Online Appendix to "Coarse Bayesian Updating"

Alexander M. Jakobsen*

November 2, 2025

Abstract

This appendix provides additional results for Jakobsen (2025). Section 1 contrasts Coarse Bayesian updating with a class of maximum-likelihood updating rules and the categorical-thinking model of Mullainathan (2002). Section 2 provides some basic results on observational learning and the identification of Coarse Bayesian Representations from state-contingent choice data.

1 Paradigm Shifts and Maximum-Likelihood Updating

In the competing-theories interpretation of the model, the agent employs subjective thresholds (the partition) for switching among candidate beliefs. It is natural to wonder if such behavior can be reformulated in terms of second-order beliefs. If an agent assigns a prior degree of confidence to each feasible theory, can Coarse Bayesian updating be reconciled with Bayesian updating of such second-order beliefs?

To answer this question, I take an approach similar to that of Ortoleva (2012), who introduces the Hypothesis-Testing (HT) model. Under HT, an agent applies Bayes' rule for signals of sufficiently high prior likelihood (that is, above some threshold $\varepsilon \geq 0$, an individual parameter). For unexpected signals (likelihood less than ε), the agent experiences a "paradigm shift" and updates beliefs by applying a maximum-likelihood criterion to a *second-order prior*, or "prior over priors." Specifically, the agent updates the second-order prior via Bayes' rule, then adopts as posterior a belief of maximal probability under the

*Kellogg MEDS, Northwestern University.

revised second-order distribution. In this section, I consider a similar maximum-likelihood procedure, adapted to the domain S of noisy signals.[1]

Definition 7. A Homogeneous, Convex updating rule μ has a **Maximum-Likelihood (ML) Representation** if there exists a probability distribution Γ over Δ (with density γ) such that, for all $s \in S$,

$$\mu^s \in \operatorname*{argmax}_{\hat{\mu} \in \Delta} \gamma(\hat{\mu}) \hat{\mu} \cdot s.$$

The function $L : \Delta \times S \to \mathbb{R}$ given by $L(\hat{\mu}|s) = \gamma(\hat{\mu})\hat{\mu} \cdot s$ is the **likelihood function**.

In a Maximum-Likelihood Representation, the agent has a second-order prior Γ that he updates, via Bayes' rule, upon arrival of signal s. Then, he selects a belief of maximal probability under the new second-order distribution. This procedure selects among beliefs $\hat{\mu}$ that maximize the likelihood function at s.[2] Intuitively, ML updating captures the behavior of an agent who assigns prior degrees of confidence to competing theories, updates these values in a Bayesian fashion, and selects the most-likely theory given available information.[3]

Proposition 11.

(i) Not every Maximum-Likelihood rule can be expressed as a Coarse Bayesian rule.

(ii) Not every Coarse Bayesian rule can be expressed as a Maximum-Likelihood rule.

(iii) If $N = 2$, then every Coarse Bayesian rule is a Maximum-Likelihood rule.

(iv) Bayesian updating is a special case of both Coarse Bayesian and Maximum-Likelihood updating. To express Bayesian updating as a Maximum-Likelihood rule, take

$$\gamma(\hat{\mu}) \propto \left\| \frac{\hat{\mu}}{\sqrt{\mu^e}} \right\|^{-1} \tag{1}$$

where $\sqrt{\mu^e} := (\sqrt{\mu^e_\omega})_{\omega \in \Omega}$.

Proposition 11 establishes that neither updating procedure subsumes the other—there

[1]As Weinstein (2017) explains, the HT model allows essentially any updating to occur for unexpected news (ie, likelihood less than ε). As we shall see, extending maximum-likelihood updating procedures to the domain of noisy signals does rule out some updating behavior.

[2]Notice that L is homogeneous (of degree 1) and convex in s. The restriction to Homogeneous, Convex updating rules, therefore, only takes effect when there are ties—multiple candidate beliefs that maximize L.

[3]There are other ways of reducing a second-order belief to a first-order belief. For example, one might use the second-order distribution to compute an average belief. However, such a procedure is continuous in s while Coarse Bayesian updating, in general, exhibits discontinuities in s.

exist updating rules that have Coarse Bayesian Representations but not ML Representations, and there exist updating rules that have ML Representations but not Coarse Bayesian Representations. These claims are demonstrated by Examples 4 and 5 below. Part (iii) establishes an important special case: if there are only two states, then every Coarse Bayesian rule can be expressed as a ML rule. Part (iv) asserts that Bayesian updating is a special case of both models and provides an explicit formula for a second-order prior generating Bayesian updating in the ML procedure. For proof of claims (iii) and (iv), see the appendix.

