a weakening of the Bayes plausibility condition (Kamenica and Gentzkow, 2011) that characterizes the behavior of a "correctly specified Bayesian." While Bayes plausibility requires the prior to equal the average posterior, the grain condition requires the prior to be a mixture of the average posterior with some other probability distribution. The paper also shows that a slightly more permissive, partition-based grain condition is both necessary and sufficient for consistency with misspecified Bayesianism.

The strength of the grain condition is highly dependent on the environment. When the state space is finite, misspecified Bayesianism imposes only a support inclusion requirement: The support of the average posterior must be contained in the support of the prior. If the prior has full support on a finite space, any distribution of posteriors can be rationalized. When the state space is compact and the prior and average posterior admit continuous densities, the same support restriction is again both necessary and sufficient (under mild regularity). The picture changes sharply if the state space is unbounded. There the grain condition rules out any set of posteriors whose tails are uniformly heavier than those of the prior. In fact, a single heavier-tailed posterior that is realized with positive probability suffices for rejection.

The paper's characterization is robust to several natural extensions. First, when the distribution of posteriors is observed across different realizations of the state, the same partition-based grain condition continues to characterize misspecified Bayesianism once one averages the distribution of posteriors across states. Second, if signals are observed alongside posteriors, the construction that rationalizes the belief sequence also rationalizes the joint distribution of signals and posteriors. Third, with more than two periods, consistency reduces to a local requirement: for almost every current belief, the next-period distribution of beliefs must satisfy the grain condition relative to that belief. Finally, with time-varying states, one can treat the whole state path as a single object, so the baseline result applies without further modification.

These findings may appear at odds with the existing results in the literature. Kamenica and Gentzkow (2011) argue that Bayesian updating requires the average posterior to equal the prior. Shmaya and Yariv (2016) argue that any belief sequence in which the prior is in the relative interior of the convex hull of posteriors is consistent with agents' use of Bayes' rule. This paper presents two additional theorems that clarify the relationship between these different conditions. The theorems adapt the existing results to general state spaces, thus making them directly comparable to this paper's results. They demonstrate that the earlier results characterize Bayesian updating only under additional assumptions on agents' subjective models. Kamenica and Gentzkow (2011) do so by requiring agents to have correct beliefs about the distribution of signals, whereas Shmaya and Yariv (2016) require the subjective belief to have the same support as the true distribution.

The paper's results have two important implications. First, they suggest that Bayesian updating is essentially unfalsifiable when priors are supported over finite or compact spaces. The only way to test Bayesian updating without invoking extra assumptions about agents' models is by comparing the tails of the prior and posterior over unbounded state spaces. Second, many non-Bayesian updating rules turn out to be observationally equivalent to Bayesian updating with misspecified models. For example, agents with diagnostic expectations (Bordalo, Gennaioli, and Shleifer, 2018) overweight representative states and violate Bayes' rule at face value. Yet they behave *as if* they were Bayesians who believed the signal was less noisy and its noise term was negatively correlated with the state.

The paper thus bridges two strands of work on departures from rational expectations. The first one preserves Bayesian updating but allows misspecification, e.g., Esponda and Pouzo (2016, 2021), Bohren (2016), Frick, Iijima, and Ishii (2020), Fudenberg, Lanzani, and Strack (2021), and Molavi, Tahbaz-Salehi, and Vedolin (2024, 2025). The second strand posits non-Bayesian heuristics, e.g., Tversky and Kahneman (1974), Rabin and Schrag (1999), Epstein, Noor, and Sandroni (2010), and Cripps (2019)—see Ortoleva (2022) for a recent survey. The results suggest that these two types of deviation are empirically hard to disentangle without information such as agents' forecasts of their own future beliefs. In contemporaneous work, Bohren and Hauser (2023) study the question of when non-Bayesian updating rules can be represented as misspecification. They extend Shmaya and Yariv (2016)'s analysis by making agents' forecasts of their future beliefs observable and deriving necessary and sufficient conditions for an updating rule and a forecast to have a misspecified-model representation. In contrast, this paper's focus is characterizing misspecified Bayesianism absent information on agents' beliefs about how they will update their beliefs.

The grain condition is borrowed from the literature on merging of opinions. The merging literature is concerned with the question of whether Bayesian learning can lead agents to forecast the future accurately (i.e., merge to truth) or to play Nash equilibrium strategies. Blackwell and Dubins (1962) show that the absolute continuity of the prior with respect to the true distribution is sufficient to ensure merging. Kalai and Lehrer (1993) introduce the notion of containing a "grain of truth" as a stronger absolute continuity notion that guarantees convergence to an approximate Nash equilibrium in repeated games of incomplete information. While the merging literature studies the long-run convergence of Bayesian learners, this paper's focus is on the finite-horizon consistency of belief sequences with Bayesian updating under potentially misspecified models.

The remainder of the paper is organized as follows: Section 2 introduces the paper's conceptual framework and formally defines misspecified Bayesianism. Section 3 presents the main result and examines several special cases. Section 4 illustrates the theoretical results of the paper in the context of three examples. Section 5 discusses several extensions of the results. Section 6 discusses the relationship to Bayes plausibility and another related notion in the literature. Section 7 discusses the implications of the results. The proofs are relegated to the appendix.

2 Conceptual Framework

This section introduces the paper's conceptual framework and defines what it means for a belief sequence to be consistent with misspecified Bayesianism.

2.1 Setup

There is a fixed state of the world. The state is denoted by x and belongs to a complete separable metric space X.[1]

The main objects of interest are a prior and a "posterior" about x.[2] The prior is a probability distribution over X denoted by $\mu_0 \in \Delta(X)$. The posterior is a random variable $\mu_1 \in \Delta(X)$ with distribution $F_1 \in \Delta(\Delta(X))$. The randomness of the posterior is due to its dependence on a random signal. I denote the signal by s and assume that it belongs to the complete separable metric space S.[3] The *true* distribution of signals given the fixed state of the world is denoted by $\mathbb{P}_S \in \Delta(S)$.

2.2 Misspecified Bayesianism

The paper's goal is to characterize the conditions under which we can interpret the pair (μ_0, F_1) as arising from Bayesian updating given some possibly misspecified subjective model of how the signal is generated. The first step is to give a definition of Bayesianism for general state spaces. One standard definition is as follows. Bayesianism requires that (i) there exists a well-defined subjective distribution over the set of state-signal pairs; (ii)

[1] I endow X with the Borel σ-algebra $\mathcal{X} \equiv \mathcal{B}(X)$ and use $\Delta(X)$ to denote the set of probability distributions over $(X, \mathcal{B}(X))$. Since X is a complete separable metric space, $\Delta(X)$ is a complete separable metric space under the weak topology. I let $\Delta(\Delta(X))$ denote the set of probability distributions over $(\Delta(X), \mathcal{B}(\Delta(X)))$, where $\mathcal{B}(\Delta(X))$ denotes the Borel σ-algebra of $\Delta(X)$.

[2] I use the term "posterior" to refer to any probability distribution obtained by updating the prior after observing new information, regardless of whether it is derived from the prior via Bayes' rule—hence the quotation marks.

[3] I endow S with the Borel σ-algebra $\mathcal{S} \equiv \mathcal{B}(S)$ and use $\Delta(S)$ to denote the set of probability distributions over $(S, \mathcal{B}(S))$.

the subjective distribution assigns positive probability to signals that occur with positive probability under the true distribution; and (iii) the distribution is updated following each signal that occurs with positive probability using Bayes' rule.[4]

I next formalize each requirement in turn. First, there needs to be a subjective distribution $\mathbb{Q} \in \Delta(X \times S)$ over the set of states and signals. Second, the true distribution of signals $\mathbb{P}_S$ needs to be absolutely continuous with respect to the S-marginal $\mathbb{Q}_S$. This requirement ensures that Bayes' rule is applicable following every contingency. When $\mathbb{P}_S$ has finite support, absolute continuity reduces to the requirement that $\mathbb{Q}_S(\{s\}) > 0$ for any signal s such that $\mathbb{P}_S(\{s\}) > 0$. Third, the posterior belief given signal s needs to be given by a regular conditional probability $\mathbb{Q}(\cdot|s) \in \Delta(X)$.[5] Regular conditional probabilities define a mapping $\varphi_{\mathbb{Q}} : s \mapsto \mathbb{Q}(\cdot|s)$ from signals to posteriors, which is unique up to signals belonging to a set of $\mathbb{Q}_S$-null probability. Bayesian updating given subjective distribution $\mathbb{Q}$ requires the prior to be updated using (a version of) this mapping. When $\mathbb{Q}(\{s\}) > 0$ for some s, then $\mathbb{Q}(D|s) = \mathbb{Q}(D \times \{s\})/\mathbb{Q}_S(\{s\})$ for any measurable subset D of X, and $\varphi_{\mathbb{Q}}$ reduces to the usual Bayes' rule.

The $\varphi_{\mathbb{Q}}$ mapping determines the posterior as a function of the realized signal, but it does not specify the distribution of posteriors. In particular, for any measurable subset D of X, $\mathbb{Q}(D|s)$ is a random variable whose distribution depends on the distribution of the signal. To determine the probability with which each posterior is realized, one needs to use the true distribution of signals. Given the true distribution of signals $\mathbb{P}_S$ and a subjective distribution $\mathbb{Q} \in \Delta(X \times S)$, the Bayesian posterior about the state is distributed according to the probability distribution $\mathbb{P}_S \circ \varphi_{\mathbb{Q}}^{-1} \in \Delta(\Delta(X))$, defined as

$$\mathbb{P}_S \circ \varphi_{\mathbb{Q}}^{-1}(E) \equiv \mathbb{P}_S\left(\{s \in S : \varphi_{\mathbb{Q}}(s) \in E\}\right) \tag{2}$$

[4] Although the second requirement is not essential for the paper's main finding, having a more demanding notion of Bayesianism strengthens the "sufficiency" part of the result by highlighting the fact that the conclusion does not rely on the inapplicability of Bayes' rule after zero-probability events. It also allows me to turn the statement of the main result into an "if and only if" statement.

[5] Given the measurable space $(X \times S, \mathcal{X} \times \mathcal{S})$ and probability distribution $\mathbb{Q} \in \Delta(X \times S)$, a *regular conditional probability* of $\mathbb{Q}$ given $\mathcal{S}$ is a mapping $\mathbb{Q}(\cdot|\cdot) : \mathcal{X} \times S \to [0,1]$ such that (i) $\mathbb{Q}(\cdot|s)$ is a probability distribution on X for every $s \in S$, (ii) the mapping $s \mapsto \mathbb{Q}(D|s)$ is measurable for all $D \in \mathcal{X}$, and (iii) the kernel $\mathbb{Q}(\cdot|\cdot)$ satisfies

$$\mathbb{Q}(D \times E) = \int_E \mathbb{Q}(D|s)\mathbb{Q}_S(ds) \tag{1}$$

for all $D \in \mathcal{X}$ and $E \in \mathcal{S}$, where $\mathbb{Q}_S$ is the S-marginal of $\mathbb{Q}$. The existence of a regular conditional probability is guaranteed when X and S are complete separable metric spaces—see Faden (1985). Regular conditional probabilities are generally not unique, but they are unique up to $\mathbb{Q}_S$-null sets.

for any measurable subset E of $\Delta(X)$.[6] This is the observed distribution of posteriors when agents who hold subjective distribution $\mathbb{Q}$ observe signals distributed according to $\mathbb{P}_S$ and update their priors using Bayes' rule.

I can now define what it means for a belief sequence to be consistent with misspecified Bayesianism.

Definition 1. Given the true distribution of signals $\mathbb{P}_S \in \Delta(S)$, a pair (μ_0, F_1), consisting of a prior and a distribution of posteriors, is *consistent with misspecified Bayesianism* if there exists a subjective distribution $\mathbb{Q} \in \Delta(X \times S)$ that satisfies the following conditions:

(a) $\mathbb{Q}_X = \mu_0$,

(b) $\mathbb{P}_S$ is absolutely continuous with respect to $\mathbb{Q}_S$,

(c) $\mathbb{P}_S \circ \varphi_{\mathbb{Q}}^{-1} = F_1$,

where $\mathbb{Q}_X$ and $\mathbb{Q}_S$ are the X- and S-marginals of the subjective distribution $\mathbb{Q}$, respectively, $\varphi_{\mathbb{Q}}$ is the Bayesian update given subjective distribution $\mathbb{Q}$, and $\mathbb{P}_S \circ \varphi_{\mathbb{Q}}^{-1}$ is defined in (2).

This definition formalizes the intuitive notion of misspecified Bayesianism. Rationalizing a belief sequence requires finding a subjective distribution $\mathbb{Q}$ that explains the observed changes in beliefs. Such a conjectured $\mathbb{Q}$ is a joint distribution for the state and signal that must satisfy three conditions: Condition (a) of Definition 1 simply requires the conjectured distribution to be consistent with the prior. Condition (b) is the requirement that the subjective distribution assigns positive probability to signals that are realized with positive probability. Condition (c) requires that the distribution of posteriors F_1 matches the distribution obtained when starting with the conjectured distribution $\mathbb{Q}$, observing signals as per $\mathbb{P}_S$, and updating the subjective distribution using Bayes' rule.

