Theoretical Economics 17 (2022), 253–289 1555-7561/20220253

Pseudo-Bayesian updating

Chen Zhao
Faculty of Business and Economics, The University of Hong Kong

I propose an axiomatic framework for belief revision when new information is qualitative, of the form "event A is at least as likely as event B." My decision maker need not have beliefs about the joint distribution of the signal she will receive and the payoff-relevant states. I propose three axioms, *Exchangeability*, *Stationarity*, and *Reduction*, to characterize the class of *pseudo-Bayesian* updating rules. The key axiom, *Exchangeability*, requires that the order in which the information arrives does not matter if the different pieces of information neither reinforce nor contradict each other. I show that adding one more axiom, *Conservatism*, which requires that the decision maker adjust her beliefs just enough to embrace new information, yields Kullback–Leibler minimization: The decision maker selects the posterior closest to her prior in terms of Kullback–Leibler divergence from the probability measures consistent with newly received information. I show that pseudo-Bayesian agents are susceptible to recency bias, which may be mitigated by repetitive learning.

Keywords. Non-Bayesian updating, qualitative information, Kullback–Leibler divergence.

JEL classification. D01, D81, D83.

1. Introduction

Throughout the development of economic theory, the Bayesian framework has played a vital role in modeling how economic agents incorporate new information into their decision making. Central to the Bayesian framework is the assumption that the decision maker has prior beliefs about how the payoff-relevant states and the set of signals are jointly distributed: Such a joint prior enables the decision maker to interpret each signal realization and extract its information content.

In many situations, however, the assumption of a well-defined joint prior seems to impose too much informational and computational demand on the decision maker. Before receiving a signal, a Bayesian decision maker must be able to think about all possible signal realizations and form beliefs about their conditional distribution given each payoff-relevant state. This is certainly not possible when the information is unexpected.

Chen Zhao: czhao@hku.hk
I am deeply indebted to Faruk Gul and Wolfgang Pesendorfer for their invaluable advice, guidance, and encouragement. I am also grateful to two anonymous referees, Dilip Abreu, Roland Benabou, Yifan Dai, Shaowei Ke, Stephen Morris, Pietro Ortoleva, Marek Pycia, Jianrong Tian, and seminar participants at Harvard, Columbia, Ohio State University, Decision Theory at Penn, Kansas Workshop in Economic Theory, and SAET (Faro) for their insightful comments. Any errors are mine.

Available at https://econtheory.org. https://doi.org/10.3982/TE4535

Even when the information is expected, decision makers may not have enough prior information to form beliefs about the signal-generating process. For example, a patient who expects to read a research paper about her condition may not be able to think about any possible results or methodologies of the study; a Go player who expects to receive a recommendation from AlphaGo may not have any idea how the recommendation is generated.

In this paper, I propose a framework for belief revision that does not require a well-defined joint prior. My decision maker has a prior over the payoff-relevant states, but need not have beliefs about the signal-generating process; she need not know the full set of possible signal realizations or their conditional distribution. As a result, whether the information is expected or a surprise does not matter. In my model, the information content of a signal is not embedded in the joint prior, but is intrinsically given by the form of the realizations. In particular, each signal realization is a qualitative probability statement about the payoff-relevant states, of the form "event A_1 is at least as likely as event A_2."

I interpret the qualitative probability statement as the decision maker's understanding based on some new information; that is, although the decision maker does not have a joint prior to interpret the new information within the Bayesian framework, she is able to obtain some qualitative understanding from it. This type of qualitative understanding naturally comes up when the decision maker sees an informed party taking an action or a bet. In addition, due to its intuitiveness and pragmatism, the concept of qualitative probability has been the main building block in the theory of subjective probability.[1] Therefore, it seems natural to focus on qualitative probability statements in situations in which sophisticated probabilistic reasoning regarding the signal-generating process may not be realistic.

To motivate the model, consider the following example: Alice lives in a city in which there is a recent virus outbreak. She firmly believes that, as the local government claims, the current virus is less contagious than SARS. To her surprise, she observes that her neighbor, a respected doctor, is taking preventive measures that are similar to, if not stronger than, those in the SARS era. Alice is not an expert, but she fully understands that this observation is contrary to her previous position. Since the news is unexpected, Alice does not have a well-defined joint prior to interpret her neighbor's action. In that case, how will/should she revise her assessment of the current epidemic? How will/should she revise her beliefs about other related events?

Below, I present an axiomatic model of belief revision relevant to the kind of situation depicted in the example. My decision maker updates her beliefs by directly moving probability mass from the event "one gets SARS if exposed" to the event "one gets the new virus if exposed." In other words, she now believes that the new virus is more contagious than she used to think, and that SARS is less contagious than she used to think.

[1] See, for example, de Finetti (1937), Villegas (1964), and Fishburn (1986). This literature takes qualitative statements as primitives and imposes axioms to characterize when they are consistent with a probability measure. By contrast, I treat these statements as signals that prompt the decision maker to update her probabilistic beliefs.

Moreover, she will move probability mass in proportion to her prior beliefs. For example, she does not adjust her assessment of the case fatality risk—that is, the probability of death from infection—of both viruses.

Consider a decision maker who has a subjective prior μ over the set of payoff-relevant states. Suppose the decision maker learns some qualitative statement $\alpha = (A_1, A_2)$ that means "event A_1 is at least as likely as event A_2," which may or may not be consistent with the decision maker's prior ranking of A_1 and A_2. An *updating rule* associates, with every μ and each qualitative statement α, a posterior μ^α that is consistent with α. A decision maker equipped with an updating rule can process any finite string of qualitative statements sequentially: Each time the decision maker learns a new qualitative statement, she applies the updating rule to her current beliefs.

I impose three axioms on the updating rule to characterize the class of *pseudo-Bayesian updating rules*: First, *Exchangeability*, which requires that the order in which the information arrives does not matter if the different pieces of information neither reinforce nor contradict each other. Second, *Stationarity*, which requires that the decision maker not change her beliefs when she hears something she already believes, that is, $\mu^\alpha = \mu$ whenever $\mu(A_1) \geq \mu(A_2)$. Third, *Reduction*, which requires that the decision maker treat "A_1 is at least as likely as A_2" and "$A_1 \backslash A_2$ is at least as likely as $A_2 \backslash A_1$" as equivalent statements.

Equipped with a pseudo-Bayesian updating rule, the decision maker derives her posterior μ^α as follows: If $\mu(A_1) \geq \mu(A_2)$, since $\alpha = (A_1, A_2)$ is consistent with her prior μ, the decision maker will simply keep her prior, and thus set $\mu^\alpha = \mu$. If $0 < \mu(A_1) < \mu(A_2)$, then the decision maker will move probability mass from $A_2 \backslash A_1$ to $A_1 \backslash A_2$ so that A_1 becomes at least as likely as A_2. In the meantime, she does not touch the probability distribution over $A_1 \cap A_2$ and outside $A_1 \cup A_2$, and maintains the likelihood ratios between subsets of $A_1 \backslash A_2$ and between subsets of $A_2 \backslash A_1$. Finally, if $0 = \mu(A_1) < \mu(A_2)$, upon receiving α, the decision maker will remove the probability mass of A_2 altogether and redistribute the mass proportionately among the states in the complement, as if she were conditioning on the event "A_2 does not happen" according to Bayes' rule.

The last case above illustrates an important connection between pseudo-Bayesian and Bayesian updating: A pseudo-Bayesian decision maker is able to update on events in the state space according to Bayes' rule, since she has a prior over the payoff-relevant states. By reinterpreting the occurrence of event A as "$\emptyset$ is at least as likely as the complement of A," pseudo-Bayesian updating yields the Bayesian posterior, as if the decision maker were conditioning on A. In this sense, a pseudo-Bayesian decision maker will always apply Bayes' rule if it is applicable given her prior; that is, if she receives degenerate statements of the form $(\emptyset, \cdot)$.

On top of *Exchangeability*, *Stationarity*, and *Reduction*, I introduce *Conservatism*, which posits that upon receiving a statement that contradicts her prior, the decision maker moves probability mass conservatively, by just enough to equate the likelihood of the events concerned. Among the class of pseudo-Bayesian updating rules, *Conservatism* characterizes *the conservative rule*: If $0 < \mu(A_1) < \mu(A_2)$, then $A_1 \backslash A_2$ and $A_2 \backslash A_1$ share the probability mass $\mu(A_1 \cup A_2) - \mu(A_1 \cap A_2)$ equally in the posterior.

I find that the conservative rule is equivalent to Kullback–Leibler minimization. That is, the decision maker selects her posterior by minimizing the Kullback–Leibler divergence—a standard distance measure between probability distributions in statistics, subject to the constraint that the posterior should be consistent with the newly received statement. Thus, the decision maker is adjusting her beliefs minimally to accommodate new information.

In Section 2, I show that *Exchangeability*, *Stationarity*, and *Reduction* characterize the class of pseudo-Bayesian updating rules, and that adding *Conservatism* yields Kullback–Leibler minimization. In Section 3, I discuss the related literature. In Section 4, I show that pseudo-Bayesian agents are susceptible to recency bias, which may be mitigated by repetitive learning. Section 5 concludes.

2. Model

I present my main characterization results in this section. First, I describe the primitives of the model and introduce the orthogonality concept that plays a key role in my main axiom, *Exchangeability*. Then I show that *Exchangeability*, together with two other axioms, *Stationarity* and *Reduction*, characterize the class of pseudo-Bayesian updating procedures. Finally, I introduce *Conservatism* and show that it uniquely identifies the procedure of minimizing Kullback–Leibler divergence within the class.

Let Σ be a σ-algebra defined on state space S. I will use capital letters A, B, and C to denote generic elements of Σ. The decision maker (DM) has a *nonatomic* probabilistic prior, denoted as μ, on (S, Σ); that is, according to μ, any nonnull event can be further divided into nonnull subevents.[2]

The DM encounters a qualitative statement (A, B): "A is at least as likely as B." The statement is interpreted as the DM's understanding based on some new information. In the leading example, Alice understands from observing her doctor neighbor's action that "the probability of her getting the new virus is at least as large as that of her getting SARS." Note that this qualitative setting nests the occurrences of events as a special case: If $\emptyset$ is at least as likely as A's complement, then event A must have occurred, and vice versa.

I will use Greek letters α and β to denote generic statements. To ease exposition, I will let $\alpha = (A_1, A_2)$ and $\beta = (B_1, B_2)$ throughout the paper.

Given prior μ, a statement $\alpha = (A_1, A_2)$ is said to be *μ-noncredible* if $\mu(A_1) = 0$ and $\mu(A_2) = 1$; α is said to be *μ-credible* if it is not μ-noncredible. The DM, like her Bayesian counterpart, cannot assign a positive posterior probability to an event that has zero prior probability; there is no coherent way to distribute probability mass within the previously null event. Therefore, no posterior could embrace a noncredible α: Doing so would require increasing the probability of either A_1 or $S \backslash (A_1 \cup A_2)$, both of which are previously null. Hence, I assume that the DM ignores noncredible statements.