Example 4. Not every ML rule can be expressed as a Coarse Bayesian rule. Suppose $N = 2$ and consider the distribution γ such that $\gamma(\mu^1) = 3/4$ and $\gamma(\mu^2) = 1/4$, where $\mu^1 = (1/3, 2/3)$ and $\mu^2 = (3/4, 1/4)$. Observe that $L(\mu^1|e) = \gamma(\mu^1)\mu^1 \cdot e = \gamma(\mu^1) > \gamma(\mu^2) = \gamma(\mu^2)\mu^2 \cdot e = L(\mu^2|e)$; thus, $\mu^e = \mu^1$. It is easy to verify that $B(\mu^e|s) = \mu^2$ if and only if $s_1/s_2 = 6$. Therefore, to be consistent with a Coarse Bayesian updating rule, we must have $L(\mu^2|s) \geq L(\mu^1|s)$ whenever $s_1/s_2 = 6$. Take $s = (1, 1/6)$. Then $L(\mu^2|s) = 19/96 < 19/72 = L(\mu^1|s)$, so that the ML rule selects μ^1 at s. This means the rule is not Confirmatory, and therefore is inconsistent with Coarse Bayesian updating.

Example 5. Not every Coarse Bayesian rule can be expressed as a ML rule. Suppose $N = 3$ and consider a Coarse Bayesian Representation where $\mathcal{P}$ has two cells, P and P', with $\mu^P = \mu^e$ and $\mu^{P'} = \mu' \neq \mu^e$. The boundary between P and P' corresponds to a hyperplane, H, in S. We will choose H (hence, $\mathcal{P}$) in such a way that no distribution γ on Δ (with support $\{\mu^e, \mu'\}$) generates the same updating behavior as $\langle \mathcal{P}, \mu^{\mathcal{P}} \rangle$ under the ML procedure.

Observe that if γ generates the same updating behavior, then $L(\mu^e|s) = L(\mu'|s)$ for all $s \in H$. In particular, $[\gamma(\mu^e)\mu^e - \gamma(\mu')\mu'] \cdot s = 0$ for all $s \in H$. Therefore, to be consistent with ML updating, the normal vector for H must be (a scalar multiple of) a member of the set $\{\lambda\mu^e - (1-\lambda)\mu' : 0 < \lambda < 1\}$; the span of this set is a 2-dimensional subset of $\mathbb{R}^3$. Thus, we may perturb the hyperplane H so that its normal does not belong to the required set.

As demonstrated by Example 4 above, ML updating rules may be incompatible with Coarse Bayesian updating due to violations of Confirmation: ML rules are measurable with respect to some partition of Δ into convex cells, but cells need not contain their representative elements. Below, I show that the categorical-thinking model of Mullainathan (2002) also violates Confirmation in some cases, and for a similar reason. Rather than employing a second-order prior to compute likelihoods and select posteriors, Mullainathan's model uses a particular formula to calculate "base rates" for candidate beliefs. Thus, the categorical-thinking model is similar in spirit to a ML procedure, and the particular functional form employed can produce violations of Confirmation.

In contrast, Example 5 shows that with three or more states, Coarse Bayesian Representations that can be represented as ML rules are unstable in that small perturbations of the updating rule render a ML representation impossible. This is so because ML representations impose rigid constraints on how representative beliefs may be positioned relative to cell boundaries. Thus, small perturbations to the boundaries or representative points render the rule incompatible with ML updating.

1.1 Relationship to Mullainathan (2002)

In a working paper, Mullainathan (2002) develops a model of categorical thinking sharing several features of Coarse Bayesian updating. In this section, I show that the categorical thinking model (adapted to my framework of states and signals) satisfies Homogeneity and Cognizance but not necessarily Confirmation.

Mullainathan works with a type space T and prior p where $p(t)$ is the prior probability of type $t \in T$. The analogous components in my model are the state space Ω and prior μ^e. Data d in Mullainathan's model is expressed as conditional probabilities $p(d|t)$ indicating the probability of observing the data given type t; in my model, data corresponds to a signal realization s, and s_ω plays the role of $p(d|t)$.