A minimal requirement for (μ_0, F_1) to be consistent with Bayesianism given $\mathbb{P}_S$ is that $F_1 = \mathbb{P}_S \circ \varphi^{-1}$ for *some* measurable mapping $\varphi : S \to \Delta(X)$. If no such φ existed, no updating rule—Bayesian or non-Bayesian—could generate a distribution F_1 of posteriors based on a signal distributed according to $\mathbb{P}_S$. To rule out such cases, in the remainder of the paper, I assume that F_1 and $\mathbb{P}_S$ can be linked via $F_1 = \mathbb{P}_S \circ \varphi^{-1}$ for some possibly

[6] Under the assumption that $\mathbb{P}_S$ is absolutely continuous with respect to $\mathbb{Q}_S$, any zero $\mathbb{Q}_S$ measure set also has zero $\mathbb{P}_S$ measure. Therefore, changing the value of $\varphi_{\mathbb{Q}}$ on a $\mathbb{Q}_S$-zero probability set does not change the expression in (2), and so, the value of $\mathbb{P}_S \circ \varphi_{\mathbb{Q}}^{-1}$ is independent of the version of $\varphi_{\mathbb{Q}}$ being used.

unknown mapping φ. I also assume without loss of generality that $S = \Delta(X)$ (by relabeling signals by the posterior that they induce). The question is then whether $F_1 = \mathbb{P}_S \circ \varphi_{\mathbb{Q}}^{-1}$ for an updating rule $\varphi_{\mathbb{Q}}$ that is Bayesian given a subjective distribution $\mathbb{Q}$ whose X-marginal agrees with the observed prior μ_0.

3 Results

The paper's main result establishes a necessary and sufficient condition for a (μ_0, F_1) pair to be consistent with misspecified Bayesianism. This section presents this characterization result and examines several special cases.

3.1 Preliminaries

Before presenting the result, I introduce two definitions that are used in its statement. The first is the following:

Definition 2. Given a measurable subset E of the set of posteriors $\Delta(X)$ with $F_1(E) > 0$, the *conditional average posterior given E* is the probability distribution over X defined as

$$\overline{\mu}_1^E(D) \equiv \frac{\int_E \mu(D) F_1(d\mu)}{F_1(E)}$$

for any measurable set D. I refer to $\overline{\mu}_1 \equiv \overline{\mu}_1^{\Delta(X)}$ as the *average posterior.*

That $\overline{\mu}_1^E$ is a probability distribution over X for any set E with $F_1(E) > 0$ is easy to see: $\overline{\mu}_1^E(D) \in [0,1]$ for any D, $\overline{\mu}_1^E(\emptyset) = 0$ for the empty set $\emptyset$, $\overline{\mu}_1^E(X) = 1$ for the entire space X, and $\overline{\mu}_1^E$ is countably additive.

The paper's main result establishes that a (μ_0, F_1) pair is consistent with misspecified Bayesianism if and only if the set of posteriors can be partitioned into a collection of measurable cells such that the conditional average posterior given a cell of positive F_1 measure satisfies an absolute continuity condition with respect to the prior. The appropriate absolute continuity notion is the following:

Definition 3 (Kalai and Lehrer, 1993)**.** For probability distributions P and Q defined over the same measurable space, P *contains a grain of* Q if $P = \epsilon Q + (1-\epsilon)Q'$ for some $\epsilon \in (0,1]$ and some probability measure Q'.

The following proposition states two alternative definitions, which are equivalent to Definition 3:

Proposition 1. *For probability distributions P and Q defined over the same measurable space, the following are equivalent:*

(i) P contains a grain of Q.

(ii) The Radon–Nikodym derivative $f \equiv \frac{dQ}{dP}$ exists and is bounded P-almost surely.

(iii) There exists a constant $c \geq 1$ such that $Q(E) \leq cP(E)$ for any measurable set E.

The proposition illustrates that the grain condition is stronger than absolute continuity. Q is absolutely continuous with respect to P if $Q(E) = 0$ for any event E for which $P(E) = 0$, whereas P contains a grain of Q when the ratio $Q(E)/P(E)$ is bounded uniformly in E. The condition in Definition 3 can thus be seen as a form of "uniform absolute continuity." The following example illustrates the wedge between absolute continuity and the grain condition:

Example 1. Consider two continuous distributions P and Q over the reals with densities f_P and f_Q. If $\operatorname{supp} Q \subseteq \operatorname{supp} P$, then Q is absolutely continuous with respect to P. However, P contains a grain of Q only if f_Q/f_P is bounded. In particular, P does *not* contain a grain of any Q that has heavier tails than P. For example, a normal distribution never contains a grain of a Laplace or exponential distribution. See Subsection 3.4 for more on tail restrictions implied by the grain condition.

3.2 Characterization of misspecified Bayesianism

With Definitions 1–3 in hand, I can state the paper's main result.

Theorem 1. *The pair (μ_0, F_1) is consistent with misspecified Bayesianism* if and only if *there exists a measurable partition of the set of posteriors $\Delta(X)$ into sets $\{E_k\}_k$ such that, for every E_k with $F_1(E_k) > 0$, the prior μ_0 contains a grain of the conditional average posterior $\overline{\mu}_1^{E_k}$ given E_k.*

Misspecified Bayesianism is thus fully characterized by the condition that the prior contains a grain of a set of average posteriors. Absent additional a priori restrictions on what constitutes a reasonable subjective distribution, any belief sequence that satisfies this condition is consistent with Bayesian updating. It is easy to see that the absolute continuity of realized posteriors with respect to the prior is necessary for Bayesianism: If the prior of a Bayesian agent assigns zero probability to an event, her posterior must also assign zero probability to the event—regardless of the agent's subjective belief and the true distribution of signals. What is more surprising is that consistency with misspecified Bayesianism requires the stronger condition that the prior contains a grain of average posteriors. Furthermore, this condition is both necessary and sufficient, so it cannot be weakened.

The partition is only required for the "only if" direction of the theorem. The belief sequence of a Bayesian agent is always a martingale given the agent's subjective signal distribution. By a change of measure, this condition can be written as a constraint on the average posterior under the true distribution. However, the change of measure introduces the likelihood ratio (or Radon–Nikodym derivative) $d\mathbb{P}_S/d\mathbb{Q}_S$.[7] As long as this likelihood ratio is bounded, being a martingale with respect to the subjective distribution implies the grain condition for the average posterior under the true distribution. The partition in the theorem groups posteriors so that the likelihood ratio is bounded within each cell, even if it is unbounded globally. When the distribution of posteriors is supported on a finite set, the likelihood ratio is globally bounded. Misspecified Bayesianism is then characterized by the two easier-to-check conditions stated in the next result.

Proposition 2. *Consider a pair (μ_0, F_1), with F_1 supported on a finite set. The following are equivalent:*

(i) The pair (μ_0, F_1) is consistent with misspecified Bayesianism.

(ii) The prior μ_0 contains a grain of any posterior μ_1 in the support of F_1.

(iii) The prior μ_0 contains a grain of the average posterior $\overline{\mu}_1 \equiv \int \mu F_1(d\mu)$.

When the support of F_1 is finite, the partition in Theorem 1 can be taken to be the trivial partition (with one cell containing the entire space) or the singleton partition (with

[7]Condition (b) of Definition 1 ensures that this Radon–Nikodym derivative exists. If Condition (b) were to be dropped, the grain condition of Theorem 1 would still be sufficient for consistency with misspecified Bayesianism (but no longer necessary).

each cell a singleton). That makes it easier to check whether a pair (μ_0, F_1) is consistent with misspecified Bayesianism. Condition (iii) is particularly useful since it only requires the knowledge of the average posterior.

Theorem 1 characterizes misspecified Bayesianism for arbitrary state spaces. I now turn to several special cases and show that the grain condition manifests itself very differently across them. When the state space is bounded—finite or compact—the grain condition adds no substantive restrictions beyond a basic support restriction. In contrast, for unbounded state spaces, it puts a real constraint on the tail behavior of the posterior.

3.3 Finite or compact state spaces

I start with the special case where the state space X is finite, a common situation in applications. The following result establishes an easy-to-check condition that characterizes consistency with misspecified Bayesianism in that case:

Proposition 3. *Suppose the state space X is finite. Then the pair (μ_0, F_1) is consistent with misspecified Bayesianism* if and only if

$$\text{supp}\,\overline{\mu}_1 \subseteq \text{supp}\,\mu_0,$$

where $\overline{\mu}_1 \equiv \int \mu F_1(d\mu)$ denotes the average posterior.

This result characterizes misspecified Bayesianism in discrete settings. The average posterior cannot assign positive probability to states that have zero probability according to the prior. The necessity of this property is apparent given Bayes' rule; the proposition goes a step further by establishing its sufficiency. A corollary of Proposition 3 is the following:

Corollary 1. *Suppose the state space X is finite and μ_0 has full support over X. Then the pair (μ_0, F_1) is consistent with misspecified Bayesianism for any distribution F_1 of posteriors.*

Misspecified Bayesianism imposes *no* restrictions on posteriors when the state space is finite and the prior assigns positive probability to every state. The result casts doubt on the possibility of deciding whether decision makers are Bayesian in many common scenarios. I discuss this implication further in Section 7.

Misspecified Bayesianism continues to impose only weak restrictions when the state space is infinite but compact and priors and posteriors have well-behaved densities. The following proposition considers that case:

Proposition 4. *Let the state space X be a compact subset of $\mathbb{R}^n$. Suppose μ_0 and $\overline{\mu}_1$ admit continuous densities m_0 and $\overline{m}_1$ with respect to the Lebesgue measure such that $m_0(x) > 0$ for every $x \in \operatorname{supp} \mu_0$. Then the pair (μ_0, F_1) is consistent with misspecified Bayesianism* if and only if

$$\operatorname{supp} \overline{\mu}_1 \subseteq \operatorname{supp} \mu_0,$$

where $\overline{\mu}_1 \equiv \int \mu F_1(d\mu)$ denotes the average posterior.

Intuitively, the uniformity in the grain condition has no bite once the state space is bounded. On a compact set and with a continuous Radon–Nikodym derivative, simple absolute continuity already implies uniform absolute continuity. Moreover, in this setting absolute continuity of $\overline{\mu}_1$ with respect to μ_0 reduces to the support condition $\operatorname{supp} \overline{\mu}_1 \subseteq \operatorname{supp} \mu_0$. Hence, when the state space is compact, misspecified Bayesianism imposes no restrictions beyond ruling out an expansion of the prior's support.

3.4 Unbounded state spaces and tail behavior

Propositions 3 and 4 suggest that misspecified Bayesianism does not impose any meaningful restrictions on belief sequences when the state space is bounded. The conclusion changes dramatically when the state space is unbounded, as this subsection shows.

If the prior has unbounded support, then misspecified Bayesianism restricts the heaviness of the posterior's tails. For concreteness, I focus on the case where the state space is $X = \mathbb{R}^n$ and use the following (partial) tail order:

Definition 4. Let P, Q be probability distributions on $\mathbb{R}^n$ with P having unbounded support. If

$$\lim_{r\to\infty} \frac{Q(\|x\| > r)}{P(\|x\| > r)} = \infty,$$

then Q has *heavier tails* than P.[8]

The following uniform version of this partial order will also be useful:

[8] The definition is independent of the choice of norm because of the equivalence of norms on $\mathbb{R}^n$.

Definition 5. Let Q be a set of probability distributions and P be a probability distribution with unbounded support on $\mathbb{R}^n$. If

$$\lim_{r\to\infty} \inf_{Q\in\mathcal{Q}} \frac{Q(\|x\| > r)}{P(\|x\| > r)} = \infty,$$

then distributions in Q have *uniformly heavier tails* than P.

The next result shows that Bayesian updating *cannot* lead to tails that are uniformly heavier:

Proposition 5. *Suppose the state space is given by $X = \mathbb{R}^n$ and the prior μ_0 has unbounded support. If there exists an F_1-positive measure set of posteriors whose tails are uniformly heavier than μ_0, then (μ_0, F_1) is* not *consistent with misspecified Bayesianism.*

If the support of the distribution of posteriors is finite, we can dispense with uniformity and get the following sharper result:

Corollary 2. *Suppose the state space is given by $X = \mathbb{R}^n$ and the prior has unbounded support. Consider a pair (μ_0, F_1), with F_1 supported on a finite set. If there exists a posterior $\mu_1 \in \operatorname{supp} F_1$ whose tails are heavier than those of μ_0, then (μ_0, F_1) is* not *consistent with misspecified Bayesianism.*

Bayesian updating can redistribute the prior's probability mass. However, shifting mass into the tail regions requires signals that were themselves extremely unlikely under the prior. Allowing for misspecified beliefs about signal probabilities enlarges the set of attainable posteriors, because signals that are rare in reality may be deemed likely by the subjective model. Yet even under such misspecification there is a hard limit: Any posterior that is heavier-tailed than the prior cannot be produced by Bayesian updating—no matter how misspecified the subjective model. Observing such a heavy-tailed posterior is thus a telltale sign of violations of Bayes' rule.

4 Illustrative Examples

This section uses several examples to illustrate the theoretical results of the paper.

4.1 An example with a finite state space

The first example illustrates the proof of the "if" direction of Theorem 1 in a discrete-state setting. The state takes values in the set $X = \{H, L\}$. The prior is the uniform distribution over X. The distribution of posteriors F_1 is as follows: with a one-quarter probability, the belief that the state is H goes up to 0.8; with the remaining three-quarters probability, the belief that the state is H goes up to 1.0. Is this belief sequence consistent with misspecified Bayesianism? The answer is yes. This conclusion follows from Corollary 1 by noting that μ_0 has full support over X. In what follows, I illustrate how (μ_0, F_1) can be rationalized.

F_1 imposes some restrictions on the true distribution of signals $\mathbb{P}_S$ and the mapping φ used to form beliefs. Since the posterior takes on two values, there are at least two signals that are realized with positive probability. The observation of one set of signals moves the belief that the state is H to 0.8. Since with a one-quarter probability, the posterior is $\mu_1(\{H\}) = 0.8$, the signals that lead to this posterior must have probability $\mathbb{P}_S(\{s : \varphi(s) = (\mu_1(\{H\}) = 0.8)\}) = 0.25$. Likewise, there is a set of signals that has true probability $\mathbb{P}_S(\{s : \varphi(s) = (\mu_1(\{H\}) = 1)\}) = 0.75$ and leads to the posterior that the state is H with certainty. With slight abuse of notation, I refer to the $\{s : \varphi(s) = (\mu_1(\{H\}) = 0.8)\}$ and $\{s : \varphi(s) = (\mu_1(\{H\}) = 1)\}$ events simply as the $s = 0.8$ and $s = 1$ signals, respectively.[9]

I illustrate how (μ_0, F_1) can be rationalized by finding a subjective distribution $\mathbb{Q}$ such that the belief sequence of a Bayesian agent with subjective distribution $\mathbb{Q}$ matches the prior and the distribution of posteriors. The distribution $\mathbb{Q}$ needs to satisfy three requirements for it to rationalize the prior μ_0 and posterior distribution F_1. First, $\mathbb{Q}$ must be consistent with μ_0; i.e., $\mathbb{Q}_X(\{H\}) = \mu_0(\{H\}) = 0.5$. Second, it must assign positive probability to the $s = 0.8$ and $s = 1$ signals for Bayes' rule to be applicable after the observation of those signals. Third, the posterior conditional on the $s = 0.8$ and $s = 1$ signals must be consistent with the corresponding posteriors; i.e., $\mathbb{Q}(\{H\}|s = 0.8) = 0.8$ and $\mathbb{Q}(\{H\}|s = 1.0) = 1.0$.