[2]The existence of a nonatomic probability measure requires that the state space S be uncountably infinite. If the state space is finite or countable, our axioms cannot identify a unique posterior.

Let $\Delta(S, \Sigma)$ be the set of nonatomic probability measures on (S, Σ). The DM's information set before updating is an element of

$$\mathbb{I}(S, \Sigma) := \{(\mu, \alpha) | \mu \in \Delta(S, \Sigma) \text{ and } \alpha \text{ is } \mu\text{-credible}\}.$$

The primitive of my model is an updating rule that associates a posterior with every $(\mu, \alpha) \in \mathbb{I}(S, \Sigma)$. I focus on the class of updating rules such that the DM successfully incorporates the new understanding α, which means "A_1 is at least as likely as A_2," into her beliefs.[3]

Definition. A function $\pi : \mathbb{I}(S, \Sigma) \to \Delta(S, \Sigma)$ is an updating rule if $\pi(\mu, \alpha)(A_1) \geq \pi(\mu, \alpha)(A_2)$ for any $(\mu, \alpha) \in \mathbb{I}(S, \Sigma)$.

Hereafter, given an updating rule π, I will use μ^α to denote the posterior $\pi(\mu, \alpha)$.

My model permits multiple stages of learning. A decision maker equipped with an updating rule can process any finite string of statements sequentially: Each time the DM learns a new qualitative statement, she applies it to her current beliefs according to the updating rule. Let $\mu^{\alpha_1\alpha_2\cdots\alpha_n}$ denote the DM's posterior, if it is well-defined, after processing statements $\alpha_1, \alpha_2, \dots, \alpha_n$ sequentially. Then $\mu^{\alpha_1\alpha_2\cdots\alpha_n} = \mu_n$, in which $\mu_0 = \mu$ and $\mu_k = \mu_{k-1}^{\alpha_k}$ for $k \geq 1$.

Note that by construction, the DM's information set does not contain a *joint* prior over the state space and the set of possible statements. As a result, Bayesian updating is not a well-defined updating rule over $\mathbb{I}(S, \Sigma)$. This setting reflects the assumption that the DM does not have enough prior information to incorporate the qualitative statements within the Bayesian framework. In fact, given a prior over the state space (S, Σ), Bayesian updating is only defined on degenerate statements that specify the occurrences of events. The following example provides an illustration.

Example (Dice). Let $S = \{1, 2, 3, 4, 5, 6\}$ and $\Sigma = 2^S$.[4] Suppose the DM initially believes that the dice is fair and does not expect any information that implies otherwise. Consider $\alpha = (\{1\}, \{3, 4, 5\})$ and $\beta = (\emptyset, \{3, 4, 5\})$. Since α does not correspond to any event in Σ, Bayesian updating is not defined for α. In contrast, by interpreting β as "$\{1, 2, 6\}$ must occur," Bayesian updating entails the posterior $(1/3, 1/3, 0, 0, 0, 1/3)$. ♢

Now, I introduce two updating rules that will be crucial in the analysis. I will use the dice example to illustrate how the updating rules behave and leave the formal definitions to Section 2.2.

The conservative rule. The DM adjusts her beliefs in proportion to her prior just enough to embrace the new understanding. In the dice example, upon receiving α, she

[3]The model permits reliability considerations regarding the information source if the state space is rich enough. For example, suppose $S = \{T, F\} \times \Omega$, where T means that the information source is reliable and F means otherwise. For $A_1, A_2 \subseteq \Omega$, statement $(\{T\} \times A_1, \{T\} \times A_2)$ can be interpreted as "if the information is reliable, A_1 is at least as likely as A_2." After the DM successfully incorporates this statement, the marginal probability of A_1 may still be lower than that of A_2.

[4]Although S is discrete, it can be viewed as a partition of a continuum on which nonatomic probability measures exist.

moves probability mass from $\{3, 4, 5\}$ to $\{1\}$ just enough to equate the likelihood of the two events, while keeping $\{3\}$, $\{4\}$, $\{5\}$ equiprobable, and does not touch $\{2, 6\}$. Under the conservative rule, $\mu^\alpha = (1/3, 1/6, 1/9, 1/9, 1/9, 1/6)$.

The radical rule. The DM radically adjusts her beliefs such that the newly incorporated understanding will not be violated upon receiving any statements thereafter. In the dice example, upon receiving α, she moves all probability mass from $\{3, 4, 5\}$ to $\{1\}$. Under the radical rule, $\mu^\alpha = (2/3, 1/6, 0, 0, 0, 1/6)$.

2.1 *Orthogonality*

My main axiom, *Exchangeability*, asserts that if two qualitative statements are *orthogonal*—that is, if they neither reinforce nor contradict each other—then the order in which the DM receives these qualitative statements does not affect her posterior. In this subsection, I provide a formal definition and discuss this orthogonality notion.

A statement α is said to be *standard* if $A_1 \cap A_2 = \emptyset$; *μ-degenerate* if $0 = \mu(A_1) < \mu(A_2)$. For any standard statement α, let

$$D_\alpha^\mu := \begin{cases} S, & \text{if } \alpha \text{ is } \mu\text{-degenerate,} \\ A_1 \cup A_2, & \text{otherwise} \end{cases}$$

be the *μ-domain* of α, which represents the set of states that are relevant to statement α. I assume that when α is degenerate, it is no different from $(\emptyset, A_2)$, which urges the DM to remove all of the probability mass from A_2 and redistribute it to the complement. Therefore, in this case, all states are forced to be relevant. For any α, let $\Pi_\alpha := \{A_1, A_2, S\backslash(A_1 \cup A_2)\}$. If α is standard, Π_α is a partition of S.

My orthogonality concept identifies pairs of *standard* qualitative statements that neither contradict nor reinforce each other. To understand what this means, consider Figure 1(a). Qualitative statement $\beta = (B_1, B_2)$ demands that the probability of B_2 be decreased; $\alpha = (A_1, A_2)$, by contrast, requires that the probability of A_1 (and thus B_2) be increased. In this sense, Figure 1(a) depicts a situation in which α and β are in conflict, and hence are not orthogonal. By contrast, in Figure 1(b), β compares two events outside the domain of α. In this case, α does not affect the relative likelihood of any state in B_1 versus any state in B_2, and thus α and β are orthogonal. Similarly, in Figure 1(b), α and $\gamma = (C_1, C_2)$ are orthogonal, since α has no bite on how C_1 and C_2 should compare, and thus should affect C_1 and C_2 equally.

The preceding observations motivate the following definition of orthogonality for two pieces of information.

Definition. Let α and β be standard and μ-credible. Then α and β are said to be μ-orthogonal if $D_\alpha^\mu \subseteq P \in \Pi_\beta$ or $D_\beta^\mu \subseteq P \in \Pi_\alpha$ for some P.

2.2 *Pseudo-Bayesian updating*

Now, I state my axioms and the main characterization theorems. I begin with my main axiom, *Exchangeability*.

Axiom 1 (Exchangeability). *For any* $(\mu, \alpha), (\mu, \beta) \in \mathbb{I}(S, \Sigma)$, *if* α *and* β *are* μ*-orthogonal, then* $\mu^{\alpha\beta} = \mu^{\beta\alpha}$.

Exchangeability asserts that if qualitative statements α and β are orthogonal, then the sequence in which the DM updates does not matter. This concept of exchangeability closely resembles the standard exchangeability notion in statistics, as examined by de Finetti: A sequence of random variables are exchangeable if the order of the sequence does not affect the joint probability distribution. In contrast to the standard notion, the one here only considers situations in which the qualitative statements are orthogonal; I allow the updating rule to be nonexchangeable otherwise.

Majumdar (2004) focuses on the standard Bayesian type of information—that is, the occurrence of some event—and axiomatizes Bayes' rule by assuming that exchangeability holds in general. Perea (2009) considers the same notion of exchangeability as Majumdar (2004), and proposes a model in which the decision maker's posterior minimizes the change of a linear objective function under the Euclidean metric. By contrast, my notion of *Exchangeability* applies to orthogonal qualitative statements.

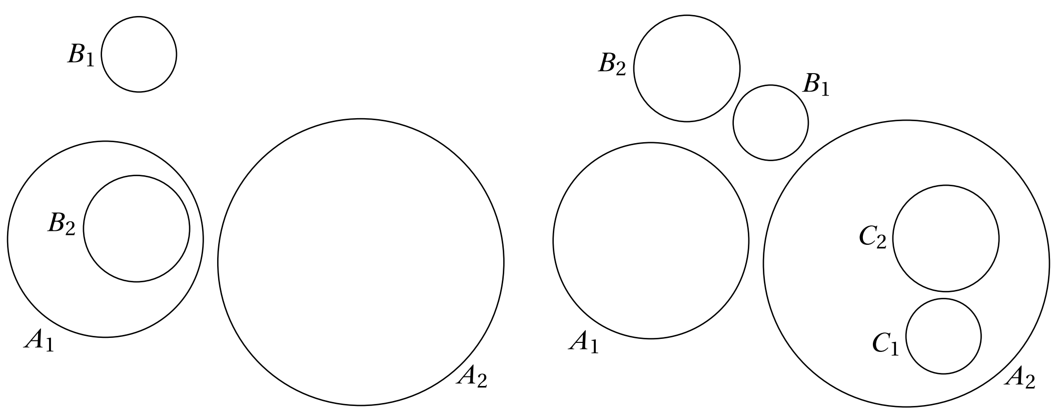
Next, I introduce *Stationarity* and *Reduction*. Recall that $\alpha = (A_1, A_2)$.

Axiom 2 (Stationarity). $\mu^{\alpha} = \mu$ *for all* $(\mu, \alpha) \in \mathbb{I}(S, \Sigma)$ *such that* $\mu(A_1) \geq \mu(A_2)$.

Stationarity posits that if the statement does not contradict the DM's prior, she will keep her beliefs unchanged. Recall that the DM has no prior belief over the qualitative statements she might receive. Thus, she interprets a statement that conforms to her prior simply as a confirmation of her prior beliefs and leaves them unchanged.

Axiom 3 (Reduction). $\mu^{\alpha} = \mu^{(A_1 \setminus A_2, A_2 \setminus A_1)}$ *for all* $(\mu, \alpha) \in \mathbb{I}(S, \Sigma)$.

Reduction posits that the DM treats "A_1 is at least as likely as A_2" and "$A_1 \setminus A_2$ is at least as likely as $A_2 \setminus A_1$" equivalently. The DM has well-defined probabilistic beliefs



(a) α and β are in conflict (b) α and β; α and γ are orthogonal

Figure 1. Orthogonality.

before and after updating, and thus understands that no matter which of the two statements she updates on, the other statement must also be consistent with her posterior. Hence, without any prior beliefs over the statements, the DM has no reason to treat them differently. Note that *Reduction* does not impose any restriction on how the DM updates on standard statements: α and $(A_1\backslash A_2, A_2\backslash A_1)$ are identical if $A_1 \cap A_2 = \emptyset$. Thus, without *Reduction*, all of the results in the paper still hold for standard statements.