In Mullainathan's model, a set C of probability distributions over T constitutes a set of "categories"; these are feasible beliefs the agent can hold. Thus, the set C is analogous to the set $\{\mu^P : P \in \mathcal{P}\}$ in my model. For a category c and data d, $p(d|c)$ is the probability of generating data d in category c; this is analogous to $s \cdot \mu^P$, which is the probability of observing signal s if μ^P is the true probability law. Finally, Mullainathan defines $p(c) := \int_t p(t)c(t)$ to be the "base rate" of category c.[4] In my model, the analogous rate is $\mu^e \cdot \mu^P$.

Like Coarse Bayesians, agents in Mullainathan's model partition the probability simplex and assign posterior beliefs as a function of the cell containing the Bayesian posterior. Any set C of categories is permitted; however, the partition is derived from C using an optimality criterion resembling that of Maximum-Likelihood rules analyzed above. In particular, let $c^*(d) \in C$ denote the agent's posterior after observing data d. Mullainathan requires that

$$c^*(d) \in \operatorname*{argmax}_{c \in C} p(d|c)p(c). \tag{2}$$

[4] I have modified the notation slightly; Mullainathan writes $q_c(\cdot)$ instead of $c(\cdot)$ to indicate the probability distribution over T associated with category $c \in C$.

In my framework, the analogous condition is

$$\mu^s \in \underset{\hat{\mu}\in\hat{C}}{\text{argmax}}(s \cdot \hat{\mu})(\mu^e \cdot \hat{\mu}), \tag{3}$$

where $\hat{C} \subseteq \Delta$ is some set of feasible posteriors. This is very similar to maximization of the likelihood function specified in Definition 7; the main difference is that my likelihood functions use a second-order belief γ instead of the base rate $p(c)$ proposed by Mullainathan.

Thus, Mullainathan's model works by specifying a set C of categories (feasible posteriors) from which criterion (2) selects posteriors after observing data d. Due to the functional forms employed, it is as if there is a partition of the probability simplex such that the agent's selected posterior only depends on which cell contains the Bayesian posterior.

Unlike Coarse Bayesians, categorical thinkers need not satisfy Confirmation because condition (2) does not guarantee that beliefs $c^*(d)$ belong to the cell containing the Bayesian posterior associated with data d.[5] Below, I prove these claims in my framework (in particular, employing condition (3)).

First, let $\hat{C}$ be a nonempty set of feasible posteriors. Suppose that some $\mu^* \in \hat{C}$ is a solution to the maximization problem in (3) for both s and t. That is, μ^* solves both

$$\max_{\hat{\mu}\in\hat{C}}(s \cdot \hat{\mu})(\mu^e \cdot \hat{\mu}) \qquad \text{and} \qquad \max_{\hat{\mu}\in\hat{C}}(t \cdot \hat{\mu})(\mu^e \cdot \hat{\mu}).$$

Then, if $\alpha, \beta \geq 0$, it follows at μ^* solves $\max_{\hat{\mu}\in\hat{C}}((\alpha s + \beta t) \cdot \hat{\mu})(\mu^e \cdot \hat{\mu})$. Consequently, the map $s \mapsto \text{argmax}_{\hat{\mu}\in\hat{C}}(s \cdot \hat{\mu})(\mu^e \cdot \hat{\mu})$ is measurable with respect to a partition of S into convex cones. As demonstrated in the proof of Theorem 1, such convex cones can be associated with convex subsets of Δ by mapping signals s to Bayesian posteriors $B(\mu^e|s)$.

Thus, any updating rule satisfying (3) satisfies Homogeneity and Cognizance if one restricts attention to signals that yield unique solutions to the optimization problem. For signals that involve ties, Homogeneity and/or Cognizance may be violated if the agent's tie-breaking selection is not Homogeneous or Convex.