One also needs to specify the subjective probability of observing signals other than 0.8

[9] This is not an abuse of notation under the assumption that φ is the identity mapping. Note that the assumption that φ is the identity mapping is innocuous in this example since F_1 only identifies $\mathbb{P}_S \circ \varphi^{-1} = F_1$—but not $\mathbb{P}_S$ or φ. Nonetheless, the construction in the example can be easily modified to allow for the possibility that F_1 and $\mathbb{P}_S$ are separately identified. I do not pursue this extension here since it would lead to additional notational complexity without offering any new insights. See the proof of Theorem 1 for the general construction.

and 1.0. I start by assuming that, according to $\mathbb{Q}$, the signal can only take values $s = 0.8$ and $s = 1.0$. This assumption constrains $\mathbb{Q}$ to satisfy $\mathbb{Q}(\{(x,s) : s \in \{0.8, 1.0\}\}) = 1$. This constraint, together with the requirements previously discussed and the requirement that $\mathbb{Q}(x,s) \geq 0$ for any (x,s), yields a mixed system of equalities and inequalities for the four unknown probabilities $\mathbb{Q}(H,0.8)$, $\mathbb{Q}(L,0.8)$, $\mathbb{Q}(H,1.0)$, and $\mathbb{Q}(L,1.0)$:

$$\mathbb{Q}(H,0.8) + \mathbb{Q}(L,0.8) > 0, \tag{3}$$

$$\mathbb{Q}(H,1.0) + \mathbb{Q}(L,1.0) > 0, \tag{4}$$

$$\frac{\mathbb{Q}(H,0.8)}{\mathbb{Q}(H,0.8) + \mathbb{Q}(L,0.8)} = 0.8, \tag{5}$$

$$\frac{\mathbb{Q}(H,1.0)}{\mathbb{Q}(H,1.0) + \mathbb{Q}(L,1.0)} = 1.0, \tag{6}$$

$$\mathbb{Q}(H,0.8), \mathbb{Q}(L,0.8), \mathbb{Q}(H,1.0), \mathbb{Q}(L,1.0) \geq 0, \tag{7}$$

$$\mathbb{Q}(H,0.8) + \mathbb{Q}(H,1.0) = 0.5, \tag{8}$$

$$\mathbb{Q}(L,0.8) + \mathbb{Q}(L,1.0) = 0.5. \tag{9}$$

It is easy to verify that this system does not have a solution.

Thus, for the belief sequence to be consistent with misspecified Bayesianism, the subjective distribution must entertain the possibility that the signal takes values outside the set $\{0.8, 1.0\}$. With slight abuse of notation, I let $s = \ominus$ denote the event that the signal takes a value outside the set $\{0.8, 1.0\}$. Constraints (8) and (9) must now be modified as follows:

$$\mathbb{Q}(H,0.8) + \mathbb{Q}(H,1.0) + \mathbb{Q}(H,\ominus) = 0.5, \tag{10}$$

$$\mathbb{Q}(L,0.8) + \mathbb{Q}(L,1.0) + \mathbb{Q}(L,\ominus) = 0.5. \tag{11}$$

The remaining requirements, expressed in equations (3)–(7), remain intact. However, $\mathbb{Q}$ must now additionally satisfy the two non-negativity requirements:

$$\mathbb{Q}(H,\ominus), \mathbb{Q}(L,\ominus) \geq 0. \tag{12}$$

Equations (3)–(7) and (10)–(12) constitute a mixed system of equalities and inequalities for the six unknown probabilities $\mathbb{Q}(H,0.8)$, $\mathbb{Q}(L,0.8)$, $\mathbb{Q}(H,1.0)$, $\mathbb{Q}(L,1.0)$, $\mathbb{Q}(H,\ominus)$, and $\mathbb{Q}(L,\ominus)$. The fact that the average posterior has the same support as the prior is sufficient to ensure that this system has a solution. One such solution—and the one corresponding to the proof of Theorem 1—is as follows:

	0.8	1.0	$\ominus$
H	0.25	0.25	0
L	0.0625	0	0.4375

Note that the pair (μ_0, F_1) can be rationalized only if we allow for a misspecified belief about the distribution of signals. If the subjective distribution were to agree with the true distribution on the probabilities of different signals, the system of equalities and inequalities that determines $\mathbb{Q}$ would have no solution.

4.2 Fattened tails and inconsistency with misspecified Bayesianism

The previous example may suggest that misspecified Bayesianism imposes no restrictions on the posterior when the prior has full support. The next example illustrates that this is not true when the state space is unbounded. I consider a pair (μ_0, F_1) that is not consistent with misspecified Bayesianism even though every posterior in the support of F_1 is absolutely continuous with respect to the prior.

The state space is the real line: $X = \mathbb{R}$. The prior is that the state is normally distributed with mean zero and unit variance. With one-half probability, the posterior is the exponential distribution with mean $1/\lambda$; with the remaining one-half probability, the posterior is the (mirrored) exponential distribution supported over $(-\infty, 0]$ with mean $-1/\lambda$. Therefore, the prior μ_0 is the standard normal distribution, and the average posterior $\overline{\mu}_1$ is the Laplace distribution with mean zero and scale parameter $1/\lambda$.

The prior and both realizations of the posterior are non-atomic probability measures over $X = \mathbb{R}$. Moreover, the support of each posterior is contained in the support of the prior. Therefore, the two realized posteriors are absolutely continuous with respect to the prior. However, the prior does *not* contain a grain of the average posterior (or either of the two realizations of the posterior), because the two posteriors in the support of F_1 both have heavier tails than the prior regardless of the value of λ. Therefore, Corollary 2 implies that (μ_0, F_1) is inconsistent with misspecified Bayesianism.

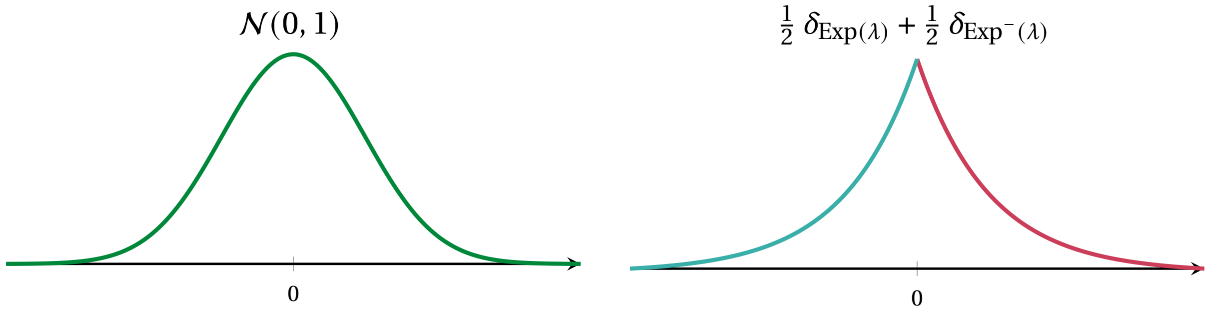

Figure 1. Subsection 4.2's example. The prior (left) and posterior (right). This prior-posterior pair is *not* consistent with misspecified Bayesianism.

The posterior variance of each realized posterior distribution is smaller than the prior variance when $\lambda > 1$. Yet this prior-posterior pair is inconsistent with Bayesian updating. This example illustrates that a shrinkage of the variance (or any finite moment of beliefs) is neither required by Bayes' rule nor is enough to guarantee consistency with it.

4.3 The necessity of using a partition in Theorem 1

When the support of F_1 is finite, Proposition 2 states that the partition in Theorem 1 can be chosen to be either the trivial partition or the singleton partition. The next example shows that this is not the case when the support of F_1 is infinite. Then a (μ_0, F_1) pair may be consistent with Bayesianism even if it violates conditions (ii) and (iii) of Proposition 2.

I start with a subjective distribution $\mathbb{Q}$ and the assumption that the posterior is generated from the prior using Bayes' rule—hence the induced prior-posterior pair will be consistent with misspecified Bayesianism by construction. I then argue that the prior does not contain a grain of any realization of the posterior or the (unconditional) average posterior. However, there exists a measurable partition of the support of the distribution of posteriors such that the prior contains a grain of the conditional average posterior given every cell of the partition.

The state belongs to the real line: $X = \mathbb{R}$. According to the subjective distribution $\mathbb{Q}$, the prior distribution of the state is the standard normal distribution and the signal equals the state with probability one. Hence, the posterior always equals the realization of the signal. The true signal distribution $\mathbb{P}_S$ is as follows: The support of $\mathbb{P}_S$ coincides with the support of the S-marginal $\mathbb{Q}_S$ of the subjective distribution. However, the true signal distribution is Laplace$(0, 1/\lambda)$, i.e., the Laplace distribution with location parameter $\mu = 0$ and scale parameter $1/\lambda$. Therefore, the posterior is always a point mass at some $Z \in \mathbb{R}$, with Z distributed according to the Laplace$(0, 1/\lambda)$ distribution.[10]

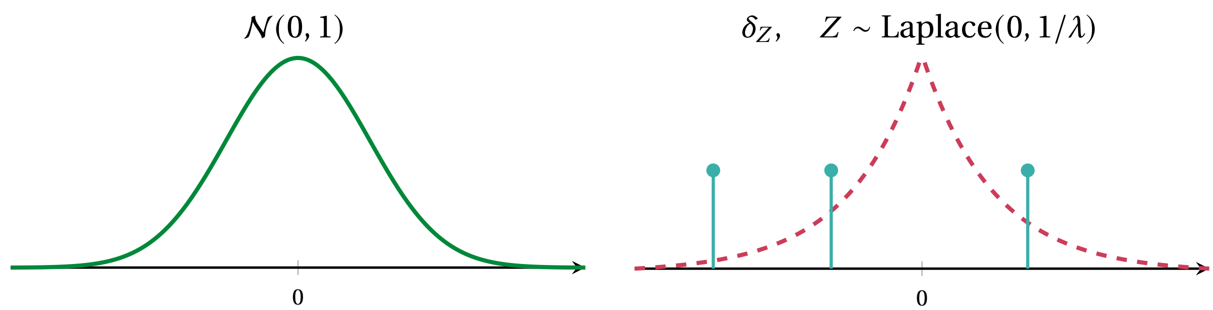

Figure 2. Subsection 4.3's example. The prior (left) and posterior (right). Each point mass in the right panel is a realization of the posterior; the dashed line shows the density of the location of those point masses.

The Radon–Nikodym derivative of the Laplace$(0, 1/\lambda)$ distribution with respect to the normal distribution is unbounded. Therefore, the prior does *not* contain a grain of the average posterior, and statement (iii) in Proposition 2 does not hold. Similarly, the prior does not contain a grain of any of the posteriors in the support of the distribution of posteriors, so statement (ii) in Proposition 2 does not hold either. Yet, by construction, the induced (μ_0, F_1) pair is consistent with Bayesianism and must therefore satisfy the grain condition in Theorem 1.

The distribution of posteriors indeed satisfies the partition version of the grain condition. To see this, consider a partition of the reals into a countable union of non-empty intervals D_k of finite length, and consider the measurable partition of the support of the distribution of posteriors into sets $E_k \equiv \{\delta_x : x \in D_k\}$. The average posterior over set E_k is equal to the truncated distribution obtained from restricting the Laplace$(0, 1/\lambda)$ distribution to D_k. Since the Radon–Nikodym derivative of the truncated Laplace$(0, 1/\lambda)$ distribution with respect to the standard normal distribution is bounded, the prior contains a grain of the conditional average posterior given any E_k.

[10]The example can be formalized as follows: $X = \mathbb{R}$, and supp $\mathbb{P}_S$ = supp $\mathbb{Q}_S = \{\delta_x : x \in \mathbb{R}\}$, where δ_x denotes the point mass at x. For any measurable set $D \subset \mathbb{R}$, $\mathbb{P}_S(\{\delta_x : x \in D\}) = F_{\text{Laplace}(0,1/\lambda)}(D)$ and $\mathbb{Q}_S(\{\delta_x : x \in D\}) = F_{\mathcal{N}(0,1)}(D)$. Finally, $\mathbb{Q}$ is defined via equation (1), where $\mathbb{Q}(D|\delta_x) = \mathbb{1}\{x \in D\}$ and $\mathbb{Q}(D|s)$ is arbitrary when s is not a point mass at some $x \in \mathbb{R}$. I am grateful to Eran Shmaya for suggesting this example.