Hereafter, an updating rule is said to be *pseudo-Bayesian* if it satisfies *Exchangeability*, *Stationarity*, and *Reduction*. Next, we characterize the properties of pseudo-Bayesian updating rules.

Given an updating rule and a *standard* statement $\alpha = (A_1, A_2)$, a probability measure $\nu \in \Delta(S, \Sigma)$ is said to be α-*connected* to $\mu \in \Delta(S, \Sigma)$ if there exists a statement β such that α and β are μ-orthogonal, and $\mu^\beta(C)/\mu^\beta(A_i) = \nu(C)/\nu(A_i)$ for any $C \subseteq A_i$ and $i = 1, 2$. In other words, ν is α-*connected* to μ if the conditional distributions on A_1 and on A_2 for μ can be transformed into the respective conditional distributions for ν by a statement that is orthogonal to α.

Example (Dice continued). Let $S = \{1, 2, 3, 4, 5, 6\}$ and $\Sigma = 2^S$. Consider the conservative rule as an example. Again let μ be uniform and $\alpha = (\{1\}, \{3, 4, 5\})$. Then any $\nu \in \Delta(S, \Sigma)$ of the form $(a, b, 2c, c, c, d)$ with $a, c > 0$ is α-connected to μ. To see that, let $\beta = (\{3\}, \{4, 5\})$. It is clear that α and β are μ-orthogonal. According to the conservative rule, $\mu^\beta = (1/6, 1/6, 1/4, 1/8, 1/8, 1/6)$. Thus, μ^β and ν have the same conditional distribution on both $\{1\}$ and $\{3, 4, 5\}$. ♢

To understand the role of α-connectedness in pseudo-Bayesian updating, consider standard statements α, β, and let ν be α-connected to μ via β. The fact that μ^β and ν have the same conditional distributions on A_1 and A_2 imposes a consistency restriction on $\mu^{\beta\alpha}$ and v^α. Then *Exchangeability*, through the requirement that $\mu^{\beta\alpha} = \mu^{\alpha\beta}$, channels the restriction to μ^α and ν^α. See (iic) in Theorem 1 for the exact statement of this restriction.

I now state the first theorem. Given any statement $\alpha = (A_1, A_2)$, it will be convenient to let $L_\alpha = A_1\backslash A_2$, $R_\alpha = A_2\backslash A_1$, and $C_\alpha = S\backslash(L_\alpha \cup R_\alpha)$, in which L, R, and C stand for "left-hand side," "right-hand side," and "complement," respectively. Clearly, $\{L_\alpha, R_\alpha, C_\alpha\}$ forms a partition of S, and α is standard if and only if $\alpha = (L_\alpha, R_\alpha)$. To ease exposition of Theorem 1, I adopt the convention that $0/0 = 0$.

Theorem 1. *An updating rule satisfies Exchangeability, Stationarity, and Reduction if and only if the following conditions are met for any* $(\mu, \alpha) \in \mathbb{I}(S, \Sigma)$*:*

(i) if $\mu(L_\alpha) \geq \mu(R_\alpha)$,

$$\mu^\alpha = \mu;$$

(ii) if $0 < \mu(L_\alpha) < \mu(R_\alpha)$,

(iia) for any $E \in \Sigma$,

$$\mu^\alpha(E) = \frac{\mu(E \cap L_\alpha)}{\mu(L_\alpha)}\mu^\alpha(L_\alpha) + \frac{\mu(E \cap R_\alpha)}{\mu(R_\alpha)}\mu^\alpha(R_\alpha) + \frac{\mu(E \cap C_\alpha)}{\mu(C_\alpha)}\mu^\alpha(C_\alpha);$$

(iib) $\mu^{\alpha}(C_{\alpha}) = \mu(C_{\alpha})$;

(iic) for $\nu \in \Delta(S, \Sigma)$ *that is* (L_{α}, R_{α})*-connected to* μ,

$$\frac{\mu(R_{\alpha})}{\mu(L_{\alpha})} = \frac{\nu(R_{\alpha})}{\nu(L_{\alpha})} \quad \Rightarrow \quad \frac{\mu^{\alpha}(R_{\alpha})}{\mu^{\alpha}(L_{\alpha})} = \frac{\nu^{\alpha}(R_{\alpha})}{\nu^{\alpha}(L_{\alpha})}.$$

(iii) if $0 = \mu(L_{\alpha}) < \mu(R_{\alpha})$*, for any* $E \in \Sigma$,

$$\mu^{\alpha}(E) = \frac{\mu(E \backslash R_{\alpha})}{\mu(S \backslash R_{\alpha})}.$$

Proof. For a proof, see Appendix A. □

Equipped with a pseudo-Bayesian updating rule, upon receiving statement α, the DM will first cancel the intersection from A_1 and A_2 and reduce α to (L_{α}, R_{α}). When $\mu(L_{\alpha}) \geq \mu(R_{\alpha})$, *Stationarity* directly requires that the DM not change her beliefs.

The salient case is when $0 < \mu(L_{\alpha}) < \mu(R_{\alpha})$. In this case, pseudo-Bayesian updating rules share the following properties: First, the probability distribution over each element of $\{L_{\alpha}, R_{\alpha}, C_{\alpha}\}$ is updated in proportion to the prior. Consider $E \subseteq L_{\alpha}$, for example. Condition (iia) requires that

$$\mu^{\alpha}(E) = \frac{\mu(E)}{\mu(L_{\alpha})} \cdot \mu^{\alpha}(L_{\alpha}),$$

which implies that the likelihood ratio between any subevents of L_{α} remains unchanged. Second, the probability mass assigned to C_{α} is unchanged, which combined with the first property, implies that the DM only moves probability mass from R_{α} to L_{α}, and does not touch the probability distribution over C_{α}. Third, within each class of priors that are (L_{α}, R_{α})-connected, the posterior likelihood ratio between R_{α} and L_{α} depends only on their prior likelihood ratio. Note that each (L_{α}, R_{α})-connected class contains a large variety of priors: ν being (L_{α}, R_{α})-connected to μ imposes no restriction on how ν is distributed over C_{α}.

The last case illustrates an important connection between pseudo-Bayesian and Bayesian updating. When $0 = \mu(L_{\alpha}) < \mu(R_{\alpha})$, the DM will remove all of the probability mass from R_{α} and redistribute it proportionately to its complement. Thus, the DM behaves as if she were conditioning on "R_{α} does not occur" according to Bayes' rule. The next corollary is an immediate consequence.

Corollary 1. *Given any pseudo-Bayesian updating rule, for any* $A \in \Sigma$ *and* $\mu \in \Delta(S, \Sigma)$ *with* $\mu(A) > 0$, $\mu^{(\emptyset, S \backslash A)}$ *is the Bayesian posterior of* μ *conditioning on* A.

Corollary 1 implies that a pseudo-Bayesian decision maker is able to update on events in the state space according to Bayes' rule. By reinterpreting the occurrence of event A as "Ø is at least as likely as the complement of A," pseudo-Bayesian updating yields the Bayesian posterior, as if the DM were conditioning on A. In this sense, a pseudo-Bayesian decision maker will always apply Bayes' rule if it is applicable; that is, if she receives degenerate statements of the form $(\emptyset, \cdot)$. Beyond degenerate statements, the DM will have to rely on pseudo-Bayesian updating rules, since she only has a prior over the payoff-relevant states.

Hence, pseudo-Bayesian updating affords the DM greater flexibility to process new information. Note that such extra flexibility does not imply that pseudo-Bayesian updating is more cognitively demanding than Bayesian updating. On the contrary, the essence of pseudo-Bayesian updating is using simple qualitative logic to replace probabilistic calculations when the DM is cognitively constrained: Pseudo-Bayesian updating rules are relevant precisely when the decision maker does not have a prior sophisticated enough to process new information according to Bayes' rule.

Next, I revisit the dice example to illustrate how pseudo-Bayesian updating works.

Example (Dice revisited). Let $S = \{1, 2, 3, 4, 5, 6\}$ and $\Sigma = 2^S$. Again let μ be uniform, $\alpha = (\{1\}, \{3, 4, 5\})$, and $\beta = (\emptyset, \{3, 4, 5\})$. Under any pseudo-Bayesian updating rule, upon receiving α, the DM will keep her beliefs over $\{2, 6\}$ unchanged. In addition, she will move probability mass from $\{3, 4, 5\}$ to $\{1\}$ to incorporate the statement, while keeping $\{3\}$, $\{4\}$, and $\{5\}$ equiprobable. If the DM receives β instead, her posterior becomes $(1/3, 1/3, 0, 0, 0, 1/3)$. ♢

The dice example demonstrates that for each statement, pseudo-Bayesian updating leaves at most one degree of freedom to the posterior—if $\mu^\alpha(\{1\}) = p$, then μ^α must be $(p, 1/6, 2/9 - p/3, 2/9 - p/3, 2/9 - p/3, 1/6)$. There are many possible ways to deplete the remaining degree of freedom; below are some examples.

The conservative rule. The DM adjusts her beliefs just enough to embrace the new understanding. That is, she sets $\mu^\alpha(L_\alpha) = \mu^\alpha(R_\alpha)$ for any $(\mu, \alpha) \in \mathbb{I}(S, \Sigma)$ such that $\mu(L_\alpha) < \mu(R_\alpha)$. In the dice example, the conservative rule implies that $\mu^\alpha = (1/3, 1/6, 1/9, 1/9, 1/9, 1/6)$.

The radical rule. The DM radically adjusts her beliefs such that the new understanding will never be violated in the future; that is, she sets $\mu^\alpha(R_\alpha) = 0$ for any $(\mu, \alpha) \in \mathbb{I}(S, \Sigma)$ such that $\mu(L_\alpha) < \mu(R_\alpha)$. As the DM's posterior has to be absolutely continuous with respect to her prior, R_α will remain null ever after, and thus α will always be consistent with the DM's beliefs despite the arrival of new information.[5] In the dice example, the radical rule requires that $\mu^\alpha = (2/3, 1/6, 0, 0, 0, 1/6)$.

The neutral rule.[6] The DM picks the middle ground between conservatism and radicalism; that is, she sets $\mu^\alpha(L_\alpha) = 2\mu^\alpha(R_\alpha)$ for any $(\mu, \alpha) \in \mathbb{I}(S, \Sigma)$ such that $\mu(L_\alpha) < \mu(R_\alpha)$. The neutral rule reflects the principle of insufficient reason—if any distribution with L_α being at least as likely as R_α is equally likely, then on average L_α is twice as likely as R_α. In the dice example, the neutral rule implies that $\mu^\alpha = (4/9, 1/6, 2/27, 2/27, 2/27, 1/6)$.