A more substantive difference between Mullainathan's model and Coarse Bayesian updating is that condition (3) does not guarantee that the updating rule satisfies Confirmation. To see this, suppose $|\Omega| = 2$ and let $\mu^e = (\frac{1}{3}, \frac{2}{3})$. Suppose $\hat{\mu}, \hat{\mu}' \in \hat{C}$ where $\hat{\mu} = (\frac{1}{4}, \frac{3}{4})$ and $\hat{\mu}' = (\frac{1}{5}, \frac{4}{5})$. Let $s = (\frac{3}{8}, \frac{9}{16})$. It follows that $B(\mu^e|s) = \hat{\mu}$; so, Confirmation requires $\hat{\mu}$ to solve $\max_{\tilde{\mu}\in\hat{C}}(s \cdot \tilde{\mu})(\mu^e \cdot \tilde{\mu})$. However, $(s \cdot \hat{\mu})(\mu^e \cdot \hat{\mu}) = \frac{77}{256} < \frac{63}{200} = (s \cdot \hat{\mu}')(\mu^e \cdot \hat{\mu}')$. Thus, $\hat{\mu}$ is not selected at s, violating Confirmation.

[5] Note that the partition in Mullainathan's model typically has convex cells. Convexity fails only if the maximization problem in (2) has more than one solution and the agent's tie-breaking criterion is not convex.

2 Observational Learning

What can be learned by observing a Coarse Bayesian? This section takes the position of an outside analyst who observes the behavior of a Coarse Bayesian and uses these observations to make inferences about the state of the world.

In such contexts, learning about the state is of course highly dependent on knowing or learning the representation $\langle \mathcal{P}, (\mu^P)_{P\in\mathcal{P}}\rangle$ of the Coarse Bayesian. The main text establishes that observing the Coarse Bayesian's beliefs μ^s at arbitrary signals s is sufficient to identify the representation. The main insights of this section involve inference when the signal s observed by the Coarse Bayesian is not observed by the analyst. It will, however, be useful to begin with a preliminary result establishing that signal-contingent *choices* (not posteriors) of the Coarse Bayesian are sufficient to pin down $\langle \mathcal{P}, (\mu^P)_{P\in\mathcal{P}}\rangle$.

Given an updating rule $\mu : S \to \Delta$, a signal $s \in S$ and a menu $A \in \mathcal{A}$, let

$$c^s(A) := \operatorname*{argmax}_{x\in A} x \cdot \mu^s.$$

Thus, the correspondence $c^s : \mathcal{A} \to \mathcal{A}$ records optimal choices from menus A conditional on signal s. For menus A, B such that $B \subseteq A$, let $S(B|A) := \{s \in S : c^s(A) = B\}$; that is, $S(B|A)$ consists of all signals s that make B the set of optimal actions in A. If $\dot{\mu}$ is another updating rule, the associated sets are denoted $\dot{S}(B|A)$.

Proposition 12. *Let $\langle \mathcal{P}, (\mu^P)_{P\in\mathcal{P}}\rangle$ and $\langle \mathcal{Q}, (\dot{\mu}^Q)_{Q\in\mathcal{Q}}\rangle$ be Coarse Bayesian Representations of updating rules μ and $\dot{\mu}$, respectively, such that $\mu^e = \dot{\mu}^e$. Then $\langle \mathcal{P}, (\mu^P)_{P\in\mathcal{P}}\rangle = \langle \mathcal{Q}, (\dot{\mu}^Q)_{Q\in\mathcal{Q}}\rangle$ if and only if, for all $A, B \in \mathcal{A}$, $S(B|A) = \dot{S}(B|A)$.*

Proof. Clearly, $\langle \mathcal{P}, (\mu^P)_{P\in\mathcal{P}}\rangle = \langle \mathcal{Q}, (\dot{\mu}^Q)_{Q\in\mathcal{Q}}\rangle$ implies $S(B|A) = \dot{S}(B|A)$ for all $A, B \in \mathcal{A}$. If $\langle \mathcal{P}, (\mu^P)_{P\in\mathcal{P}}\rangle \neq \langle \mathcal{Q}, (\dot{\mu}^Q)_{Q\in\mathcal{Q}}\rangle$, there is a signal s such that $\mu^s \neq \dot{\mu}^s$. Thus, there is a hyperplane in $\mathbb{R}^N$ that strictly separates μ^s and $\dot{\mu}^s$, which immediately implies there is a binary menu $A = \{x, y\}$ such that $c^s(A) = \{x\}$ and $\dot{c}^s(A) = \{y\}$; that is, $S(\{x\}|A) \neq \dot{S}(\{x\}|A)$. $\square$

By Proposition 12, a Coarse Bayesian Representation is pinned down by observing signal-contingent action choices; crucially, this requires variation in the menu A, as (for example) singleton menus A reveal nothing about behavior. The result implies that if two Coarse Bayesians differ in any way—different cells, different representative points, or both—there will be a menu A (in fact, a binary menu) and signal s where their choices differ. Since Bayesian updating is a special case of Coarse Bayesian updating, this means (proper) Coarse

Bayesians can be distinguished from standard Bayesians via their choice behavior.