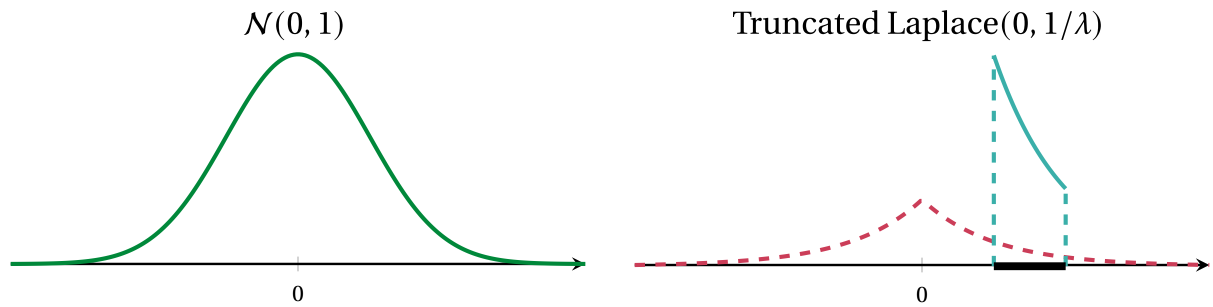

Figure 3. Subsection 4.3's example. The prior (left) and the conditional average posterior given a cell of the partition of supp F_1 (right).

Although both the prior and the average posterior are identical across the last two examples, only the current example is consistent with misspecified Bayesianism. This shows that knowledge of the average posterior is, in general, insufficient to determine consistency. It is only when the distribution of posteriors has finite support that one can decide consistency simply by examining the prior and the average posterior.

5 Extensions

This section discusses several extensions of the paper's main result.

5.1 Rationalizing belief sequences across different states

In the baseline model, the distribution of posteriors is observed given a fixed state x. Here, I consider a generalization where the distribution of posteriors is observed given every possible realizations of the state. Let $\mathbb{P} \in \Delta(X \times S)$ denote the true joint distribution of the state and signal. When the realized state is x, signals are drawn from $\mathbb{P}(\cdot|x)$, where $\mathbb{P}(\cdot|\cdot)$ is a regular conditional probability of $\mathbb{P}$ given the Borel sigma-algebra on X. I let $\mathbb{P}_X$ and $\mathbb{P}_S$ denote the X- and S-marginals of $\mathbb{P}$, respectively.

The goal is to rationalize the pair $(\mu_0, \{F_{1x}\}_{x\in X})$, where F_{1x} is the distribution of posteriors when the realized state of the world is x. The beliefs are updated using the same mapping $\varphi : S \to \Delta(X)$ regardless of the realized state of the world; that is, $F_{1x} = \mathbb{P}(\cdot|x) \circ \varphi^{-1}$ for some unknown mapping φ and for all $x \in X$. This restriction captures the idea that the signal summarizes any information that can be used to update beliefs.

The following definition extends the notion of consistency introduced in Definition 1 to this setting:

Definition 6. Given the true joint distribution of the state and signal $\mathbb{P} \in \Delta(X \times S)$, the pair $(\mu_0, \{F_{1x}\}_{x \in X})$, consisting of a prior and a distribution of posteriors in each state of the world, is *consistent with misspecified Bayesianism* if there exists a subjective distribution $\mathbb{Q} \in \Delta(X \times S)$ that satisfies the following conditions:

(a) $\mathbb{Q}_X = \mu_0$,

(b) $\mathbb{P}_S$ is absolutely continuous with respect to $\mathbb{Q}_S$,

(c) $\mathbb{P}(\cdot|x) \circ \varphi_{\mathbb{Q}}^{-1} = F_{1x}$ for $\mathbb{P}_X$-almost all x,

where $\varphi_{\mathbb{Q}}$ is the Bayesian update given subjective distribution $\mathbb{Q}$.

Misspecified Bayesianism in this case turns out to be characterized by a condition similar to the condition that characterizes it in the fixed-state setting. The only difference is that F_1 in Theorem 1 needs to be replaced with the average of $\{F_{1x}\}_{x \in X}$ across different states. Define

$$\overline{F}_1 \equiv \int_X F_{1x} \mathbb{P}_X(dx),$$

and let

$$\overline{\mu}_1^E \equiv \frac{\int_E \mu \overline{F}_1(d\mu)}{\overline{F}_1(E)}$$

for any measurable subset E of the set of posteriors such that $\overline{F}_1(E) > 0$. The following result generalizes Theorem 1:

Theorem 2. *The pair* $(\mu_0, \{F_{1x}\}_{x \in X})$ *is consistent with misspecified Bayesianism* if and only if *there exists a measurable partition of the set of posteriors* $\Delta(X)$ *into sets* $\{E_k\}_k$ *such that, for every* E_k *with* $\overline{F}_1(E_k) > 0$*, the prior* μ_0 *contains a grain of* $\overline{\mu}_1^{E_k}$.

Requiring state-by-state rationalization does not significantly alter the set of belief sequences that are consistent with Bayesian updating. A version of the grain condition continues to fully characterize misspecified Bayesianism. The only difference is that the grain condition is now imposed on the *average* distribution of posteriors $\overline{F}_1$. When the true distribution of the state is degenerate, i.e., $\mathbb{P}_X(\{x^*\}) = 1$ for some $x^* \in X$, then Theorem 2 reduces to Theorem 1.

5.2 Rationalizing the joint distribution of the posterior and signal

Suppose signal realizations are observed (in addition to posteriors), and the goal is to rationalize the pair (μ_0, G_1), where $G_1 \in \Delta(S \times \Delta(X))$ is the joint distribution of signal-posterior pairs. Under the assumption that the posterior is a function φ of the realized signal, we have $G_1 = \mathbb{P}_S \circ \Phi^{-1}$, where $\mathbb{P}_S$ is the true signal distribution and the Φ mapping is defined as $\Phi : s \mapsto (s, \varphi(s))$.

If the $\Delta(X)$-marginal of G_1, i.e., $F_1 \equiv \mathbb{P}_S \circ \varphi^{-1}$, satisfies the grain condition of Theorem 1, then (μ_0, G_1) is consistent with misspecified Bayesianism given a subjective distribution $\mathbb{Q}$ that is identical to the one used to prove the "if" part of Theorem 1. The reason is that the construction in Theorem 1 defines a mapping $\varphi_{\mathbb{Q}}$ that coincides ($\mathbb{P}_S$-almost everywhere) with the true mapping φ used to update beliefs. Therefore, $\mathbb{P}_S \circ \Phi^{-1} = \mathbb{P}_S \circ \Phi_{\mathbb{Q}}^{-1}$, where $\Phi_{\mathbb{Q}} : s \mapsto (s, \varphi_{\mathbb{Q}}(s))$.

5.3 More than two periods

Consider a sequence of beliefs $(\mu_t)_t$, where each μ_t is now a random variable. Let $F_{t+1}(\cdot|\mu_t)$ denote the distribution of μ_{t+1} conditional on μ_t. A straightforward generalization of Theorem 1 characterizes misspecified Bayesianism in this case: The sequence $(\mu_t)_t$ is consistent with misspecified Bayesianism if (and only if) for almost every μ_t, the pair $(\mu_t, F_{t+1}(\cdot|\mu_t))$ satisfies the grain condition of Theorem 1.

5.4 Time-varying states and filtering

Suppose there is a time-varying state x_t, which belongs to a separable metric space. Theorem 1 already covers this case by defining $x \equiv (x_t)_t$ and taking X to be the separable metric space of infinite sequences.

6 Bayes plausibility and related notions

Definition 1 puts no restrictions on what constitutes a reasonable subjective distribution $\mathbb{Q}$. The assumption that any well-defined subjective distribution is permissible is made in keeping with Savage (1972)'s idea of purely subjective probability. I assume that any subjective distribution $\mathbb{Q}$ that can rationalize (μ_0, F_1) is a valid subjective distribution. In other words, rationality of beliefs is not judged by what those beliefs are but by how they are updated. I next discuss two alternatives to this assumption proposed in the literature and how they change the conclusion of Theorem 1.

The first alternative I consider imposes a correctly specified belief about the distribution of signals. This assumption leads to Bayes plausibility (or the martingale property of Bayesian beliefs): The average posterior must equal the prior. Aumann and Maschler (1995) and Kamenica and Gentzkow (2011) show that this is indeed the only restriction Bayesian updating imposes on beliefs. The following theorem adapts this result to general metric spaces. More importantly, however, it highlights the fact that Bayes plausibility characterizes Bayesianism *only* under the assumption of correct beliefs about the distribution of signals.

Theorem 3 (Kamenica and Gentzkow, 2011)**.** *The pair (μ_0, F_1) is consistent with Bayesianism given a subjective distribution $\mathbb{Q}$ with the S-marginal satisfying $\mathbb{Q}_S = \mathbb{P}_S$* if and only if $\mu_0 = \overline{\mu}_1 \equiv \int \mu F_1(d\mu)$.

A more permissive notion of Bayesianism is proposed by Shmaya and Yariv (2016). They allow for incorrect beliefs about the distribution of signals—as long as the supports of those beliefs coincide with the support of the true distribution. The following theorem generalizes Shmaya and Yariv (2016)'s Lemma 1 to general metric state spaces and arbitrary true signal distributions. It reduces to their result when both X and $\operatorname{supp} \mathbb{P}_S$ are finite sets. However, its main significance is to clarify that Shmaya and Yariv (2016)'s conclusion relies on an a priori restriction on what constitutes a reasonable subjective distribution.

Theorem 4 (Shmaya and Yariv, 2016)**.** *The following statements are equivalent:*

(i) The pair (μ_0, F_1) is consistent with Bayesianism given a subjective distribution $\mathbb{Q}$ whose S-marginal $\mathbb{Q}_S$ is equivalent to $\mathbb{P}_S$.[11]

(ii) There exists a probability measure $\lambda \in \Delta(\Delta(X))$ such that λ and F_1 are equivalent and $\mu_0 = \int \mu\lambda(d\mu)$.

[11] Probability distributions P and Q are *equivalent* if Q is absolutely continuous with respect to P and P is absolutely continuous with respect to Q.

Theorems 3 and 4 show that misspecified Bayesianism is a less restrictive notion than either of the notions considered by Kamenica and Gentzkow (2011) and Shmaya and Yariv (2016). The following table summarizes the relationship between different notions of Bayesianism and the conditions that characterize them:

Bayes plausibility (KG, 2011)		Shmaya and Yariv (2016)		misspecified Bayesianism
$\mu_0 = \int \mu F_1(d\mu)$	$\Longrightarrow$	$\mu_0 = \int \mu\lambda(d\mu), \quad \lambda \sim F_1$	$\Longrightarrow$	$\mu_0 = \epsilon \int \mu F_1(d\mu) + (1-\epsilon)\mu_1'$

7 Implications

7.1 Tests of Bayesian updating

Consider an econometrician who wants to decide whether a group of agents updates their beliefs about a state of the world $x \in X$ using Bayes' rule. There are two periods $t = 0, 1$, and a large number of agents indexed by $i \in I$. Agent i's time-t belief about the value of x is a probability distribution, denoted by $\mu_{it} \in \Delta(X)$. Agents' beliefs may evolve between the periods due to new information. Let s_i denote the signal observed by agent i between the two periods.

The econometrician does not know the realization of signals agents use to update their beliefs. However, he can elicit what each agent believes about the state of the world in each of the two periods and perfectly observe μ_{it} for all i and $t = 0, 1$. The econometrician's question is then whether the agents' observed belief paths $\{(\mu_{i0}, \mu_{i1})\}_{i \in I}$ are consistent with Bayesian updating given some subjective belief about the distribution of signals.

I make several assumptions. First, agents are ex ante identical. In particular, $\mu_{i0} = \mu_0$ for some $\mu_0 \in \Delta(X)$ and all $i \in I$. Second, agents' signals are independent and identically distributed, with $\mathbb{P}_S \in \Delta(S)$ denoting the true distribution of signals given the fixed state of the world. Third, agents all use the same mapping $\varphi = \varphi_i : s_i \mapsto \mu_{i1}$ to form their beliefs as a function of their signals. Fourth, the number of agents is large enough that the empirical distribution of observed posteriors $\{\mu_{i1}\}_{i \in I}$ provides an arbitrarily good approximation to the corresponding population distribution (by the law of large numbers). Specifically, I assume that the econometrician can perfectly observe the population distribution of agents' posterior beliefs, denoted by $F_1 \in \Delta(\Delta(X))$. Fifth, the econometrician is assumed to know everything described so far—except for the mapping φ used by agents to update their beliefs. Finally and for simplicity, I assume that $\mathbb{P}_S$ and F_1 have finite support. This assumption serves no purpose other than to allow me to rely on Proposition 2 instead of the more complicated characterization in Theorem 1.

The first five assumptions all make it easier for the econometrician to conclude that agents must *not* be Bayesian. The assumptions that agents have identical priors, observe i.i.d. signals, and use identical mappings to update their beliefs all help with the identification of agents' subjective beliefs. The econometrician observes each agent only after the realization of a single signal. Without these homogeneity assumptions, how an agent behaves after a signal would not be informative of how other agents would have behaved if they had observed that same signal; that would make it harder to conclude that an agent has behaved in a non-Bayesian way. The assumption that $\mathbb{P}_S$, μ_0, and F_1 are perfectly observed by the econometrician limits what he can freely choose in order to rationalize his observations. This assumption too makes it easier for him to reject an agent's Bayesianism.

Yet the econometrician can only reject agents' misspecified Bayesianism if he observes a violation of the grain condition. Proposition 2 implies that the econometrician's observations are consistent with Bayesian updating by agents if and only if μ_0 contains a grain of the average posterior $\overline{\mu}_1 \equiv \int \mu F_1(d\mu)$. If the state space X is finite and agents' prior has full support over X, then *any* distribution of posteriors is consistent with agents being Bayesian. On the other hand, if the prior is supported on an unbounded subset of the Euclidean space, misspecified Bayesianism disallows the average posterior from having heavier tails than the prior's tails.

These observations have two practical implications. First, they suggest that essentially any test of Bayesian updating proposed in the literature is a joint test of Bayesian updating and the assumption that agents have a correctly specified model of the data-generating process. Therefore, any rejection by those tests could be due to non-Bayesian updating, misspecification, or a combination thereof. Second, they suggest that one can indeed test Bayesian updating, in isolation, by examining the change in belief tails. In particular, the posterior having heavier tails than the prior provides strong evidence against Bayesian updating.