The common theme of these examples is that the posterior likelihood ratio between R_α and L_α is constant across all priors inconsistent with α, which need not be the case

[5] By conditions (i), (iia), and (iii) of Theorem 1, μ^α has to be absolutely continuous with respect to μ for any $(\mu, \alpha) \in \mathbb{I}(S, \Sigma)$.

[6] I thank Jacob Sagi for suggesting this example.

for general pseudo-Bayesian updating rules. Although these examples are highly non-generic, they serve as benchmarks in the analysis of the DM's belief dynamics. Moreover, the conservative rule is highly related to the minimization of Kullback–Leibler divergence.

2.3 *Conservatism and Kullback–Leibler minimization*

Formally, the following axiom characterizes the conservative rule.

Axiom 4 (Conservatism). $\mu^\alpha(A_1) = \mu^\alpha(A_2)$ *for all* $(\mu, \alpha) \in \mathbb{I}(S, \Sigma)$ *such that* $\mu(A_1) < \mu(A_2)$.

Conservatism posits that the DM adjusts her beliefs just enough to accommodate statements that are contradictory to her prior. It restricts the relative likelihood between A_1 and A_2, but is silent on the DM's beliefs about any other event.

The idea of having minimal changes in belief revision is not new. In the belief revision literature in philosophy, this idea is referred to as *the criterion of informational economy*, which asserts that when revising the set of propositions we believe, we want to avoid unnecessary loss of information by retaining our old propositions as much as possible.[7] Imagine that the DM's prior is summarized by (the deductive closure of) the set of all qualitative probability statements she believes. Believing that A_1 is strictly less likely than A_2 amounts to believing (A_2, A_1) but not (A_1, A_2). When the DM receives $\alpha = (A_1, A_2)$, the criterion of informational economy suggests that she keep (A_2, A_1), since (A_2, A_1) and (A_1, A_2) are not contradictory. If the DM believes in both (A_2, A_1) and (A_1, A_2), she will have to set A_1 and A_2 to be equally as likely. Further discussion of how my model is related to the belief revision literature in philosophy is provided in Section 3.

By Theorem 1, among the class of pseudo-Bayesian updating rules, *Conservatism* characterizes the following updating rule, which I call *the conservative rule*:

$$\mu^\alpha(E) = \begin{cases} \mu(E), & \text{if } \mu(L_\alpha) \geq \mu(R_\alpha) \\ \mu(E \cap C_\alpha) + \left(\dfrac{\mu(E \cap L_\alpha)}{\mu(L_\alpha)} \right. & \\ \qquad \left. + \dfrac{\mu(E \cap R_\alpha)}{\mu(R_\alpha)}\right) \dfrac{\mu(L_\alpha) + \mu(R_\alpha)}{2}, & \text{if } 0 < \mu(L_\alpha) < \mu(R_\alpha) \\ \dfrac{\mu(E \backslash R_\alpha)}{\mu(S \backslash R_\alpha)}, & \text{if } 0 = \mu(L_\alpha) < \mu(R_\alpha) \end{cases}$$

for any $E \in \Sigma$ and $(\mu, \alpha) \in \mathbb{I}(S, \Sigma)$. In addition to the properties given by Theorem 1, the conservative rule requires that if $0 < \mu(L_\alpha) < \mu(R_\alpha)$, then in the posterior, L_α and R_α each shares half of their original, combined probability mass.

Example (Uniform prior). Let $S = [0, 1]$. Suppose the DM, who has a uniform prior, encounters the statement $([0, 0.2], [0.6, 1])$. Given the conservative rule, the density of her

[7] See Gärdenfors (1984), Alchourrón et al. (1985) and Rott (2000).

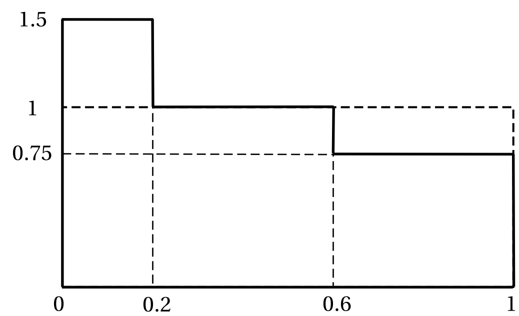


FIGURE 2. Uniform prior and $\alpha = ([0, 0.2], [0.6, 1])$.

posterior is shown in Figure 2. The density at states in $(0.2, 0.6)$ will remain unchanged and the density will be shifted proportionately from the interval $[0.6, 1]$ to the interval $[0, 0.2]$ just enough to equate the probabilities of $[0, 0.2]$ and $[0.6, 1]$. ♢

Perhaps surprisingly, the conservative rule is closely related to the minimization of Kullback–Leibler divergence (hereafter K-L divergence), which is widely used in statistics as a measure of distance between probability distributions. For $\mu, \nu \in \Delta(S, \Sigma)$, write $\nu \ll \mu$ if ν is absolutely continuous with respect to μ. For μ, ν such that $\nu \ll \mu$, the *Radon–Nikodym derivative* of ν with respect to μ, denoted as $d\nu/d\mu$, is defined as the measurable function $f : S \to [0, \infty)$ such that $\nu(A) = \int_A f\, d\mu$ for all $A \in \Sigma$.[8]

DEFINITION. For $\mu, \nu \in \Delta(S, \Sigma)$ such that $\nu \ll \mu$, the K-L divergence of ν from μ is given by

$$d(\mu \| \nu) = -\int_S \ln\left(\frac{d\nu}{d\mu}\right) d\mu.$$

It is well known that $d(\mu\|\nu)$ is strictly convex in ν, and that $d(\mu\|\nu) \geq 0$ with equality if and only if $\mu = \nu$ μ-almost surely. Now we are ready for the next theorem.

THEOREM 2. *The following statements are equivalent regarding a given updating rule:*

(i) the updating rule satisfies Exchangeability, Stationarity, Reduction, and Conservatism;

(ii) the updating rule is the conservative rule;

(iii) for any $(\mu, \alpha) \in \mathbb{I}(S, \Sigma)$,

$$\mu^\alpha = \arg\min_{\nu \ll \mu} d(\mu\|\nu) \tag{P}$$

$$\textit{subject to} \quad \nu(A_1) \geq \nu(A_2).$$

PROOF. For a proof, see Appendix B. □

[8]By the Radon–Nikodym theorem, such f always exists and is unique up to a zero μ-measure set.

Thus, equipped with the conservative rule, the DM selects the posterior closest to her prior in terms of K-L divergence, subject to the constraint that the posterior is consistent with the newly received statement. In information theory, constrained optimization (P) is sometimes called the *method of maximum relative entropy* (hereafter MrE), since $-d(\mu\|\nu)$ is the relative entropy of ν with respect to μ.

K-L minimization plays an important role in the literature in statistics on estimation with a misspecified prior. A strand of this literature aims to characterize the asymptotic behavior of Bayes estimators under the assumption that the model may be misspecified; that is, the support of the prior used may not contain the true parameter value. Berk (1966, 1970), Bunke and Milhaud (1998) consider a Bayesian agent with a misspecified (second-order) prior receiving i.i.d. data and shows that under certain regularity conditions, her (second-order) posterior will asymptotically concentrate on the distributions in the support of her (second-order) prior that minimizes K-L divergence from the true distribution. To illustrate the connection of Theorem 2 with this literature, suppose the DM has prior μ and let Δ_α be the set of all probability distributions consistent with $\alpha = (A_1, A_2)$ and absolutely continuous with respect to μ. Thus, equipped with the conservative rule, the DM has $\mu^\alpha = \arg\min_{\nu\in\Delta_\alpha} d(\mu\|\nu)$. If the Bayesian agent in Berk (1966) has a (second-order) prior with support Δ_α and the true model is μ, then her posterior converges to μ^α as she receives more data. In other words, a conservative pseudo-Bayesian decision maker with information set (μ, α) will have the asymptotic beliefs of a Bayesian who thinks the true distribution is in Δ_α but it in fact is μ. In a certain sense, our DM behaves as if she thinks her prior μ is true but is forced to reason within Δ_α, which matches the general perception of a conservative person.

3. Related literature

The earliest version of K-L minimization is proposed by Kullback (1959) as the *principle of minimum discrimination information*: Given new facts, the posterior distribution should be the one that is hardest to discriminate from the prior, in the sense of Kullback–Leibler divergence.[9] This principle is further developed in the literature of information theory on maximum entropy methods. This literature aims to find universally applicable algorithms for probablistic inference when new information imposes constraints on the probability distribution. Papers in this literature posit a well-parameterized class of constrained optimization models. In particular, the statistician is *assumed* to be able to use standard Lagrangian arguments to optimize his posterior subject to the constraints. In contrast, I consider a choice-theoretic state space and start from a general mapping that assigns a posterior to each piece of information given a prior. Furthermore, the class of updating rules I characterize is beyond the scope of constrained optimization models. In fact, a pseudo-Bayesian updating rule that violates *Conservatism* cannot be rationalized by any constrained optimization model. Within this literature, Caticha (2004)

[9]When first introduced by Kullback and Leibler (1951), the K-L divergence between probability measures μ and ν is symmetric, defined as $d(\mu\|\nu) + d(\nu\|\mu)$. It is designed to measure how difficult it is for a statistician to discriminate between distributions with the best test. In Kullback (1959), the directed version, $d(\nu\|\mu)$, is used, where μ represents the prior and ν represents the posterior.

axiomatizes the minimization of $d(\mu\|\cdot)$ with conditions that require, for example, invariance to coordinates changes. In a similar vein, Shore and Johnson (1980), Skilling (1988), Caticha (2004), and Caticha and Giffin (2006) propose to minimize $d(\cdot\|\mu)$ instead of $d(\mu\|\cdot)$. When μ is uniform, minimizing $d(\cdot\|\mu)$ reduces to the original maximum entropy method proposed by Jaynes (1957), which appears in Spiegler (2017, 2020) in economic contexts.

A strand of papers in decision theory considers a decision maker who knows that the objective probability lies in some set and models how she forms her subjective prior by selecting from the set; see, for example, Chambers and Hayashi (2010) and Damiano (2006). The structure of information in these papers nests my qualitative setting; however, the belief selection rules they consider takes the set of possible objective probabilities as the only argument. In other words, the decision maker in these studies does not have a prior to begin with before knowing the probability-possibility set. Probability-possibility sets also appear in Ahn (2008) and Gajdos et al. (2008), among others, as a mechanism to generate ambiguity.