I now turn attention to the learning problem when signal realizations are not directly observed by the analyst. Some additional notation is needed for this exercise. Given an experiment σ and a Coarse Bayesian Representation $\langle \mathcal{P}, (\mu^P)_{P\in\mathcal{P}} \rangle$ of an updating rule μ, let $\sigma_{\mathcal{P}}$ denote the experiment formed by composing σ with $\mathcal{P}$; that is, for each $P \in \mathcal{P}$, let $s^P := \sum_{s\in\sigma:\mu^s=\mu^P} s$ and set $\sigma_{\mathcal{P}} := [s^P : P \in \mathcal{P}]$. Intuitively, $\sigma_{\mathcal{P}}$ is formed by merging signals in σ that result in the same Coarse Bayesian belief, which means merged signals have Bayesian posteriors belonging to a common cell P. This way, $\mathcal{P}$ effectively partitions signals in σ, and signals $s^1, \ldots, s^n \in \sigma$ belonging to the same cell are merged into a single column $s^1 + \ldots + s^n$. Clearly, this makes $\sigma_{\mathcal{P}}$ a Blackwell garbling of σ.

To study observational learning, I consider three different scenarios differing in what is known to the analyst in advance and what the analyst observes. In each case, I assume $\langle \mathcal{P}, (\mu^P)_{P\in\mathcal{P}} \rangle$ is known to the analyst.

Scenario 1: The analyst knows σ and observes posterior beliefs.
In this case, the analyst can learn from the Coarse Bayesian as follows. By observing the posterior $\hat{\mu}$ of the Coarse Bayesian, the analyst deduces via $\langle \mathcal{P}, (\mu^P)_{P\in\mathcal{P}} \rangle$ that $\hat{\mu} \in P$ and therefore that the Coarse Bayesian saw a signal s such that $\mu^s = \mu^P$; thus, the analyst effectively observes s^P as defined above. In general, observing the beliefs of a Coarse Bayesian is equivalent to observing those of a Bayesian with information $\sigma_{\mathcal{P}}$; Coarse Bayesians with finer partitions therefore provide Blackwell more-informative information to observers.

Scenario 2: The analyst knows σ and observes action choices only.
Relative to Scenario 1, this adds an additional layer of garbling from the analyst's perspective. Suppose the Coarse Bayesian's choices from A, but not the associated posterior beliefs $\hat{\mu}$, are available to the analyst. By choosing $x \in A$, the Coarse Bayesian reveals posterior beliefs $\hat{\mu}$ such that x is optimal at $\hat{\mu}$ (equivalently, the analyst learns that $s \in S(x|A)$), which amounts to a set of cells $P^1, \ldots, P^n \in \mathcal{P}$ spanning all cells $P^i \in \mathcal{P}$ where the representative μ^{P^i} makes x optimal in A. Consequently, the analyst effectively observes the signal $s^{P^1} + \ldots + s^{P^n}$; varying over all $x \in A$, then, clearly generates an overall information structure for the analyst that is a Blackwell garbling of the information $\sigma_{\mathcal{P}}$ generated in Scenario 1.

Note that expanding A need not improve the information generated by action choices. For example, expanding to $A \cup \{z\}$ eliminates all learning for the observer if z strictly dominates all actions in A. More generally, adding actions to A need not improve the information generated by a *Bayesian* decision maker, so it follows immediately that expanding A need not improve the information generated by Coarse Bayesian choices.[6]

Scenario 3: The analyst does not know σ.
The previous scenarios assumed that σ, but not its realization, was known to the analyst. If σ is not known, then observations of the Coarse Bayesian become incomplete (or ambiguous) signals. For example, observing the Coarse Bayesian's posterior $\hat{\mu}$ reveals the cell P containing the Bayesian posterior, but without knowledge of σ this only reveals that the observed signal s satisfies $B(\mu^e|s) \in \mathcal{P}$, yielding a convex cone of possible signals. Unless the analyst holds probabilistic beliefs about σ, this set of signals will not be reduced to a single signal and the analyst only learns that $B(\mu^e|s) \in \mathcal{P}$. Observing actions (but not posteriors) amplifies this problem.