7.2 Equivalence of Bayesian and non-Bayesian updating rules

The paper's results suggest that many non-Bayesian updating rules are observationally equivalent to Bayesian updating under misspecified subjective models. Consider an agent who observes signals generated according to some probability distribution $\mathbb{P}_S$ and updates her belief using some non-Bayesian updating rule φ. As long as φ does not lead to violations of the grain condition, it is *as if* the agent used Bayes' rule to update her belief—but under a misspecified subjective model $\mathbb{Q}$ of her observations (with $\mathbb{Q}_S \neq \mathbb{P}_S$).

The following example makes this point concrete in the context of diagnostic expectations (Bordalo, Gennaioli, and Shleifer, 2018; Bordalo, Gennaioli, Ma, and Shleifer, 2020). I focus on the updating step of diagnostic expectations since the deviation from the benchmark framework appears in that step. There is a fixed state $x \in \mathbb{R}$ and a signal $s \in \mathbb{R}$ about the state.[12] The true distribution of the signal $\mathbb{P}_S$ is given by

$$s = x + \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2). \tag{13}$$

All agents—whether correctly specified Bayesian, diagnostic, or misspecified Bayesian—are assumed to share the prior that x is normally distributed with some mean $\bar{x}$ and some variance σ^2. Let μ_0 denote that prior. According to the posterior μ_1 of a correctly specified Bayesian, x is normally distributed with mean $\bar{x} + K(s - \bar{x})$ and variance $(1-K)\sigma^2$, where $K \equiv \sigma^2/(\sigma^2 + \sigma_\epsilon^2)$ is the Kalman gain.

Next consider a diagnostic agent who starts with the same prior as a correctly specified Bayesian but updates her belief by overweighting representative states. The agent's posterior probability of state x is given by $\mu_1^\theta(x) = \mu_1(x)R(x)^\theta \frac{1}{Z}$, where $R(x)$ is a measure of the representativeness of state x given the signal, θ is the diagnosticity parameter, and Z is a normalizing constant. Proposition 1 of Bordalo et al. (2020) shows that μ_1^θ is normal with variance $(1-K)\sigma^2$ and mean $\bar{x} + (1+\theta)K(s - \bar{x})$.[13] That is, the posterior uncertainty of the diagnostic agent matches that of a correctly specified Bayesian, whereas the mean of her posterior is more sensitive to the signal.

The belief sequence of the diagnostic agent satisfies the grain condition. The prior belief is normal, while the posterior belief is normal with a fixed variance and a mean

[12] In Bordalo et al. (2018, 2020), the state is time-varying, and the prior itself is obtained from Kalman filtering. However, the dynamics of the state are orthogonal to the distortion from diagnostic expectations.

[13] Note that both μ_1 and μ_1^θ are random since s is a random variable.

that is itself normally distributed. Consider a partition of the set of posteriors in which each cell consists of posteriors whose means belong to a bounded interval. The prior contains a grain of the average diagnostic posterior over every cell of such a partition.[14] Therefore, Theorem 1 implies that the diagnostic agent's belief sequence is consistent with Bayesianism given some subjective distribution $\mathbb{Q}$.

The following parsimonious subjective model rationalizes diagnostic expectations. The prior belief about the state is given by $\mathbb{Q}_X = \mu_0$, while the conditional signal distribution $\mathbb{Q}(\cdot|x)$ is given by

$$s = x + \epsilon, \qquad \epsilon \sim \mathcal{N}\left(\frac{-\theta}{1+\theta}(x - \bar{x}), \frac{\sigma_\epsilon^2}{(1+\theta)^2}\right). \tag{14}$$

The X-marginal $\mathbb{Q}_X$ and conditional distribution $\mathbb{Q}(\cdot|x)$ define a subjective distribution $\mathbb{Q} \in \Delta(X \times S)$ as follows:

$$\mathbb{Q}(D \times E) \equiv \int_D \mathbb{Q}(E|x)\mu_0(dx) \tag{15}$$

for measurable sets $D \subseteq X$ and $E \subseteq S$. It is straightforward to verify that the distribution of posteriors of a misspecified Bayesian with subjective model $\mathbb{Q}$ coincides with that of a diagnostic agent with parameter θ under the true signal distribution $\mathbb{P}_S$.[15]

This rationalization offers an alternative interpretation of diagnostic agents' behavioral bias. Diagnostic agents behave *as if* they believe that the signal is less noisy conditional on the state and corrupted by a noise term that is negatively correlated with the state. This belief increases the Kalman gain without altering the agents' posterior uncertainty.

[14] When $\theta < \sqrt{1 + \sigma^2/\sigma_\epsilon^2} - 1$, the prior contains a grain of the average posterior, and the partition in Theorem 1 can be taken to be the trivial partition.

[15] In fact, a stronger notion of observational equivalence holds here: The *joint* distribution of signal-posterior pairs is matched *for every value of* x. Suppose equation (13) describes $\mathbb{P}(\cdot|x)$, i.e., the true distribution of the signal *conditional* on the state x. The Bayesian update $\varphi_{\mathbb{Q}}$ given the subjective model in (14) and (15) equals the diagnostic update φ for every signal s. Therefore, $\mathbb{P}(\cdot|x) \circ \Phi_{\mathbb{Q}}^{-1} = \mathbb{P}(\cdot|x) \circ \Phi^{-1}$ for every x, where $\Phi_{\mathbb{Q}} : s \mapsto (s, \varphi_{\mathbb{Q}}(s))$ and $\Phi : s \mapsto (s, \varphi(s))$. See Subsections 5.1 and 5.2 for generalizations of Theorem 1 that cover two stronger notions of observational equivalence.

Proofs

Proof of Proposition 1

The proof involves showing (i) $\implies$ (ii) $\implies$ (iii) $\implies$ (i).

Proof of (i) $\implies$ (ii). If $P = \epsilon Q + (1-\epsilon)Q'$, then $Q(E) \leq \frac{1}{\epsilon}P(E)$ for any measurable set E. Therefore, Q is absolutely continuous with respect to P, and so, by the Radon–Nikodym theorem, there exists a derivative $f \equiv \frac{dQ}{dP}$. I finish the proof by showing that f is bounded P-almost surely. Toward a contradiction, suppose that for any positive constant C there exists a measurable set E with $P(E) > 0$ such that $f > C$ on E. Then,

$$Q(E) = \int_E dQ = \int_E f\,dP > C\int_E dP = CP(E).$$

Since C is arbitrary, there exists no constant $\epsilon > 0$ such that $Q(E) \leq \frac{1}{\epsilon}P(E)$ for all E, a contradiction.

Proof of (ii) $\implies$ (iii). Suppose the Radon–Nikodym derivative $f \equiv \frac{dQ}{dP}$ satisfies $f \leq c$ for some $c \geq 1$ and up to sets of zero P measure. For any measurable set E,

$$Q(E) = \int_E dQ = \int_E f\,dP \leq c\int_E dP = cP(E).$$

Proof of (iii) $\implies$ (i). Let $\epsilon \equiv 1/c \leq 1$. When $\epsilon = 1$, then Q' can be chosen arbitrarily. This is because $Q(E) \leq P(E)$ implies $Q(E^c) \geq P(E^c)$, where E^c denotes the complement of E. But $Q(E^c) \leq P(E^c)$ by assumption. Therefore, $Q(E) = P(E)$. Since E is an arbitrary measurable set, $Q = P$. When $\epsilon < 1$, set $Q' = \frac{1}{1-\epsilon}P - \frac{\epsilon}{1-\epsilon}Q$. To finish the proof, I need to argue that such a Q' is a probability measure. Note that $Q'(E) = \frac{1}{1-\epsilon}P(E) - \frac{\epsilon}{1-\epsilon}Q(E) \geq \frac{1}{1-\epsilon}P(E) - \frac{1}{1-\epsilon}P(E) = 0$. Moreover, since P and Q are probability measures, Q' returns zero for the empty set, returns one for the entire space, and is countably additive. Therefore, Q' is a probability measure. $\square$

Proof of Theorem 1

Proof of the "if" direction. The proof of this direction is constructive. Given the measurable space $(X, \mathcal{X})$ and the true signal distribution $\mathbb{P}_S$, I construct the subjective distribution $\mathbb{Q}$

that rationalizes an observed pair (μ_0, F_1) satisfying the assumption of the theorem. By assumption, $F_1 = \mathbb{P}_S \circ \varphi^{-1}$ for some φ, and there exists a measurable partition $\{E_k\}_k$ of the set of posteriors $\Delta(X)$ into sets such that, for every E_k with $F_1(E_k) > 0$, μ_0 contains a grain of the conditional average posterior $\overline{\mu}_1^{E_k}$ given E_k. Let K denote the indices of the cells E_k for which $F_1(E_k) > 0$. Since at most countably many of E_k have positive measure, K is a countable set. For any $k \in K$, since μ_0 contains a grain of $\overline{\mu}_1^{E_k}$, there exist some $\epsilon_k \in (0, 1]$ and some probability measure $\mu'_k \in \Delta(X)$ such that $\mu_0 = \epsilon_k \overline{\mu}_1^{E_k} + (1 - \epsilon_k)\mu'_k$. Define $\hat{S} \equiv \bigcup_{k \in K} E_k$, and note that $F_1(\hat{S}) = 1$.

I start by constructing the regular conditional probability $\mathbb{Q}(\cdot|\cdot) : \mathcal{X} \times S \to [0, 1]$ that represents the posterior about state $x \in X$ conditional on signal s. Let $\ominus \in S$ denote a signal such that $\mathbb{P}_S(\{\ominus\}) = 0$. Such a signal always exists since $S = \Delta(X)$ is uncountable, but there are at most countably many signals $s \in S$ such that $\mathbb{P}_S(\{s\}) > 0$. For any $s \in S$ such that $\varphi(s) \in \hat{S}$ and $s \neq \ominus$, set $\mathbb{Q}(D|s) = \varphi(s)(D)$ for all $D \in \mathcal{X}$. Set $\mathbb{Q}(\cdot|\ominus) = \mu'$, where μ' is an arbitrary probability measure over X if $1 - \sum_{k \in K} \epsilon_k F_1(E_k) = 0$ and is given by

$$\mu'(D) \equiv \sum_{k \in K} \frac{(1 - \epsilon_k)F_1(E_k)}{1 - \sum_{k \in K} \epsilon_k F_1(E_k)} \mu'_k(D)$$

for all $D \in \mathcal{X}$ if $1 - \sum_{k \in K} \epsilon_k F_1(E_k) > 0$. Note that μ' is always a probability measure over X since $\epsilon_k \leq 1$ for all $k \in K$ and $\sum_{k \in K} \frac{(1-\epsilon_k)F_1(E_k)}{1-\sum_{k \in K} \epsilon_k F_1(E_k)} = 1$. Finally, set $\mathbb{Q}(D|s) = \mu_0(D)$ for any $s \in S$ such that $s \neq \ominus$ and $\varphi(s) \notin \hat{S}$ and all $D \in \mathcal{X}$, indicating that the posterior equals the prior conditional on any signal realized with zero probability. Note that, by construction, the mapping $s \mapsto \mathbb{Q}(D|s)$ is measurable for any $D \in \mathcal{X}$, and $\mathbb{Q}(\cdot|s)$ is a probability distribution on $(X, \mathcal{X})$ for all $s \in S$.

I can now define the subjective distribution $\mathbb{Q}$, starting with its S-marginal distribution $\mathbb{Q}_S$. Let

$$\mathbb{Q}_S(E) \equiv \sum_{k \in K} \epsilon_k \mathbb{P}_S\left(E \cap \varphi^{-1}(E_k)\right) + \left(1 - \sum_{k \in K} \epsilon_k F_1(E_k)\right) \mathbb{1}\{\ominus \in E\}, \tag{16}$$

for all $E \in \mathcal{S}$. The fact that $\epsilon_k \leq 1$ for all $k \in K$ implies that $\sum_{k \in K} \epsilon_k F_1(E_k) \leq 1$. Therefore, $\mathbb{Q}_S(E)$ is a probability distribution over S. Next, let

$$\mathbb{Q}(D \times E) \equiv \int_E \mathbb{Q}(D|s)\mathbb{Q}_S(ds) \tag{17}$$

for all $D \in \mathcal{X}$ and $E \in \mathcal{S}$. Since the sigma-algebra $(\mathcal{X} \times \mathcal{S})$ over $(X \times S)$ is generated by sets of the form $D \times E$ with $D \in \mathcal{X}$ and $E \in \mathcal{S}$, the above expression fully specifies the probability distribution $\mathbb{Q}$. Furthermore, comparing equations (1) and (17) shows that $\mathbb{Q}(\cdot|\cdot)$ is indeed a regular conditional probability of $\mathbb{Q}$ given $\mathcal{S}$.