Models of non-Bayesian updating in economics typically consider a setting in which Bayes' rule is applicable but the DM deviates from it due to a bias, bounded rationality, or temptation. By contrast, I assume that the agent uses Bayes' rule when it is applicable, but extends updating to situations in which Bayes' rule does not apply. In particular, given a prior over the state space (S, Σ), Bayes' rule only applies to information of the form "$A \in \Sigma$ has occurred." While nesting such information as $(\emptyset, S \backslash A)$, my qualitative setting also permits statements that are not events in the state space. For behavioral models of non-Bayesian updating, see, for example, Rabin and Schrag (1999), Rabin (2002), Mullainathan et al. (2008), and Gennaioli and Shleifer (2010). In the decision theory literature, Zhao (2020) formally links the concept of similarity with belief updating to explain a wide class of non-Bayesian fallacies. Ortoleva (2012) proposes a hypothesis-testing model in which agents reject their prior when a rare event occurs. Epstein (2006) and Epstein et al. (2008) build on Gul and Pesendorfer's temptation theory and show that the DM might be tempted to use a posterior that differs from the Bayesian update.

In philosophy, belief revision refers to the process of revising a theory (a set of deductively closed propositions) to accommodate an inconsistent proposition. In this literature, the AGM model (Alchourrón et al. (1985)) is the dominant theory.[10] Bonanno (2009) and Basu (2019) apply the AGM model to probabilistic belief updating by treating events as propositions. In their models, the AGM postulates are used to regulate the support of the decision maker's posterior given that some event occurs. In contrast, I consider qualitative statements, which prompt a more general propositional language than the algebra of events. Furthermore, compared with the AGM setting, my decision maker's prior beliefs have more structure than a deductively closed set of propositions—that is, the set of qualitative statements the DM believes to be true before revision uniquely define a probability measure.[11]

[10] See Costa and Pedersen (2011) and Fermé and Hansson (2011) for surveys of the literature.

[11] This is because the DM's prior is nonatomic. See Villegas (1964), Section 4, Theorem 3, for a proof.

4. Belief dynamics of pseudo-Bayesian agents

In this section, I explore the belief dynamics of pseudo-Bayesian agents. I show that pseudo-Bayesian agents are in general susceptible to persistent recency bias, but repetitive learning may mitigate the bias.

4.1 *Recency bias and repetitive learning*

The DM employs a sequential updating procedure and views each piece of new information as a constraint. Suppose the DM learns α then β. Recall that *Exchangeability* requires that if α and β are orthogonal, the order in which they arrive does not matter. It follows that the DM's posterior, $\mu^{\alpha\beta}$, must be consistent with both statements. However, if α and β are in conflict, $\mu^{\alpha\beta}$ must be consistent with β but need not be consistent with α. When the arrival of a second statement prompts, the DM to "forget" the first statement she has already learned, *recency bias* has been induced. Formally, an ordered pair of statements (α, β) is said to induce *recency bias* on μ if $\mu(A_1) < \mu(A_2)$ and $\mu^{\alpha\beta}(A_1) < \mu^{\alpha\beta}(A_2)$.

Example (Recency bias). Let $S = \{1, 2, 3\}$. Suppose the DM with prior beliefs $\mu = (0.2, 0.3, 0.5)$ receives $\alpha = (\{2\}, \{3\})$ then $\beta = (\{1\}, \{2\})$. The neutral rule, for example, requires that $\mu^{\alpha} = (0.2, 0.533, 0.267)$ and $\mu^{\alpha\beta} = (0.489, 0.244, 0.267)$. Notice that now state 2 has a lower probability than state 3—statement α becomes contradictory to the DM's beliefs again after she learns β. ◊

Moreover, (α, β) is said to induce *persistent recency bias on* μ if $\mu^{(\alpha\beta)^k\alpha}(B_1) < \mu^{(\alpha\beta)^k\alpha}(B_2)$ for $k = 0, 1, 2, \dots,$ in which $\mu^{(\alpha\beta)^k\alpha}$ denotes the DM's posterior after she learns α, β repeatedly for k times and then one more α. Thus, when persistent recency bias is present, if the DM learns α and β alternately, whenever she processes α, statement β is no longer consistent with her beliefs, and vice versa.[12]

Example (Persistent recency bias). Let $S = \{1, 2, 3\}$. Suppose the DM with prior beliefs $\mu = (0.2, 0.3, 0.5)$ receives $\alpha = (\{2\}, \{3\})$ then $\beta = (\{1\}, \{2\})$. The conservative rule, for example, requires $\mu^{\alpha} = (0.2, 0.4, 0.4)$ and $\mu^{\alpha\beta} = (0.3, 0.3, 0.4)$. Notice that now α is violated. Moreover, $\mu^{\alpha\beta\alpha} = (0.3, 0.35, 0.35)$, and thus β is no longer consistent with the DM's beliefs. As the DM learns α and β alternately, her beliefs will keep oscillating. ◊

An updating rule is said to be *susceptible to* (*persistent*) *recency bias* if there exists $\mu \in \Delta(S, \Sigma)$ and statements α, β such that (α, β) induces (persistent) recency bias on μ. Clearly, the radical rule is not susceptible to any recency bias because the DM's posterior has to be absolutely continuous with respect to her prior. A pseudo-Bayesian updating rule is said to be *never-radical* if $\mu^{\alpha}(R_\alpha) > 0$ for any $(\mu, \alpha) \in \mathbb{I}(S, \Sigma)$ such that $0 < \mu(L_\alpha) < \mu(R_\alpha)$. Thus, a pseudo-Bayesian updating rule is never-radical if the DM creates new null events only upon encountering statements with a degenerate reduction.

[12] By *Stationarity*, once the DM's beliefs are consistent with α after processing β, her beliefs will remain consistent with both statements ever after, and thus, persistent recency bias cannot have been induced.

Proposition 1. *A pseudo-Bayesian updating rule is susceptible to recency bias if and only if it is not the radical rule. Moreover, any never-radical pseudo-Bayesian updating rule is susceptible to persistent recency bias.*

Proof. For a proof, see the Appendix C. □

In fact, given a never-radical pseudo-Bayesian updating rule, for any prior μ, there always exists a pair of statements such that the DM cannot accommodate both at the same time with finite stages of alternate learning. In the limit, however, it is possible that the DM's beliefs converge to a probability distribution that is consistent with both statements. In particular, if the DM adopts the conservative rule, this kind of repetitive learning is *always* effective in the limit. Consider the following example.

Example (Limiting beliefs). Let $S = \{1, 2, 3\}$. Suppose the DM with prior beliefs $\mu = (0.2, 0.3, 0.5)$ receives $\alpha = (\{2\}, \{3\})$ then $\beta = (\{1\}, \{2\})$. The conservative rule implies that $\lim_{k\to\infty} \mu^{(\alpha\beta)^k} = \lim_{k\to\infty} \mu^{(\alpha\beta)^k\alpha} = (1/3, 1/3, 1/3)$. Thus, both statements are accommodated in the limit. ◊

It follows that repetition plays an important role in learning for pseudo-Bayesian agents. Without the joint prior over states and signals, the only information the DM can store after learning a statement is her posterior beliefs over the states. By contrast, once a Bayesian decision maker forms a joint prior, she has already thought about not only the possibilities of the new signal and how they interact with the payoff-relevant states, but also all the past signals she has received or might have received. Thus, hearing an old piece of information again does nothing to a Bayesian decison maker's beliefs.

DeMarzo et al. (2003) consider a setting in which agents treat any information they receive as new and independent information. Since the agents do not adjust properly for repetition, repeated exposure to an opinion has a cumulative effect on their beliefs. In my model, however, repetition plays a role only in the presence of contradictory information and, in particular, when recency bias occurs. If the DM learns a single statement repeatedly, she will simply stop revising her beliefs after the first time—my DM does not have the joint beliefs needed to treat these repetitive occurrences differently.

More generally, let $\alpha_i = (A_{i1}, A_{i2})$. A finite set of statements $\{\alpha_i\}_{i=1}^n$ is said to be μ-*compatible* if there exists $\nu \ll \mu$ such that $\nu(A_{i1}) \geq \nu(A_{i2})$ for all i. Such ν is called a μ-*solution* to $\{\alpha_i\}_{i=1}^n$. Put differently, $\{\alpha_i\}_{i=1}^n$ is μ-compatible if all of the statements are sampled from a probability measure ν that is absolutely continuous with respect to μ.

Given n statements, a *learning sequence* is a sequence of elements in $\{1, 2, \ldots, n\}$ that represents the order of the indices in which the DM learns the statements. A learning sequence $\{i_k\}_{k=1}^\infty$ is said to be *comprehensive* if there is $N \in \mathbb{N}$ such that $\{1, 2, \ldots, n\} \subseteq \bigcup_{j=1}^N i_{mN+j}$ for $m \in \mathbb{N}$. In other words, within each block of N steps, the DM learns each statement at least once.

Theorem 3 (Qualitative law of large numbers). *Suppose the DM adopts the conservative rule and let $\{\alpha_i\}_{i=1}^n$ be μ-compatible. If $\{i_k\}_{k=1}^\infty$ is a comprehensive learning sequence, then $\mu^{\alpha_{i_1}\alpha_{i_2}\cdots\alpha_{i_k}}$ converges in total variation to a μ-solution to $\{\alpha_i\}_{i=1}^n$ as $k \to \infty$.*

Proof. For a proof, see Appendix D. □

Thus, the conservative rule allows the DM to digest any finite collection of objectively true qualitative statements with repetitive learning, provided she has not neglected the possibility of any truly probable event. By learning a statement with the conservative rule, the DM essentially projects her beliefs onto the closed and convex set of probabilities that are consistent with the statement. Bregman (1967) proves that if the notion of distance is well behaved, cyclically projecting onto a finite collection of closed and convex sets converges to a point in the intersection. Although optimization (P) in Theorem 2 is not exactly Bregman's type, the proof is similar in spirit to Bregman's.

It is worth emphasizing that the radical rule and the neutral rule do not have this convergence property: For example, let $A \in \Sigma$ be such that $0 < \mu(A) < 1/2$. Consider statements $\alpha = (A, S\backslash A)$ and $\beta = (S\backslash A, A)$. The radical rule implies that after learning α, statement β becomes noncredible, and vice versa. It follows that the DM can never accommodate both statements at the same time. The neutral rule implies that $\mu^{(\alpha\beta)^k}(A) = 1/3$ and $\mu^{(\alpha\beta)^k\alpha}(A) = 2/3$ for any $k \geq 1$, which means that the DM's beliefs do not converge and are never consistent with both statements.

5. Final remarks

In this paper, I consider a situation in which the decision maker receives qualitative information. Three simple axioms, *Exchangeability*, *Stationarity*, and *Reduction*, deliver the class of pseudo-Bayesian updating rules. With the addition of *Conservatism*, the pseudo-Bayesian posterior turns out to be the closest probability measure to the decision maker's prior that is consistent with the newly received information. I show that the DM is susceptible to recency bias and that repetition may enable her to overcome it.