A Proof of Proposition 11

Proof of part (iii). If every cell of $\langle \mathcal{P}, \mu^{\mathcal{P}} \rangle$ is a singleton, then the agent is Bayesian and the ML representation is established independently by the proof of part (iv) below. So, let $P^* \in \mathcal{P}$ be a non-singleton cell. Let I denote the set of all Coarse Bayesian Representations $i = \langle \mathcal{Q}(i), \dot{\mu}^{\mathcal{Q}(i)} \rangle$ such that $\mathcal{Q}(i)$ is finite, $\mathcal{P}$ is finer than $\mathcal{Q}(i)$, $\dot{\mu}^{\mathcal{Q}(i)} \subseteq \mu^{\mathcal{P}}$, and $P^* \in \mathcal{Q}(i)$. Define a partial order $\geq_I$ on I by $i \geq_I i'$ if and only if $\mathcal{Q}(i)$ is finer than $\mathcal{Q}(i')$ and $\dot{\mu}^{\mathcal{Q}(i)} \supseteq \dot{\mu}^{\mathcal{Q}(i')}$. It is straightforward to verify that $\geq_I$ is a partial order and that for all $i, i' \in I$, there exists $i^* \in I$ such that $i^* \geq_I i$ and $i^* \geq_I i'$. Thus, $(I, \geq_I)$ is a directed set.

For each $\langle \mathcal{Q}, \dot{\mu}^{\mathcal{Q}} \rangle \in I$, define a function $\gamma : \Delta \to [0, \infty)$ as follows. Since $N = 2$, the (finite) set $\dot{\mu}^{\mathcal{Q}}$ can be arranged in decreasing order of state 1: $\dot{\mu}^{\mathcal{Q}} = \{\dot{\mu}^{Q_1}, \ldots, \dot{\mu}^{Q_M}\}$, where $\dot{\mu}_1^{Q_1} > \dot{\mu}_1^{Q_2} > \ldots > \dot{\mu}_1^{Q_M}$. Since $P^* \in \mathcal{Q}$, there exists m^* such that $\dot{\mu}^{Q_{m^*}} = \mu^{P^*}$. For $1 \leq m < M$, let $\dot{\mu}^m$ denote the (unique) belief belonging to $\partial Q_m \cap \partial Q_{m+1}$ (the boundaries of Q_m and Q_{m+1}) and choose a signal s^m such that $B(\mu^e|s^m) = \dot{\mu}^m$. Now choose scalars $\alpha_m > 0$ such that, for all $1 \leq m < M$, $\alpha_m \dot{\mu}^{Q_m} \cdot s^m = \alpha_{m+1} \dot{\mu}^{Q_{m+1}} \cdot s^m$; taking $\alpha_{m^*} = 1$ pins down the α_m uniquely. Now define γ by

$$\gamma(\hat{\mu}) = \begin{cases} \alpha_m & \text{if } \hat{\mu} = \dot{\mu}^{Q_m} \\ 0 & \text{otherwise} \end{cases}.$$

[6]For example, in $A = \{x, y\}$, Bayesian choices partition Δ into three convex regions: beliefs where x is optimal, beliefs where y is optimal, and beliefs where x and y are tied. Adding a third option, z, can introduce a region where z is strictly optimal that intersects the original regions where x and y were optimal. Thus, the new partition contains a larger number of cells but it does not refine the original partition. This means information generated by the Bayesian need not improve at $A \cup \{z\}$, and it is not difficult to construct cases where information generated by the Coarse Bayesian does not improve, either.

By construction, $\dot{\mu}^{Q_m} \in \operatorname{argmax}_{\hat{\mu}} \gamma(\hat{\mu})\hat{\mu} \cdot s$ (that is, $\dot{\mu}^{Q_m}$ maximizes the likelihood function associated with γ) if and only if $B(\mu^e|s) \in Q_m$. Moreover, every point $\gamma(\hat{\mu})\hat{\mu}$, viewed as a point in $\mathbb{R}^2$, is contained in the half-space bounded above by the line with normal s^* passing through μ^{P^*}, where s^* is any signal such that $B(\mu^e|s^*) = \mu^{P^*}$. Thus, there exists a scalar $\overline{\gamma} > 0$ such that $\gamma(\hat{\mu}) \in [0, \overline{\gamma}]$ for all $\hat{\mu}$. Observe that the bound $\overline{\gamma}$ is independent of i.