I next show that the true distribution $\mathbb{P}_S$ is absolutely continuous with respect to $\mathbb{Q}_S$. It suffices to show that if $\mathbb{Q}_S(E) = 0$ for some $E \in \mathcal{S}$, then also $\mathbb{P}_S(E) = 0$. Fix such a set E. Since $\mathbb{Q}_S(E) \geq \sum_{k\in K} \epsilon_k \mathbb{P}_S(E \cap \varphi^{-1}(E_k))$ and E has zero $\mathbb{Q}_S$ measure, $\mathbb{P}_S(E \cap \varphi^{-1}(E_k)) = 0$ for all $k \in K$. Therefore, $\mathbb{P}_S(E \cap \varphi^{-1}(\hat{S})) = 0$. On the other hand, $F_1(\hat{S}) = 1$ implies that $\mathbb{P}_S(\varphi^{-1}(\hat{S})) = 1$, and so, $\mathbb{P}_S(S \setminus \varphi^{-1}(\hat{S})) = 0$. Therefore,

$$\mathbb{P}_S(E) = \mathbb{P}_S(E \cap \varphi^{-1}(\hat{S})) + \mathbb{P}_S(E \cap (S \setminus \varphi^{-1}(\hat{S}))) \leq \mathbb{P}_S(E \cap \varphi^{-1}(\hat{S})) + \mathbb{P}_S(S \setminus \varphi^{-1}(\hat{S})) = 0.$$

It remains to show that $\mathbb{Q}_X = \mu_0$ and that the distribution of posteriors $\mathbb{P}_S \circ \varphi_{\mathbb{Q}}^{-1}$, defined in equation (2), coincides with the observed posterior distribution F_1. Equation (16) implies that $\mathbb{Q}_S(\varphi^{-1}(\hat{S}) \cup \{\ominus\}) = 1$. Therefore, for any $D \in \mathcal{X}$,

$$\begin{aligned}
\mathbb{Q}_X(D) &= \int_S \mathbb{Q}(D|s)\mathbb{Q}_S(ds) \\
&= \int_{\varphi^{-1}(\hat{S})\setminus\{\ominus\}} \varphi(s)(D)\mathbb{Q}_S(ds) + \mathbb{Q}_S(\{\ominus\})\mathbb{Q}(D|\ominus) \\
&= \sum_{k\in K} \epsilon_k \int_{\varphi^{-1}(E_k)} \varphi(s)(D)\mathbb{P}_S(ds) + \left(1 - \sum_{k\in K} \epsilon_k F_1(E_k)\right) \mu'(D) \\
&= \sum_{k\in K} \epsilon_k \int_{E_k} \mu(D)\mathbb{P}_S \circ \varphi^{-1}(d\mu) + \left(1 - \sum_{k\in K} \epsilon_k F_1(E_k)\right) \mu'(D) \\
&= \sum_{k\in K} \epsilon_k \int_{E_k} \mu(D)F_1(d\mu) + \left(1 - \sum_{k\in K} \epsilon_k F_1(E_k)\right) \mu'(D) \\
&= \sum_{k\in K} F_1(E_k)\epsilon_k \overline{\mu}_1^{E_k}(D) + \left(1 - \sum_{k\in K} \epsilon_k F_1(E_k)\right) \mu'(D) \\
&= \sum_{k\in K} F_1(E_k)\left(\mu_0(D) - (1-\epsilon_k)\mu'_k(D)\right) + \left(1 - \sum_{k\in K} \epsilon_k F_1(E_k)\right) \mu'(D) \\
&= \mu_0(D) - \sum_{k\in K} (1-\epsilon_k)F_1(E_k)\mu'_k(D) + \left(1 - \sum_{k\in K} \epsilon_k F_1(E_k)\right) \mu'(D).
\end{aligned}$$

If $\epsilon_k = 1$ for all $k \in K$, then the last two terms in the above display are both zero, and so, $\mathbb{Q}_X(D) = \mu_0(D)$. If, on the other hand, $\epsilon_k < 1$ for some $k \in K$, then $1 - \sum_{k\in K} \epsilon_k F_1(E_k) > 0$, and the last two terms cancel out given the definition of μ', again resulting in $\mathbb{Q}_X(D) = \mu_0(D)$.

Lastly, I show that $\mathbb{P}_S \circ \varphi_{\mathbb{Q}}^{-1} = F_1$. Since $F_1 = \mathbb{P}_S \circ \varphi^{-1}$, $F_1(\hat{S}) = 1$, and $\mathbb{P}_S(\{\ominus\}) = 0$,

$$\mathbb{P}_S(\{s \in S : \varphi(s) \in \hat{S}, s \neq \ominus\}) = 1.$$

Therefore, for any $E \in \mathcal{S}$,

$$\begin{aligned}
\mathbb{P}_S \circ \varphi_{\mathbb{Q}}^{-1}(E) &= \mathbb{P}_S\left(\{s \in S : \mathbb{Q}(\cdot|s) \in E\}\right) \\
&= \mathbb{P}_S\left(\{s \in S : \mathbb{Q}(\cdot|s) \in E, \varphi(s) \in \hat{S}, s \neq \ominus\}\right) \\
&= \mathbb{P}_S\left(\{s \in S : \varphi(s) \in E\}\right) = F_1(E).
\end{aligned}$$

This completes the proof of the first direction.

Proof of the "only if" direction. Let $\mathbb{Q}$ denote the subjective distribution on $X \times S$, and let $\mathbb{Q}(\cdot|\cdot)$ denote a regular conditional probability of $\mathbb{Q}$ given $\mathcal{S}$. Since the signal labels have no inherent meaning, I can label any signal by the posterior belief it induces. More specifically, I assume without loss of generality that $\mathbb{Q}(D|\mu) = \mu(D)$ for any $\mu \in S = \Delta(X)$.[16] Given those signal labels, $\varphi_{\mathbb{Q}}$ is the identity mapping, and $F_1 = \mathbb{P}_S \circ \varphi_{\mathbb{Q}}^{-1} = \mathbb{P}_S$.

Since $\mathbb{Q}$ satisfies condition (a) of Definition 1 and $\mathbb{Q}(\cdot|\cdot)$ is a regular conditional probability of $\mathbb{Q}$ given $\mathcal{S}$,

$$\mu_0(D) = \mathbb{Q}_X(D) = \int_S \mathbb{Q}(D|s)\mathbb{Q}_S(ds) \tag{18}$$

for all $D \in \mathcal{X}$. On the other hand, by condition (b) of Definition 1, $\mathbb{P}_S$ is absolutely continuous with respect to $\mathbb{Q}_S$. Hence, by the Radon–Nikodym theorem, there exists a Radon–Nikodym derivative $f \equiv \frac{d\mathbb{P}_S}{d\mathbb{Q}_S}$. For $k \in \mathbb{N}$, define

$$E_k \equiv \left\{s \in S : f(s) \in [k-1, k)\right\}.$$

Since f is a measurable function, E_k is a measurable subset of S for any $k \in \mathbb{N}$. Furthermore,

[16]Given a subjective distribution $\mathbb{Q}$ on $X \times S$ with regular conditional probability $\mathbb{Q}(\cdot|\cdot)$ and a true distribution for signals $\mathbb{P}_S \in \Delta(S)$, define $\tilde{\mathbb{Q}}(D|\mu) \equiv \mu(D)$, $\tilde{\mathbb{Q}}_S(E) \equiv \mathbb{Q}_S(\{s \in S : \mathbb{Q}(\cdot|s) \in E\})$, $\tilde{\mathbb{Q}}(D \times E) = \int_E \tilde{\mathbb{Q}}(D|s)\tilde{\mathbb{Q}}_S(ds)$, and $\tilde{\mathbb{P}}_S(E) \equiv \mathbb{P}_S(\{s \in S : \mathbb{Q}(\cdot|s) \in E\})$ for any $\mu \in S = \Delta(X)$ and any measurable sets $D \subseteq X$ and $E \subseteq S$. Then $\tilde{\mathbb{Q}}_X = \mathbb{Q}_X$ and $\tilde{\mathbb{P}}_S \circ \varphi_{\tilde{\mathbb{Q}}}^{-1} = \mathbb{P}_S \circ \varphi_{\mathbb{Q}}^{-1}$; that is, the prior and distribution of posteriors induced by $\tilde{\mathbb{Q}}$ and $\tilde{\mathbb{P}}_S$ coincide with those induced by $\mathbb{Q}$ and $\mathbb{P}_S$.

the sets $\{E_k\}_{k\in\mathbb{N}}$ partition the set of posteriors $S = \Delta(X)$. For any k such that $F_1(E_k) > 0$,

$$\overline{\mu}_1^{E_k} = \frac{1}{F_1(E_k)} \int_{E_k} \mu F_1(d\mu) = \frac{1}{F_1(E_k)} \int_{E_k} \mu \, \mathbb{P}_S \circ \varphi_{\mathbb{Q}}^{-1}(d\mu) = \frac{1}{F_1(E_k)} \int_{\varphi_{\mathbb{Q}}^{-1}(E_k)} \mathbb{Q}(\cdot|s)\mathbb{P}_S(ds),$$

where the last equality uses the change-of-variables formula for pushforward measures. Therefore, since $\varphi_{\mathbb{Q}}^{-1}(E_k) = E_k$,

$$\overline{\mu}_1^{E_k}(D) = \frac{1}{F_1(E_k)} \int_{E_k} \mathbb{Q}(D|s)\mathbb{P}_S(ds) \tag{19}$$

for any $D \in \mathcal{X}$. Since f is the Radon–Nikodym derivative of $\mathbb{P}_S$ with respect to $\mathbb{Q}_S$,

$$\int_{E_k} \mathbb{Q}(D|s)\mathbb{P}_S(ds) = \int_{E_k} \mathbb{Q}(D|s)f(s)\mathbb{Q}_S(ds) \le k \int_{E_k} \mathbb{Q}(D|s)\mathbb{Q}_S(ds) \le k \int_S \mathbb{Q}(D|s)\mathbb{Q}_S(ds), \tag{20}$$

where the first inequality is by the definition of set E_k, and the second inequality is due to the fact that $\int_{S\setminus E_k} \mathbb{Q}(D|s)\mathbb{Q}_S(ds) \ge 0$. Equations (18)–(20) imply

$$\overline{\mu}_1^{E_k}(D) = \frac{1}{F_1(E_k)} \int_{E_k} \mathbb{Q}(D|s)\mathbb{P}_S(ds) \le \frac{k}{F_1(E_k)} \int_S \mathbb{Q}(D|s)\mathbb{Q}_S(ds) = \frac{k}{F_1(E_k)}\mu_0(D).$$

The $k/F_1(E_k)$ constant in the above inequality is independent of D. Therefore, by Proposition 1, μ_0 contains a grain of $\overline{\mu}_1^{E_k}$. □

Proof of Proposition 2

Theorem 1 establishes (iii) $\Longrightarrow$ (i) by choosing the trivial partition, so I only need to show that (i) $\Longrightarrow$ (ii) $\Longrightarrow$ (iii).

Proof of (i) $\Longrightarrow$ (ii). Fix a posterior μ_1 in the support of F_1. I show that μ_0 contains a grain of μ_1. Let $\mathbb{Q}$ denote the subjective distribution on $X \times S$, and let $\mathbb{Q}(\cdot|\cdot)$ denote the regular conditional probability of $\mathbb{Q}$ given $\mathcal{S}$. Let $S_{\mu_1} = \{s \in S : \mathbb{Q}(\cdot|s) = \mu_1\}$ denote the set of signals that engender μ_1 as the posterior. The assumptions that the support of F_1 is finite and μ_1 is in the support of F_1 imply that $F_1(\{\mu_1\}) > 0$, and so, $\mathbb{P}_S(S_{\mu_1}) > 0$. Since $\mathbb{P}_S$ is absolutely continuous with respect to $\mathbb{Q}_S$, it must also be that $\mathbb{Q}_S(S_{\mu_1}) > 0$. Since $\mathbb{Q}(\cdot|\cdot)$

is the regular conditional probability of $\mathbb{Q}$ given $\mathcal{S}$,

$$\mathbb{Q}(D \times S_{\mu_1}) = \int_{S_{\mu_1}} \mathbb{Q}(D|s)\mathbb{Q}_S(ds) = \mu_1(D)\mathbb{Q}_S(S_{\mu_1})$$

for any $D \in \mathcal{X}$. On the other hand,

$$\mu_0(D) = \mathbb{Q}_X(D) = \mathbb{Q}(D \times S) \geq \mathbb{Q}(D \times S_{\mu_1}).$$

Therefore,

$$\mu_1(D) = \frac{\mathbb{Q}(D \times S_{\mu_1})}{\mathbb{Q}_S(S_{\mu_1})} \leq \frac{1}{\mathbb{Q}_S(S_{\mu_1})}\mu_0(D),$$

and by Proposition 1, μ_0 contains a grain of μ_1.

Proof of (ii) $\Longrightarrow$ (iii). By assumption, for any $\mu_1 \in \operatorname{supp} F_1$, there exist some $\epsilon_{\mu_1} \in (0,1]$ and some probability measure $\mu'_{\mu_1} \in \Delta(X)$ such that $\mu_0 = \epsilon_{\mu_1}\mu_1 + (1-\epsilon_{\mu_1})\mu'_{\mu_1}$. Therefore,

$$\overline{\mu}_1 = \sum_{\mu_1 \in \operatorname{supp} F_1} F_1(\{\mu_1\})\mu_1 = \sum_{\mu_1 \in \operatorname{supp} F_1} F_1(\{\mu_1\})\left(\frac{\mu_0}{\epsilon_{\mu_1}} - \frac{1-\epsilon_{\mu_1}}{\epsilon_{\mu_1}}\mu'_{\mu_1}\right),$$

and so,

$$\mu_0 = \frac{1}{\sum_{\mu_1 \in \operatorname{supp} F_1} \frac{F_1(\{\mu_1\})}{\epsilon_{\mu_1}}}\left(\overline{\mu}_1 + \sum_{\mu_1 \in \operatorname{supp} F_1} \frac{(1-\epsilon_{\mu_1})F_1(\{\mu_1\})}{\epsilon_{\mu_1}}\mu'_{\mu_1}\right). \tag{21}$$

Let $\epsilon \equiv \left(\sum_{\mu_1 \in \operatorname{supp} F_1} \frac{F_1(\{\mu_1\})}{\epsilon_{\mu_1}}\right)^{-1}$ denote the weighted harmonic mean of $\{\epsilon_{\mu_1}\}_{\mu_1 \in \operatorname{supp} F_1}$. Since all ϵ_{μ_1} are in the $(0,1]$ interval, so is ϵ. If $\epsilon = 1$, choose $\mu' \in \Delta(X)$ arbitrarily. Otherwise, let

$$\mu' \equiv \frac{\epsilon}{1-\epsilon} \sum_{\mu_1 \in \operatorname{supp} F_1} \frac{(1-\epsilon_{\mu_1})F_1(\{\mu_1\})}{\epsilon_{\mu_1}}\mu'_{\mu_1}$$

denote the weighted average of probability measures μ'_{μ_1}. Since the weights are positive and add up to one, μ' is a probability distribution over X. Equation (21) can thus be written as $\mu_0 = \epsilon\overline{\mu}_1 + (1-\epsilon)\mu'$, establishing that μ_0 contains a grain of $\overline{\mu}_1$. □

Proof of Proposition 3

Proof of the "if" direction. If $\operatorname{supp}\overline{\mu}_1 \subseteq \operatorname{supp}\mu_0$, then

$$\mu_0(\{x\}) = 0 \implies \overline{\mu}_1(\{x\}) = 0. \tag{22}$$

Define

$$c \equiv \max_{\{x\in X:\mu_0(\{x\})>0\}} \frac{\overline{\mu}_1(\{x\})}{\mu_0(\{x\})}.$$

Since X is a finite set, c is well-defined. Furthermore, since μ_0 and $\overline{\mu}_1$ are probability distributions over X, which also satisfy (22), $c \geq 1$. Therefore, for all $x \in X$,

$$\overline{\mu}_1(\{x\}) \leq c\mu_0(\{x\}),$$

where the inequality follows the definition of c for any x for which $\mu_0(\{x\}) > 0$ and follows equation (22) for other x. The above display establishes that μ_0 contains a grain of $\overline{\mu}_1$. The result then follows Theorem 1 by choosing the trivial partition of $\Delta(X)$.