The primitive of my model is an updating rule that associates a posterior with every credible prior-statement pair. Thus, to verify the axioms (in particular, *Exchangeability*), the analyst must be able to observe the DM's posterior for every such prior-statement pair, which is admittedly difficult without involving a population of agents. However, if the analyst can observe the belief path of a single agent as she receives information sequentially, all four axioms in my model are falsifiable. In particular, if the agent exhibits recency bias upon receiving orthogonal statements, then *Exchangeability* must be violated.

Several extensions of the model can be pursued. First, with a slight modification of *Exchangeability*, noncredible statements can be incorporated. In the model, I assume that the decision maker ignores noncredible statements, since the interaction between *Exchangeability* and *Stationarity* requires that the DM's posterior be absolutely continuous with respect to her prior. However, this absolute continuity assumption can be relaxed: If we allow α and β to be nonexchangeable when A_1 is null and $B_1, B_2 \subseteq A_1$, then A_1 does not have to remain null when the decision maker updates on α. In that case, all we need is another axiom to regulate how the posterior should be distributed over previously null events.

Second, pseudo-Bayesian updating can easily be extended to *quantitative* statements of the form "A is x times as likely as B" where $A \cap B = \emptyset$. If we extend our definitions of updating rules and orthogonality in the obvious way, *Exchangeability* on its own yields the following updating rule: The decision maker sets A to be x times as likely as B by moving probability mass between A and B in proportion to her prior, and again does not touch the probability distribution outside $A \cup B$. It is easy to see that here the DM again selects her posterior using Kullback–Leibler minimization.

Appendix A: Proof of Theorem 1

Proof. Since μ is nonatomic and countably additive, it is also *convex-ranged*; that is, for any $a \in (0, 1)$ and A such that $\mu(A) > 0$, there is $B \subseteq A$ such that $\mu(B) = a\mu(A)$. This property is exploited from time to time throughout the proof. Also, for any $\mu \in \Delta(S, \Sigma)$ and any $A \in \Sigma$ such that $\mu(A) > 0$, it will be convenient to define $\mu_A \in \Delta(S, \Sigma)$ to be such that

$$\mu_A(E) = \frac{\mu(E \cap A)}{\mu(A)}$$

for any $E \in \Sigma$; that is, μ_A is the Bayesian posterior of μ conditioning on A. To ease exposition, I will maintain the convention that $0/0 = 0$ throughout the proof.

I will first prove the "only if" part. By *Reduction*, it suffices to focus on standard statements. Recall that when α is standard, $L_\alpha = A_1$, $R_\alpha = A_2$, and $C_\alpha = S \backslash (A_1 \cup A_2)$. By *Stationarity*, condition (i) of Theorem 1 is trivial. The proofs of (ii) and (iii) are broken into a series of lemmas. Recall that $\alpha = (A_1, A_2)$ and $\beta = (B_1, B_2)$.

Lemma 1. *If an updating rule satisfies Stationarity and Exchangeability, then*

$$\mu(B_1) \geq \mu(B_2) \quad \Rightarrow \quad \mu^\alpha(B_1) \geq \mu^\alpha(B_2)$$

for standard, μ-credible statements α, β that are μ-orthogonal.

Proof. By way of contradiction, suppose $\mu(B_1) \geq \mu(B_2)$ but $\mu^\alpha(B_1) < \mu^\alpha(B_2)$. *Stationarity* and $\mu(B_1) \geq \mu(B_2)$ together imply $\mu^\beta = \mu$, and thus $\mu^{\beta\alpha} = \mu^\alpha$. Hence, $\mu^{\beta\alpha}(B_1) < \mu^{\beta\alpha}(B_2)$. However, by definition of an updating rule, since $\mu^{\alpha\beta} = (\mu^\alpha)^\beta$, we have $\mu^{\alpha\beta}(B_1) \geq \mu^{\alpha\beta}(B_2)$. It follows that we cannot have $\mu^{\alpha\beta} = \mu^{\beta\alpha}$, a contradiction to *Exchangeability*. □

Lemma 2. *If an updating rule satisfies Stationarity and Exchangeability, then for any $(\mu, \alpha) \in \mathbb{I}(S, \Sigma)$ such that α is standard,*

$$\mu^\alpha(E) = \frac{\mu(E \cap A_1)}{\mu(A_1)}\mu^\alpha(A_1) + \frac{\mu(E \cap A_2)}{\mu(A_2)}\mu^\alpha(A_2) + \frac{\mu\big(E \backslash (A_1 \cup A_2)\big)}{\mu\big(S \backslash (A_1 \cup A_2)\big)}\mu^\alpha\big(S \backslash (A_1 \cup A_2)\big)$$

for any $E \in \Sigma$.

Proof. Recall that for any standard statement α, $\Pi_\alpha = \{A_1, A_2, S\backslash(A_1 \cup A_2)\}$. Pick any partition element $P \in \Pi_\alpha$. Note that for any $B_1, B_2 \subseteq P$ such that $B_1 \cap B_2 = \emptyset$ and $\mu(B_1) \geq \mu(B_2)$, we always have $\mu^\alpha(B_1) \geq \mu^\alpha(B_2)$. This is because, since $\mu(B_1) \geq \mu(B_2)$, β is μ-credible and not μ-degenerate. Therefore, $D^\mu_\beta = B_1 \cup B_2 \subseteq P \in \Pi_\alpha$, which implies α and β are μ-orthogonal. Then, by Lemma 1, $\mu^\alpha(B_1) \geq \mu^\alpha(B_2)$. It then follows that for any $B_1, B_2 \subseteq P$ such that $B_1 \cap B_2 = \emptyset$ and $\mu(B_1) = \mu(B_2)$, we always have $\mu^\alpha(B_1) = \mu^\alpha(B_2)$.

Now we are ready to prove the lemma. By $\mu^\alpha(E) = \sum_{P \in \Pi_\alpha} \mu^\alpha(E \cap P)$ for any $E \in \Sigma$, it suffices to show

$$\mu^\alpha(E \cap P) = \frac{\mu(E \cap P)}{\mu(P)} \mu^\alpha(P)$$

for any $E \in \Sigma$ and $P \in \Pi_\alpha$. There are three cases:

Case 1: $\mu(P) = 0$. As $\mu(\emptyset) = \mu(P) = 0$, we have $\mu^\alpha(\emptyset) = \mu^\alpha(P)$, which means $\mu^\alpha(P) = 0$. In this case, for any $E \in \Sigma$, we have

$$\mu^\alpha(E \cap P) = 0 = \frac{\mu(E \cap P)}{\mu(P)} \mu^\alpha(P),$$

under the convention that $0/0 = 0$.

Case 2: $\mu(P) > 0$ and $\mu^\alpha(P) = 0$. In this case, we always have

$$\mu^\alpha(E \cap P) = 0 = \frac{\mu(E \cap P)}{\mu(P)} \mu^\alpha(P),$$

for any $E \in \Sigma$, since $\mu^\alpha(E \cap P) = \mu^\alpha(P) = 0$.

Case 3: $\mu(P) > 0$ and $\mu^\alpha(P) > 0$. For any $B_1, B_2 \subseteq P$ such that $B_1 \cap B_2 = \emptyset$ and $\mu(B_1) = \mu(B_2)$, we have $\mu^\alpha(B_1) = \mu^\alpha(B_2)$. Since μ is nonatomic (and thus convex-ranged), for any $n \in \mathbb{N}$, there exists a partition of P, denoted as $\{C^n_i\}^n_{i=1}$, such that $\mu(C^n_i) = \mu(P)/n$ for any i. Call such $\{C^n_i\}^n_{i=1}$ an n-fold μ-equipartition of P. By $\mu(C^n_i) = \mu(C^n_j)$ for any i, j, it follows that $\mu^\alpha(C^n_i) = \mu^\alpha(C^n_j)$ for any i, j, which implies $\mu^\alpha(C^n_i) = \mu^\alpha(P)/n$. Therefore, any n-fold μ-equipartition of P is also an n-fold μ^α-equipartition of P. Thus, for any n-fold μ-equipartition $\{C^n_i\}^n_{i=1}$ and any i,

$$\frac{\mu(C^n_i)}{\mu(P)} = \frac{\mu^\alpha(C^n_i)}{\mu^\alpha(P)}.$$

It then follows that for any $C \subseteq P$ such that $\mu(C)/\mu(P) = m/n$ for some $m, n \in \mathbb{N}$ with $m \leq n$ and $n > 0$, we always have

$$\frac{\mu(C)}{\mu(P)} = \frac{\mu^\alpha(C)}{\mu^\alpha(P)} = \frac{m}{n}.$$

To see why, note that by the convex-rangedness of μ, any such C can be expressed as the union of m distinct elements of some n-fold μ-equipartition of P. As any n-fold μ-equipartition of P is also an n-fold μ^α-equipartition of P, the additivity of μ^α yields

$\mu^\alpha(C) = (m/n)\cdot\mu^\alpha(P)$. Since any nonnegative real number can be expressed as a countable sum of nonnegative rational numbers, the convex-rangedness of μ and the countable additivity of μ and μ^α together yield

$$\frac{\mu(C)}{\mu(P)} = \frac{\mu^\alpha(C)}{\mu^\alpha(P)}$$

for any $C \subseteq P$. Hence, if $\mu(P) > 0$ and $\mu^\alpha(P) > 0$, then

$$\mu^\alpha(E\cap P) = \frac{\mu(E\cap P)}{\mu(P)}\mu^\alpha(P).$$

for any $E \in \Sigma$. □

Note that Lemma 2 completes the proof of condition (iii) in Theorem 1. To see that, let α be a standard statement such that $0 = \mu(A_1) < \mu(A_2)$. By Lemma 2, it is clear that $\mu^\alpha(A_1) = 0$ (note the convention that $0/0 = 0$). The definition of updating rules requires that $\mu^\alpha(A_1) \geq \mu^\alpha(A_2)$, which implies $\mu^\alpha(A_2) = 0$. Thus, by Lemma 2

$$\mu^\alpha(E) = \frac{\mu\big(E\backslash(A_1\cup A_2)\big)}{\mu\big(S\backslash(A_1\cup A_2)\big)}\mu^\alpha\big(S\backslash(A_1\cup A_2)\big)$$

for any $E \in \Sigma$. Then $\mu(A_1) = \mu^\alpha(A_1) = \mu^\alpha(A_2) = 0$ implies that

$$\mu^\alpha(E) = \frac{\mu(E\backslash A_2)}{\mu(S\backslash A_2)}\mu^\alpha(S) = \frac{\mu(E\backslash A_2)}{\mu(S\backslash A_2)}$$

for any $E \in \Sigma$. Thus, the DM behaves as if she is conditioning on $S\backslash A_2$ according to Bayes' rule. Condition (iii) in Theorem 1 is established.

I now turn to condition (ii). Note that (iia) is directly implied by Lemma 2. Thus, it suffices to show (iib) and (iic).