Having defined a function $\gamma^i : \Delta \to [0, \overline{\gamma}]$ for every $i \in I$, the family $\{\gamma^i\}_{i \in I}$ forms a net. Each γ^i is an element of the (compact) product set $[0, \overline{\gamma}]^\Delta$, so that $\{\gamma^i\}_{i \in I}$ has a convergent subnet. This means there is a directed set $(J, \geq_J)$ and a function $\iota : J \to I$ such that (a) $j \geq_J j'$ implies $\iota(j) \geq_I \iota(j')$, (b) for every $i \in I$, there exists $j \in J$ such that $\iota(j') \geq_I i$ for all $j' \geq_J j$, and (c) the net $\{\gamma^{\iota(j)}\}_{j \in J}$ converges to some γ^*. Thus, for every $\hat{\mu} \in \Delta$, $\gamma^{\iota(j)}(\hat{\mu})$ converges to a point $\gamma^*(\hat{\mu})$.

Let $P \in \mathcal{P}$. By definition of $(I, \geq_I)$ and properties (a) and (b) of $(J, \geq_J)$, there exists $j^P \in J$ such that $P \in \mathcal{Q}(\iota(j))$ and $\mu^P \in \dot{\mu}^{\mathcal{P}(\iota(j))}$ for all $j \geq_J j^P$. Suppose s satisfies $B(\mu^e|s) \in P$. By construction, μ^P maximizes the likelihood function associated with $\gamma^{\iota(j)}$ at s if $j \geq_J j^P$: for every $\hat{\mu} \in \Delta$, $\gamma^{\iota(j)}(\mu^P)\mu^P \cdot s \geq \gamma^{\iota(j)}(\hat{\mu})\hat{\mu} \cdot s$. Taking the limit of both sides with respect to j yields $\gamma^*(\mu^P)\mu^P \cdot s \geq \gamma^*(\hat{\mu})\hat{\mu} \cdot s$; thus, μ^P maximizes the likelihood function associated with γ^* at s. □

Proof of part (iv). Notice that $B(\mu^e|s) = \mu'$ if and only if $s \approx \mu'/\mu^e := (\mu'_\omega/\mu^e_\omega)_{\omega \in \Omega}$. Thus, it will suffice to verify that $L(\cdot|s)$ is maximized at μ' for such signals s. This is done as follows. Let $s \in S$. Then, for any $\hat{\mu} \in \Delta$, we have

$$L(\hat{\mu}|s) = \gamma(\hat{\mu})\hat{\mu} \cdot s = \frac{\hat{\mu}}{\|\hat{\mu}/\sqrt{\mu^e}\|} \cdot s = \frac{\hat{\mu}/\sqrt{\mu^e}}{\|\hat{\mu}/\sqrt{\mu^e}\|} \cdot s\sqrt{\mu^e} = \left\| \frac{\hat{\mu}/\sqrt{\mu^e}}{\|\hat{\mu}/\sqrt{\mu^e}\|} \right\| \|s\sqrt{\mu^e}\| \cos\theta$$
$$= \|s\sqrt{\mu^e}\| \cos\theta$$

where θ is the angle (in radians) between $\hat{\mu}/\sqrt{\mu^e}$ and $s\sqrt{\mu^e}$. Thus, $L(\cdot|s)$ is maximized at $\hat{\mu}$ where $\hat{\mu}/\sqrt{\mu^e} \approx s\sqrt{\mu^e}$ (because then $\theta = 0$), implying $\hat{\mu} \approx s\mu^e \approx \frac{\mu'}{\mu^e}\mu^e = \mu'$. □

References

Jakobsen, A. M. (2025). Coarse bayesian updating. *Working paper*.

Mullainathan, S. (2002). Thinking through categories. *Working Paper*.

Ortoleva, P. (2012). Modeling the change of paradigm: Non-Bayesian reactions to unexpected news. *American Economic Review 102*(6), 2410–36.

Weinstein, J. (2017). Bayesian inference tempered by classical hypothesis testing. *Working paper*.