Proof of the "only if" direction. Let $\mathbb{Q}$ denote the subjective distribution on $X \times S$, and let $\mathbb{Q}(\cdot|\cdot)$ denote the regular conditional probability of $\mathbb{Q}$ given $\mathcal{S}$. By the argument in the proof of the "only if" direction of Theorem 1,

$$\mu_0(D) = \int_S \mathbb{Q}(D|s)\mathbb{Q}_S(ds), \tag{23}$$

$$\overline{\mu}_1(D) = \int_S \mathbb{Q}(D|s)\mathbb{P}_S(ds). \tag{24}$$

for any $D \in \mathcal{X}$. Fix some $x \in X$ *not* in the support of μ_0, and let $\hat{S} \equiv \{s \in S : \mathbb{Q}(\{x\}|s) > 0\}$. Since $\mu_0(\{x\}) = 0$, by equation (23), $\mathbb{Q}_S(\hat{S}) = 0$. Therefore, since $\mathbb{P}_S$ is absolutely continuous with respect to $\mathbb{Q}_S$, it must be that $\mathbb{P}_S(\hat{S}) = 0$. Equation (24) then implies that $\overline{\mu}_1(\{x\}) = 0$; that is, x is not in the support of $\overline{\mu}_1$. $\square$

Proof of Proposition 4

Proof of the "if" direction. By assumption, m_0 and $\overline{m}_1$ are continuous functions and $m_0(x) > 0$ for all $x \in \operatorname{supp}\mu_0$. Therefore, $\overline{m}_1(x)/m_0(x)$ is a continuous function over the compact support of μ_0, and so, it is bounded. But since μ_0 and $\overline{\mu}_1$ have densities, the

Radon–Nikodym derivative $d\overline{\mu}_1/d\mu_0$ is equal to the ratio of densities $\overline{m}_1(x)/m_0(x)$, μ_0-almost everywhere. (Since $\operatorname{supp}\overline{\mu}_1 \subseteq \operatorname{supp}\mu_0$, the Radon–Nikodym derivative is arbitrary and irrelevant off the support of μ_0.) Thus, Proposition 1 implies that μ_0 contains a grain of $\overline{\mu}_1$. The result then follows Theorem 1 by choosing the trivial partition.

Proof of the "only if" direction. Theorem 1 implies that there exists a measurable partition of the set of posteriors $\Delta(X)$ into sets $\{E_k\}_k$ such that, for every E_k with $F_1(E_k) > 0$, the prior μ_0 contains a grain of the conditional average posterior $\overline{\mu}_1^{E_k}$ given E_k. Therefore, for every E_k with $F_1(E_k) > 0$, $\overline{\mu}_1^{E_k}$ is absolutely continuous with respect to μ_0, and consequently, $\operatorname{supp}\overline{\mu}_1^{E_k} \subseteq \operatorname{supp}\mu_0$. Let K denote the indices of the cells E_k for which $F_1(E_k) > 0$. Since at most countably many of E_k have positive measure, K is a countable set. Therefore,

$$\overline{\mu}_1 = \sum_{k\in K} F_1(E_k)\overline{\mu}_1^{E_k}.$$

Because the support of each $\overline{\mu}_1^{E_k}$ is contained in the support of μ_0 and $\overline{\mu}_1$ is their convex combination, the support of $\overline{\mu}_1$ is contained in the closure of $\bigcup_{k\in K}\operatorname{supp}\overline{\mu}_1^{E_k}$, which is contained in $\operatorname{supp}\mu_0$. □

Proof of Proposition 5

By assumption, there exists a positive F_1 measure set of posteriors H such that posteriors in H have uniformly heavier tails than μ_0. Consider an arbitrary measurable partition of the set of posteriors $\Delta(X)$ into sets $\{E_k\}_k$. Since $F_1(H) > 0$, there exists a cell E_{k^*} of the partition such that $F_1(H \cap E_{k^*}) > 0$. Since the distributions in H have uniformly heavier tails than μ_0, for every M, there exists some R such that, for all $r > R$,

$$\frac{\mu_1(\|x\| > r)}{\mu_0(\|x\| > r)} > M,$$

for all $\mu_1 \in H$. Therefore, for every M, there exists some R such that

$$\begin{aligned}\frac{\overline{\mu}_1^{E_{k^*}}(\|x\| > r)}{\mu_0(\|x\| > r)} &= \frac{1}{F_1(E_{k^*})}\int_{E_{k^*}} \frac{\mu_1(\|x\| > r)}{\mu_0(\|x\| > r)} F_1(d\mu_1)\\ &\geq \frac{1}{F_1(E_{k^*})}\int_{H\cap E_{k^*}} \frac{\mu_1(\|x\| > r)}{\mu_0(\|x\| > r)} F_1(d\mu_1)\end{aligned}$$

$$> \frac{MF_1(H \cap E_{k^*})}{F_1(E_{k^*})},$$

for any $r > R$. Hence,

$$\lim_{r\to\infty} \frac{\overline{\mu}_1^{E_{k^*}}(\|x\| > r)}{\mu_0(\|x\| > r)} = \infty.$$

That is, $\overline{\mu}_1^{E_{k^*}}$ has heavier tails than μ_0. Towards a contradiction, suppose μ_0 contains a grain of $\overline{\mu}_1^{E_{k^*}}$. Proposition 1 then implies that there exists a constant $c \geq 1$ such that $\overline{\mu}_1^{E_{k^*}}(E) \leq c\mu_0(E)$ for any measurable set E, a contradiction to the fact that $\overline{\mu}_1^{E_{k^*}}(\|x\| > r)/\mu_0(\|x\| > r)$ grows without bound as r goes to infinity. Thus, μ_0 does not contain a grain of $\overline{\mu}_1^{E_{k^*}}$. Since the partition was arbitrary and $F_1(E_{k^*}) > 0$, by Theorem 1, the pair (μ_0, F_1) is not consistent with misspecified Bayesianism. □

Proof of Theorem 2

The proof closely follows the proof of Theorem 1.

Proof of the "if" direction. Given the true distribution $\mathbb{P} \in \Delta(X \times S)$, I construct the subjective distribution $\mathbb{Q} \in \Delta(X \times S)$ that rationalizes an observed pair $(\mu_0, \{F_{1x}\}_{x\in X})$ satisfying the assumption of the theorem. By assumption, $F_{1x} = \mathbb{P}(\cdot|x) \circ \varphi^{-1}$ for some mapping φ and for all $x \in X$. Therefore, for any $E \in \mathcal{S}$,

$$\overline{F}_1(E) = \int_X F_{1x}(E)\mathbb{P}_X(dx) = \int_X \mathbb{P}\left(\varphi^{-1}(E)\middle|x\right)\mathbb{P}_X(dx) = \mathbb{P}_S\left(\varphi^{-1}(E)\right),$$

and so, $\overline{F}_1 = \mathbb{P}_S \circ \varphi^{-1}$. Furthermore, by assumption, there exists a measurable partition $\{E_k\}_k$ of the set of posteriors $\Delta(X)$ into sets such that, for every E_k with $\overline{F}_1(E_k) > 0$, μ_0 contains a grain of the conditional average posterior $\overline{\overline{\mu}}_1^{E_k}$ given E_k. Let K denote the indices of the cells E_k for which $\overline{F}_1(E_k) > 0$. Since at most countably many of E_k have positive measure, K is a countable set. For any $k \in K$, since μ_0 contains a grain of $\overline{\overline{\mu}}_1^{E_k}$, there exist some $\epsilon_k \in (0,1]$ and some probability measure $\mu_k' \in \Delta(X)$ such that $\mu_0 = \epsilon_k\overline{\overline{\mu}}_1^{E_k} + (1-\epsilon_k)\mu_k'$. Define $\hat{S} \equiv \bigcup_{k\in K} E_k$, and note that $\overline{F}_1(\hat{S}) = 1$.

I construct $\mathbb{Q}$ as in the proof of the "if" direction of Theorem 1 with F_1 replaced by $\overline{F}_1$ throughout. In particular, I let $\ominus \in S$ denote a signal such that $\mathbb{P}_S(\{\ominus\}) = 0$; for any $s \in S$ such that $\varphi(s) \in \hat{S}$ and $s \neq \ominus$, I set $\mathbb{Q}(D|s) = \varphi(s)(D)$ for all $D \in \mathcal{X}$; I set $\mathbb{Q}(\cdot|\ominus) = \mu'$, where

μ' is an arbitrary probability measure over X if $1 - \sum_{k \in K} \epsilon_k \overline{F}_1(E_k) = 0$ and is given by

$$\mu'(D) \equiv \sum_{k \in K} \frac{(1-\epsilon_k)\overline{F}_1(E_k)}{1 - \sum_{k \in K} \epsilon_k \overline{F}_1(E_k)} \mu'_k(D)$$

for all $D \in \mathcal{X}$ if $1 - \sum_{k \in K} \epsilon_k \overline{F}_1(E_k) > 0$; I set $\mathbb{Q}(D|s) = \mu_0(D)$ for any $s \in S$ such that $s \neq \ominus$ and $\varphi(s) \notin \hat{S}$ and all $D \in \mathcal{X}$; and I let

$$\mathbb{Q}_S(E) \equiv \sum_{k \in K} \epsilon_k \mathbb{P}_S\left(E \cap \varphi^{-1}(E_k)\right) + \left(1 - \sum_{k \in K} \epsilon_k \overline{F}_1(E_k)\right) \mathbb{1}\{\ominus \in E\}, \tag{25}$$

for all $E \in \mathcal{S}$ and let

$$\mathbb{Q}(D \times E) \equiv \int_E \mathbb{Q}(D|s)\mathbb{Q}_S(ds) \tag{26}$$

for all $D \in \mathcal{X}$ and $E \in \mathcal{S}$.

By an identical argument to that in the proof of Theorem 1, $\mathbb{P}_S$ is absolutely continuous with respect to $\mathbb{Q}_S$ and $\mathbb{Q}_X = \mu_0$. I only need to show that $\mathbb{P}(\cdot|x) \circ \varphi_{\mathbb{Q}}^{-1} = F_{1x}$ for $\mathbb{P}_X$-almost all x. By definition, $F_{1x} = \mathbb{P}(\cdot|x) \circ \varphi^{-1}$ for all x. On the other hand, since $\mathbb{P}_S(\varphi^{-1}(\hat{S}) \setminus \{\ominus\}) = 1$, for $\mathbb{P}_X$-almost all x,

$$\mathbb{P}(\varphi^{-1}(\hat{S}) \setminus \{\ominus\}|x) = 1.$$

Therefore, for any $E \in \mathcal{S}$ and $\mathbb{P}_X$-almost all x,

$$\begin{aligned}
\left(\mathbb{P}(\cdot|x) \circ \varphi_{\mathbb{Q}}^{-1}\right)(E) &= \mathbb{P}\left(\{s \in S : \mathbb{Q}(\cdot|s) \in E\}|x\right) \\
&= \mathbb{P}\left(\{s \in S : \mathbb{Q}(\cdot|s) \in E, \varphi(s) \in \hat{S}, s \neq \ominus\}|x\right) \\
&= \mathbb{P}\left(\{s \in S : \varphi(s) \in E\}|x\right) = F_{1x}(E).
\end{aligned}$$

This completes the proof of the first direction.