To show condition (iib), it suffices to show that for any $(\mu, \alpha) \in \mathbb{I}(S, \Sigma)$ such that α is standard and $0 < \mu(A_1) < \mu(A_2)$, we have $\mu(A_1\cup A_2) = \mu^\alpha(A_1\cup A_2)$. If $\mu(A_1\cup A_2) = 1$, then $\mu(S\backslash(A_1\cup A_2)) = 0$. Then by Lemma 2, $\mu^\alpha(S\backslash(A_1\cup A_2)) = 0$, which implies $\mu^\alpha(A_1\cup A_2) = 1 = \mu(A_1\cup A_2)$.

Hereafter, I will maintain the assumption that α is a standard statement with $0 < \mu(A_1) < \mu(A_2) < 1-\mu(A_1)$. Consider statement $\beta = (\emptyset, B_2)$, in which $B_2 \subseteq S\backslash(A_1\cup A_2)$ and $0 < \mu(B_2) < 1-\mu(A_1\cup A_2)$. Such B_2 exists, since μ is nonatomic. Clearly, β is μ-credible and standard. Furthermore, by $D^\mu_\alpha = A_1\cup A_2 \subseteq S\backslash B_2 \in \Pi_\beta$, we know that α and β are μ-orthogonal. Hence, *Exchangeability* requires that $\mu^{\alpha\beta} = \mu^{\beta\alpha}$, which implies $\mu^{\alpha\beta}(A_1\cup A_2) = \mu^{\beta\alpha}(A_1\cup A_2)$.

If $\mu^\alpha(B_2) > 0$, β belongs to condition (iii) (which we have established) if the DM has prior μ^α. In this case, updating on β is equivalent to conditioning on $S\backslash B_2$ according to Bayes's rule. If $\mu^\alpha(B_2) = 0$, then by *Stationarity*, $\mu^{\alpha\beta} = \mu^\alpha$. In both cases, since $A_1\cup A_2 \subseteq S\backslash B_2$,

$$\mu^{\alpha\beta}(A_1\cup A_2) = \frac{\mu^\alpha(A_1\cup A_2)}{1-\mu^\alpha(B_2)}.$$

Moreover, by Lemma 2, since $B_2 \subseteq S\backslash(A_1 \cup A_2)$,

$$\mu^\alpha(B_2) = \frac{\mu(B_2)}{1-\mu(A_1 \cup A_2)}\big(1-\mu^\alpha(A_1 \cup A_2)\big).$$

Combining the preceding equations yields

$$\mu^{\alpha\beta}(A_1 \cup A_2) = \frac{\mu^\alpha(A_1 \cup A_2)}{1-\dfrac{1-\mu^\alpha(A_1 \cup A_2)}{1-\mu(A_1 \cup A_2)}\mu(B_2)}. \tag{A.1}$$

We also know that

$$\mu^\beta = \mu_{S\backslash B_2},$$

in which $\mu_{S\backslash B_2}$ denotes the Bayesian conditional distribution given $S\backslash B_2$ for μ. Hence, the requirement that $\mu^{\alpha\beta}(A_1 \cup A_2) = \mu^{\beta\alpha}(A_1 \cup A_2)$ reduces to

$$\frac{\mu^\alpha(A_1 \cup A_2)}{1-\dfrac{1-\mu^\alpha(A_1 \cup A_2)}{1-\mu(A_1 \cup A_2)}\mu(B_2)} = (\mu_{S\backslash B_2})^\alpha(A_1 \cup A_2). \tag{A.2}$$

Note that equation (A.2) is true for any $(\mu, \alpha) \in \mathbb{I}(S, \Sigma)$ and $B_2 \subseteq S\backslash(A_1 \cup A_2)$ such that α is standard, $0 < \mu(A_1) < \mu(A_2) < 1 - \mu(A_1)$, and $0 < \mu(B_2) < 1 - \mu(A_1 \cup A_2)$.

The next lemma will turn equation (A.2) into a functional equation. It states that the posterior probability of $A_1 \cup A_2$ upon receiving a standard statement α only depends on how the prior behaves over $A_1 \cup A_2$.

Lemma 3. *Let $(\mu, \alpha) \in \mathbb{I}(S, \Sigma)$ be such that α is standard and $0 < \mu(A_1) < \mu(A_2) < 1 - \mu(A_1)$. If the updating rule satisfies Stationarity and Exchangeability, then $\mu^\alpha(A_1 \cup A_2) = \nu^\alpha(A_1 \cup A_2)$ for any $\nu \in \Delta(S, \Sigma)$ such that $\nu(E) = \mu(E)$ for any $E \subseteq A_1 \cup A_2$.*

Proof. Let α be a standard statement such that $0 < \mu(A_1) < \mu(A_2) < 1 - \mu(A_1)$. Clearly, α is μ-credible. Pick A such that $A \subseteq S\backslash(A_1 \cup A_2)$ and $0 < \mu(A) < 1 - \mu(A_1 \cup A_2)$. Such A exists since μ is nonatomic and $\mu(A_1 \cup A_2) < 1$. Let $B = S\backslash(A \cup A_1 \cup A_2)$. Write $\gamma = (\emptyset, A)$. The rest of the proof is broken into three steps.

Step 1: If $\mu(E) = \nu(E)$ for any $E \subseteq S\backslash A$, then $\mu^\alpha(A_1 \cup A_2) = \nu^\alpha(A_1 \cup A_2)$.

This claim basically states that $\mu^\alpha(A_1 \cup A_2)$, as a function of μ, does not depend on how μ behaves over A. First, note that $A_2 \subseteq S\backslash A$ implies that $\nu(A_2) = \mu(A_2) < 1$, and thus α is ν-credible.

Second, note that since $\mu(A) < 1$, $\gamma = (\emptyset, A)$ is a standard, μ-credible statement. Since $\mu(S\backslash A) = \nu(S\backslash A)$, we have $\nu(A) = \mu(A) < 1$, which implies that $\gamma = (\emptyset, A)$ is also ν-credible. Observe that $D^\mu_\alpha = A_1 \cup A_2 \subseteq S\backslash A \in \Pi_\gamma$, implying that α and γ are μ-orthogonal. Furthermore, by $A_1 \subseteq S\backslash A$, we have $\nu(A_1) = \mu(A_1) > 0$, which implies that α is not ν-degenerate. It follows that $D^\nu_\alpha = A_1 \cup A_2 \subseteq S\backslash A \in \Pi_\gamma$, implying that α and γ are ν-orthogonal. By *Exchangeability*, we have both $\mu^{\alpha\gamma} = \mu^{\gamma\alpha}$ and $\nu^{\alpha\gamma} = \nu^{\gamma\alpha}$.

Third, note that since $\mu(E) = \nu(E)$ for any $E \subseteq S\backslash A$, by condition (iii) (which we have established), $\mu^\gamma = \mu_{S\backslash A} = \nu_{S\backslash A} = \nu^\gamma$. It follows that $\mu^{\gamma\alpha} = \nu^{\gamma\alpha}$. Thus, $\mu^{\alpha\gamma} = \mu^{\gamma\alpha} =$

$\nu^{\gamma\alpha} = \nu^{\alpha\gamma}$. In particular, we have $\mu^{\alpha\gamma}(A_1 \cup A_2) = \nu^{\alpha\gamma}(A_1 \cup A_2)$. By the same logic in the derivation of equation (A.1), $\mu^{\alpha\gamma}(A_1 \cup A_2) = \nu^{\alpha\gamma}(A_1 \cup A_2)$ implies that

$$\frac{\mu^{\alpha}(A_1 \cup A_2)}{1 - \dfrac{1-\mu^{\alpha}(A_1 \cup A_2)}{1-\mu(A_1 \cup A_2)} \cdot \mu(A)} = \frac{\nu^{\alpha}(A_1 \cup A_2)}{1 - \dfrac{1-\nu^{\alpha}(A_1 \cup A_2)}{1-\nu(A_1 \cup A_2)} \cdot \nu(A)}. \tag{A.3}$$

Since $A_1 \cup A_2, S\backslash A \subseteq S\backslash A$, we have $\nu(A_1 \cup A_2) = \mu(A_1 \cup A_2) < 1$ and $\nu(A) = \mu(A) < 1 - \mu(A_1 \cup A_2)$. Thus, (A.3) reduces to $\mu^{\alpha}(A_1 \cup A_2) = \nu^{\alpha}(A_1 \cup A_2)$.

Step 2: If $\mu(A) = \nu(A)$ and $\mu(E) = \nu(E)$ for any $E \subseteq A_1 \cup A_2$, then $\mu^{\alpha}(A_1 \cup A_2) = \nu^{\alpha}(A_1 \cup A_2)$.

The claim in Step 2 is stronger than the claim in Step 1, in that now μ and ν only need to agree on A and any event within $A_1 \cup A_2$.

To prove the claim, let $\hat{\mu}$ be given by

$$\hat{\mu} = \mu_{S\backslash A} \cdot \mu(S\backslash A) + \nu_A \cdot \mu(A),$$

in which $\mu_{S\backslash A}$ is the Bayesian conditional distribution given $S\backslash A$ for μ, and ν_A is the Bayesian conditional distribution given A for ν. Observe that $\hat{\mu}$ is a well-defined nonatomic probability measure. Moreover, for any $E \subseteq S\backslash A$, $\mu(E) = \hat{\mu}(E)$; by $\mu(A) = \nu(A)$, for any $E \subseteq A$, $\nu(E) = \hat{\mu}(E)$. In other words, $\hat{\mu}$ agrees with μ within $S\backslash A$ and agrees with ν within A.

The fact that $\hat{\mu}$ agrees with μ within $S\backslash A$, by Step 1, implies that $\mu^{\alpha}(A_1 \cup A_2) = \hat{\mu}^{\alpha}(A_1 \cup A_2)$. Since ν and μ agree within $A_1 \cup A_2$, and $A_1 \cup A_2 \subseteq S\backslash A$, we have $\hat{\mu}$ agrees with ν within $A_1 \cup A_2$. Combined with the fact that $\hat{\mu}$ agrees with ν within A, we conclude that $\hat{\mu}$ agrees with ν within $A \cup A_1 \cup A_2$. Recall that we let $B = S\backslash(A \cup A_1 \cup A_2)$. Thus, $\hat{\mu}$ agrees with ν within $S\backslash B$. Then, applying Step 1 with A replaced by B and μ replaced by $\hat{\mu}$ yields $\nu^{\alpha}(A_1 \cup A_2) = \hat{\mu}^{\alpha}(A_1 \cup A_2)$, by which we conclude that $\mu^{\alpha}(A_1 \cup A_2) = \nu^{\alpha}(A_1 \cup A_2)$.[13]

Step 3: If $\mu(E) = \nu(E)$ for any $E \subseteq A_1 \cup A_2$, then $\mu^{\alpha}(A_1 \cup A_2) = \nu^{\alpha}(A_1 \cup A_2)$.