Proof of the "only if" direction. Let $\mathbb{Q}$ denote the subjective distribution on $X \times S$, and let $\mathbb{Q}(\cdot|\cdot)$ denote a regular conditional probability of $\mathbb{Q}$ given $\mathcal{S}$. Since the signal labels have no inherent meaning, I assume without loss of generality that $\mathbb{Q}(D|\mu) = \mu(D)$ for any $\mu \in S = \Delta(X)$. Given those signal labels, $\varphi_{\mathbb{Q}}$ is the identity mapping, and $F_{1x} = \mathbb{P}(\cdot|x) \circ \varphi_{\mathbb{Q}}^{-1} = \mathbb{P}(\cdot|x)$. Therefore,

$$\overline{F}_1 \equiv \int_X F_{1x}\mathbb{P}_X(dx) = \int_X \mathbb{P}(\cdot|x)\mathbb{P}_X(dx) = \mathbb{P}_S.$$

Since $\mathbb{Q}$ satisfies condition (a) of Definition 6 and $\mathbb{Q}(\cdot|\cdot)$ is a regular conditional probability of $\mathbb{Q}$ given $\mathcal{S}$,

$$\mu_0(D) = \mathbb{Q}_X(D) = \int_S \mathbb{Q}(D|s)\mathbb{Q}_S(ds) \tag{27}$$

for all $D \in \mathcal{X}$. On the other hand, by condition (b) of Definition 6, $\mathbb{P}_S$ is absolutely continuous with respect to $\mathbb{Q}_S$. Hence, by the Radon–Nikodym theorem, there exists a Radon–Nikodym derivative $f \equiv \frac{d\mathbb{P}_S}{d\mathbb{Q}_S}$. For $k \in \mathbb{N}$, define

$$E_k \equiv \left\{s \in S : f(s) \in [k-1, k)\right\}.$$

Since f is a measurable function, E_k is a measurable subset of S for any $k \in \mathbb{N}$. Furthermore, the sets $\{E_k\}_{k\in\mathbb{N}}$ partition the set of posteriors $S = \Delta(X)$. For any k such that $\overline{F}_1(E_k) > 0$,

$$\overline{\overline{\mu}}_1^{E_k} = \frac{1}{\overline{F}_1(E_k)} \int_{E_k} \mu \overline{F}_1(d\mu) = \frac{1}{\overline{F}_1(E_k)} \int_{E_k} \mu \, \mathbb{P}_S \circ \varphi_{\mathbb{Q}}^{-1}(d\mu) = \frac{1}{\overline{F}_1(E_k)} \int_{\varphi_{\mathbb{Q}}^{-1}(E_k)} \mathbb{Q}(\cdot|s)\mathbb{P}_S(ds).$$

Therefore, since $\varphi_{\mathbb{Q}}^{-1}(E_k) = E_k$,

$$\overline{\overline{\mu}}_1^{E_k}(D) = \frac{1}{\overline{F}_1(E_k)} \int_{E_k} \mathbb{Q}(D|s)\mathbb{P}_S(ds) \tag{28}$$

for any $D \in \mathcal{X}$. Since f is the Radon–Nikodym derivative of $\mathbb{P}_S$ with respect to $\mathbb{Q}_S$,

$$\int_{E_k} \mathbb{Q}(D|s)\mathbb{P}_S(ds) = \int_{E_k} \mathbb{Q}(D|s)f(s)\mathbb{Q}_S(ds) \le k \int_{E_k} \mathbb{Q}(D|s)\mathbb{Q}_S(ds) \le k \int_S \mathbb{Q}(D|s)\mathbb{Q}_S(ds), \tag{29}$$

where the first inequality is by the definition of set E_k, and the second inequality is due to the fact that $\int_{S\setminus E_k} \mathbb{Q}(D|s)\mathbb{Q}_S(ds) \ge 0$. Equations (27)–(29) imply

$$\overline{\overline{\mu}}_1^{E_k}(D) = \frac{1}{\overline{F}_1(E_k)} \int_{E_k} \mathbb{Q}(D|s)\mathbb{P}_S(ds) \le \frac{k}{\overline{F}_1(E_k)} \int_S \mathbb{Q}(D|s)\mathbb{Q}_S(ds) = \frac{k}{\overline{F}_1(E_k)} \mu_0(D).$$

The $k/\overline{F}_1(E_k)$ constant in the above inequality is independent of D. Therefore, by Proposition 1, μ_0 contains a grain of $\overline{\overline{\mu}}_1^{E_k}$. □

Proof of Theorem 3

Proof of the "if" direction. The proof of this direction is constructive. I define the regular conditional probability $\mathbb{Q}(\cdot|\cdot) : \mathcal{X} \times S \to [0,1]$ by setting $\mathbb{Q}(D|s) = \varphi(s)(D)$ for all $s \in S$ and all $D \in \mathcal{X}$. By construction, $\mathbb{Q}(\cdot|s)$ is a probability distribution on $(X, \mathcal{X})$ for any s, and the mapping $s \mapsto \mathbb{Q}(D|s)$ is measurable for all $D \in \mathcal{X}$. I set the S-marginal $\mathbb{Q}_S$ of the subjective distribution equal to the true distribution $\mathbb{P}_S$ of signals and define $\mathbb{Q}$ as in (17) with $\mathbb{Q}_S = \mathbb{P}_S$. By construction, $\mathbb{Q}(\cdot|\cdot)$ is a regular conditional probability of $\mathbb{Q}$ given $\mathcal{S}$. Next, note that, for any $D \in \mathcal{X}$,

$$\mathbb{Q}_X(D) = \int_S \mathbb{Q}(D|s)\mathbb{P}_S(ds) = \int_S \mu(D)F_1(d\mu) = \overline{\mu}_1(D) = \mu_0(D),$$

where the last equality follows the assumption that $\overline{\mu}_1 = \mu_0$. Moreover, by an argument similar to the one in the proof of Theorem 1,

$$\mathbb{P}_S \circ \varphi_{\mathbb{Q}}^{-1}(E) = \mathbb{P}_S\left(\{s \in S : \mathbb{Q}(\cdot|s) \in E\}\right) = \mathbb{P}_S\left(\{s \in S : \varphi(s) \in E\}\right) = F_1(E)$$

for all $E \in \mathcal{S}$. This shows that the subjective distribution $\mathbb{Q}$ constructed above rationalizes the observed pair (μ_0, F_1).

Proof of the "only if" direction. By the argument in the proof of the "only if" direction of Theorem 1, $\mu_0(D) = \int_S \mathbb{Q}(D|s)\mathbb{Q}_S(ds)$ and $\overline{\mu}_1(D) = \int_S \mathbb{Q}(D|s)\mathbb{P}_S(ds)$ for any $D \in \mathcal{X}$. The assumption that $\mathbb{Q}_S = \mathbb{P}_S$ completes the proof. □

Proof of Theorem 4

Proof of (i) $\implies$ (ii). Suppose (μ_0, F_1) is consistent with Bayesianism given a subjective distribution $\mathbb{Q}$ with an S-marginal $\mathbb{Q}_S$ that is equivalent to $\mathbb{P}_S$, and let $\mathbb{Q}(\cdot|\cdot)$ denote the regular conditional probability of $\mathbb{Q}$ given $\mathcal{S}$. I define $\lambda \in \Delta(S) = \Delta(\Delta(X))$ as follows:

$$\lambda(E) \equiv \mathbb{Q}_S(\{s \in S : \mathbb{Q}(\cdot|s) \in E\})$$

for all $E \in \mathcal{S}$. I next show that λ and F_1 are equivalent and $\mu_0 = \int \mu\lambda(d\mu)$. Since $\mathbb{Q}$ satisfies condition (a) and $\mathbb{Q}(\cdot|\cdot)$ is a regular conditional probability of $\mathbb{Q}$ given $\mathcal{S}$,

$$\mu_0 = \mathbb{Q}_X = \int_S \mathbb{Q}(\cdot|s)\mathbb{Q}_S(ds) = \int_S \mu\lambda(d\mu),$$

where the last equality is by the definition of λ. On the other hand, for all $E \in \mathcal{S}$,

$$\lambda(E) = \mathbb{Q}_S(\{s \in S : \mathbb{Q}(\cdot|s) \in E\})$$

and

$$F_1(E) = \mathbb{P}_S \circ \varphi_{\mathbb{Q}}^{-1}(E) = \mathbb{P}_S(\{s \in S : \mathbb{Q}(\cdot|s) \in E\}),$$

where the first equality above is by condition (c) of Definition 1. Since $\mathbb{Q}_S$ and $\mathbb{P}_S$ are equivalent, so are λ and F_1.

Proof of (ii) $\implies$ (i). Suppose there exists a probability measure $\lambda \in \Delta(\Delta(X))$ such that λ and F_1 are equivalent and $\mu_0 = \int \mu\lambda(d\mu)$. By the Radon–Nikodym theorem, there are derivatives $f \equiv \frac{d\lambda}{dF_1}$ and $\frac{1}{f} \equiv \frac{dF_1}{d\lambda}$. Set $\mathbb{Q}(D|s) = \varphi(s)(D)$ for all $s \in S$ and $D \in \mathcal{X}$, and set $\mathbb{Q}_S(ds) = f(\varphi(s))\mathbb{P}_S(ds)$. I need to show that $\mathbb{Q}_S$, as defined above, is indeed a probability distribution on $(S, \mathcal{S})$. By construction, $\mathbb{Q}_S(E) \geq 0$ for all $E \in \mathcal{S}$, and $\mathbb{Q}_S(\emptyset) = 0$. Next, note that

$$\int_S \mathbb{Q}_S(ds) = \int_S f(\varphi(s))\mathbb{P}_S(ds) = \int_S f(\mu)\mathbb{P}_S \circ \varphi^{-1}(d\mu) = \int_S f(\mu)F_1(d\mu) = \int_S \lambda(d\mu) = 1,$$

where the first equality is by definition, the second one uses the change-of-variables formula for pushforward measures, the third equality is due to the fact that $F_1 = \mathbb{P}_S \circ \varphi^{-1}$, the fourth one uses the definition of f, and the last equality is because λ is a probability measure on S. Finally, $\mathbb{Q}_S$ is countably additive since $\mathbb{P}_S$ is countably additive. Therefore, $\mathbb{Q}_S$ is a well-defined probability distribution. I finish the construction by defining $\mathbb{Q}$ as in equation (17). Note that, by construction, $\mathbb{Q}(\cdot|\cdot)$ is a regular conditional probability of $\mathbb{Q}$ given $\mathcal{S}$. Furthermore, by an argument similar to the one in the above display,

$$\mathbb{Q}_X = \int_S \mathbb{Q}(\cdot|s)\mathbb{Q}_S(ds) = \int_S \varphi(s)f(\varphi(s))\mathbb{P}_S(ds) = \int_S \mu f(\mu)F_1(d\mu) = \int_S \mu\lambda(d\mu) = \mu_0,$$

where the last equality is by assumption. Therefore, condition (a) of Definition 1 is satisfied. Furthermore, since $\mathbb{Q}_S(ds) = f(\varphi(s))\mathbb{P}_S(ds)$ and $\mathbb{P}_S(ds) = \frac{1}{f(\varphi(s))}\mathbb{Q}_S(ds)$, probability distributions $\mathbb{Q}_S$ and $\mathbb{P}_S$ are equivalent. That is, condition (b) of Definition 1 is satisfied, and $\mathbb{Q}_S$ is absolutely continuous with respect to $\mathbb{P}_S$. On the other hand,

$$\mathbb{P}_S \circ \varphi_{\mathbb{Q}}^{-1}(E) = \mathbb{P}_S\left(\{s \in S : \mathbb{Q}(\cdot|s) \in E\}\right) = \mathbb{P}_S\left(\{s \in S : \varphi(s) \in E\}\right) = \mathbb{P}_S \circ \varphi^{-1}(E) = F_1(E)$$

for all $E \in \mathcal{S}$, implying that condition (c) is also satisfied. □

References

Aumann, Robert J. and Michael Maschler (1995): *Repeated Games with Incomplete Information*, MIT press.

Blackwell, David and Lester Dubins (1962): "Merging of Opinions with Increasing Information," *The Annals of Mathematical Statistics*, 33 (3), 882–886.

Bohren, J. Aislinn (2016): "Informational Herding with Model Misspecification," *Journal of Economic Theory*, 163, 222–247.

Bohren, J. Aislinn and Daniel N. Hauser (2023): "Behavioral Foundations of Model Misspecification," Working Paper.

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer (2020): "Overreaction in Macroeconomic Expectations," *American Economic Review*, 110 (9), 2748–2782.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer (2018): "Diagnostic Expectations and Credit Cycles," *The Journal of Finance*, 73 (1), 199–227.

Coibion, Olivier and Yuriy Gorodnichenko (2015): "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts," *American Economic Review*, 105 (8), 2644–78.

Cripps, Martin W. (2019): "Divisible Updating," Working Paper.

Epstein, Larry G. and Michel Le Breton (1993): "Dynamically Consistent Beliefs Must Be Bayesian," *Journal of Economic Theory*, 61 (1), 1–22.

Epstein, Larry G., Jawwad Noor, and Alvaro Sandroni (2010): "Non-Bayesian Learning," *The BE Journal of Theoretical Economics*, 10 (1), 1–20.

Esponda, Ignacio and Demian Pouzo (2016): "Berk–Nash Equilibrium: A Framework for Modeling Agents with Misspecified Models," *Econometrica*, 84 (3), 1093–1130.

——— (2021): "Equilibrium in Misspecified Markov Decision Processes," *Theoretical Economics*, 16 (2), 717–757.

Faden, Arnold M. (1985): "The Existence of Regular Conditional Probabilities: Necessary and Sufficient Conditions," *Annals of Mathematical Statistics*, 13 (1), 288–298.

Frick, Mira, Ryota Iijima, and Yuhta Ishii (2020): "Misinterpreting Others and the Fragility of Social Learning," *Econometrica*, 88 (6), 2281–2328.

Fudenberg, Drew, Giacomo Lanzani, and Philipp Strack (2021): "Limit Points of Endogenous Misspecified Learning," *Econometrica*, 89 (3), 1065–1098.

Kalai, Ehud and Ehud Lehrer (1993): "Rational Learning Leads to Nash Equilibrium," *Econometrica*, 61 (5), 1019–1045.

Kamenica, Emir and Matthew Gentzkow (2011): "Bayesian Persuasion," *American Economic Review*, 101 (6), 2590–2615.

Molavi, Pooya, Alireza Tahbaz-Salehi, and Andrea Vedolin (2024): "Model Complexity, Expectations, and Asset Prices," *Review of Economic Studies*, 91 (4), 2462–2507.

——— (2025): "Expectations and the Term Structure of Interest Rates," *Working paper*.

Ortoleva, Pietro (2022): "Alternatives to Bayesian Updating," *Annual Review of Economics*, 15, 1–28.

Rabin, Matthew and Joel L. Schrag (1999): "First Impressions Matter: A Model of Confirmatory Bias," *The Quarterly Journal of Economics*, 114 (1), 37–82.

Savage, Leonard J. (1972): *The Foundations of Statistics*, Courier Corporation.

Shmaya, Eran and Leeat Yariv (2016): "Experiments on Decisions Under Uncertainty: A Theoretical Framework," *American Economic Review*, 106 (7), 1775–1801.

Tversky, Amos and Daniel Kahneman (1974): "Judgment Under Uncertainty: Heuristics and Biases," *Science*, 185 (4157), 1124–1131.