The claim in Step 3 is even stronger. Now μ and ν only need to agree within $A_1 \cup A_2$.

Suppose $\nu(A) = 0$ or $\nu(B) = 0$. WLOG, assume that $\nu(B) = 0$. Then it is clear that $\nu(A) = 1 - \nu(A_1 \cup A_2) = 1 - \mu(A_1 \cup A_2) > 0$. Let $\hat{\mu}$ be defined in the same way as in Step 2. Since $\hat{\mu}$ and μ agree within $S\backslash A$, by Step 1, $\mu^{\alpha}(A_1 \cup A_2) = \hat{\mu}^{\alpha}(A_1 \cup A_2)$.

Pick $A' \subseteq A$ such that $\nu(A') = \nu(A)/2$ and $B' \subseteq B$ such that $\mu(B') = \mu(B)/2$. Such A' and B' exist because μ, ν are both convex-ranged. Then

$$\nu(A' \cup B') = \nu(A') = \frac{1 - \nu(A_1 \cup A_2)}{2} = \frac{1 - \mu(A_1 \cup A_2)}{2} = \hat{\mu}(A' \cup B'),$$

in which the last equality follows from the following facts: $\hat{\mu}_A = \nu_A$, which implies that $\hat{\mu}(A') = \hat{\mu}(A)/2$; $\hat{\mu}$ agrees with μ within $S\backslash A$, which implies that $\hat{\mu}(B') = \hat{\mu}(B)/2$ and $\hat{\mu}(A_1 \cup A_2) = \mu(A_1 \cup A_2)$; and $A \cup B = S\backslash(A_1 \cup A_2)$. Note that by construction, $\hat{\mu}$ and μ agree within $A_1 \cup A_2$, which together with the assumption of Step 3, implies that $\hat{\mu}$ and

[13]The fact that $\hat{\mu}$ agrees with μ within $S\backslash A$ implies that $0 < \hat{\mu}(A_1) < \hat{\mu}(A_2) < 1 - \hat{\mu}(A_1)$, and $0 < \hat{\mu}(B) < 1 - \hat{\mu}(A_1 \cup A_2)$, which together ensure that the claim in Step 1 is applicable.

ν agree within $A_1 \cup A_2$. Then, applying Step 2 with A replaced by $A' \cup B'$ and μ replaced by $\hat{\mu}$ yields $\nu^\alpha(A_1 \cup A_2) = \hat{\mu}^\alpha(A_1 \cup A_2) = \mu^\alpha(A_1 \cup A_2)$.

Now suppose $\nu(A) > 0$ and $\nu(B) > 0$. Construct $\hat{\nu}$ as follows:

$$\hat{\nu} = \mu_{A_1 \cup A_2} \cdot \mu(A_1 \cup A_2) + \nu_A \cdot \mu(A) + \nu_B \cdot \mu(B).$$

Observe that $\hat{\nu} \in \Delta(S, \Sigma)$, and $\hat{\nu}$ agrees with μ (and thus ν) within $A_1 \cup A_2$. By Step 2, since $\hat{\nu}$ agrees with μ within $A_1 \cup A_2$, and $\hat{\nu}(A) = \mu(A)$, it is clear that $\mu^\alpha(A_1 \cup A_2) = \hat{\nu}^\alpha(A_1 \cup A_2)$.

Pick $A' \subseteq A$ such that $\nu(A') = \nu(A)/2$ and $B' \subseteq B$ such that $\nu(B') = \nu(B)/2$. Such A' and B' exist because ν is convex-ranged. Then

$$\nu(A' \cup B') = \frac{1 - \nu(A_1 \cup A_2)}{2} = \frac{1 - \mu(A_1 \cup A_2)}{2} = \hat{\nu}(A' \cup B'),$$

in which the last equality follows from the following facts: $\hat{\nu}_A = \nu_A$, which implies that $\hat{\nu}(A') = \hat{\nu}(A)/2$; $\hat{\nu}_B = \nu_B$, which implies that $\hat{\nu}(B') = \hat{\nu}(B)/2$; $\hat{\nu}$ agrees with μ within $A_1 \cup A_2$, which implies that $\hat{\nu}(A_1 \cup A_2) = \mu(A_1 \cup A_2)$; and $A \cup B = S \backslash (A_1 \cup A_2)$. Then, since $\hat{\nu}$ agrees with ν within $A_1 \cup A_2$, and $\nu(A' \cup B') = \hat{\nu}(A' \cup B')$, applying Step 2 with A replaced by $A' \cup B'$ and μ replaced by $\hat{\nu}$ finishes the proof. □

Next, I apply Lemma 3 to turn equation (A.2) into a functional equation.

Fix some $(\mu, \alpha) \in \mathbb{I}(S, \Sigma)$ such that α is standard, and $0 < \mu(A_1) < \mu(A_2) < 1 - \mu(A_1)$. Define $g : (0, 1) \to [0, 1]$ as the following mapping:

$$\nu(A_1 \cup A_2) \mapsto \nu^\alpha(A_1 \cup A_2)$$

for any $\nu \in \Delta(S, \Sigma)$ such that $\nu_{A_1 \cup A_2} = \mu_{A_1 \cup A_2}$.

To see why g is well-defined, note that for any $\nu \in \Delta(S, \Sigma)$ such that $\nu(A_1 \cup A_2) \in (0, 1)$, by $\nu(A_2) < \nu(A_1 \cup A_2)$, α is ν-credible. Furthermore, for any $\nu, \tilde{\nu} \in \Delta(S, \Sigma)$ such that $\nu_{A_1 \cup A_2} = \tilde{\nu}_{A_1 \cup A_2} = \mu_{A_1 \cup A_2}$ and $\nu(A_1 \cup A_2) = \tilde{\nu}(A_1 \cup A_2) \in (0, 1)$, by Lemma 3, $\nu^\alpha(A_1 \cup A_2) = \tilde{\nu}^\alpha(A_1 \cup A_2)$. Here, Lemma 3 is applicable because $\nu_{A_1 \cup A_2} = \mu_{A_1 \cup A_2}$ and $\nu(A_1 \cup A_2) < 1$ together imply that $0 < \nu(A_1) < \nu(A_2) < 1 - \nu(A_1)$.

Given function g, equation (A.2) is equivalent to

$$\frac{g\big(\mu(A_1 \cup A_2)\big)}{1 - \dfrac{1 - g\big(\mu(A_1 \cup A_2)\big)}{1 - \mu(A_1 \cup A_2)} \cdot \mu(B_2)} = g\left(\frac{\mu(A_1 \cup A_2)}{1 - \mu(B_2)}\right),$$

in which we have used the facts that $\mu_{A_1 \cup A_2} = (\mu_{S \backslash B_2})_{A_1 \cup A_2}$ and $\mu_{S \backslash B_2}(A_1 \cup A_2) = \mu(A_1 \cup A_2)/(1 - \mu(B_2))$. Verify that since $\mu(B_2) < 1 - \mu(A_1 \cup A_2)$, the RHS is well-defined.

Recall that equation (A.2) is true for any $(\mu, \alpha) \in \mathbb{I}(S, \Sigma)$ and $B_2 \subseteq S \backslash (A_1 \cup A_2)$ such that α is standard, $0 < \mu(A_1) < \mu(A_2) < 1 - \mu(A_1)$, and $0 < \mu(B_2) < 1 - \mu(A_1 \cup A_2)$. Fixing α, we can always vary $\mu(A_1 \cup A_2)$ within $(0, 1)$ and $\mu(B_2)$ within $(0, 1 - \mu(A_1 \cup A_2))$ continuously without affecting the conditional distribution over $A_1 \cup A_2$. In the process, since the statement α and the conditional distribution over $A_1 \cup A_2$ are both fixed, the same g applies. Thus, by letting $\mu(B_2) = b$ and $\mu(A_1 \cup A_2) = x$, the equation above reads

$$\frac{g(x)}{1 - \frac{1-g(x)}{1-x} \cdot b} = g\left(\frac{x}{1-b}\right), \tag{A.4}$$

which holds for $0 < x < 1$ and $0 < b < 1 - x$.

Lemma 4. *The solutions to functional equation (A.4) are of the form*

$$g(x) = \frac{a}{2a - 1 + \frac{1-a}{x}}, \tag{A.5}$$

where $a \in [0, 1]$.

Proof. Let $g(1/2) = a \in [0, 1]$. For any $0 < x < 1/2$, substituting $b = 1 - 2x$ into (A.4) yields

$$a = g\left(\frac{1}{2}\right) = \frac{g(x)}{1 - \frac{1-g(x)}{1-x} \cdot (1-2x)},$$

which implies equation (A.5). For $1/2 < y < 1$, substituting $x = 1/2$ and $b = 1 - 1/(2y)$ into (A.4) yields

$$\frac{g\left(\frac{1}{2}\right)}{1 - 2\left(1 - g\left(\frac{1}{2}\right)\right) \cdot \left(1 - \frac{1}{2y}\right)} = g(y),$$

which reads

$$g(y) = \frac{a}{2a - 1 + \frac{1-a}{y}}.$$

Clearly, there are no other solutions to the functional equation, since $g(1/2)$ uniquely defines $g(x)$ on $(0, 1/2)$ and $(1/2, 1)$. □

The final step to establish condition (iib) is to show that $a = 1/2$, so that $g(x) = x$. Recall that in the definition of g we have fixed μ to be such that $0 < \mu(A_1) < \mu(A_2) < 1 - \mu(A_1)$. Construct $\nu \in \Delta(S, \Sigma)$ as follows:

$$\nu = \frac{1}{2}\mu_{A_1 \cup A_2} + \frac{1}{2}\mu_{S \backslash (A_1 \cup A_2)}.$$

Observe that ν is well-defined, $\nu_{A_1 \cup A_2} = \mu_{A_1 \cup A_2}$, and $\nu(A_1 \cup A_2) = \nu(S \backslash (A_1 \cup A_2)) = 1/2$.

Suppose $a > 1/2$. Then we have $\nu^\alpha(A_1 \cup A_2) = g(\nu(A_1 \cup A_2)) = g(1/2) = a > 1/2 = \nu(A_1 \cup A_2)$. By $\nu^\alpha(A_1 \cup A_2) > \nu(A_1 \cup A_2)$, we have

$$\nu^\alpha\big(S \backslash (A_1 \cup A_2)\big) < \nu\big(S \backslash (A_1 \cup A_2)\big).$$

It follows that $\nu^\alpha(S\backslash(A_1 \cup A_2)) < \nu^\alpha(A_1 \cup A_2)$. Consider the standard statement $\gamma = (S\backslash(A_1 \cup A_2), A_1 \cup A_2)$. By $\nu(S\backslash(A_1 \cup A_2)) = \nu(A_1 \cup A_2)$, γ is ν-credible. Then, by *Stationarity*, $\nu^\gamma = \nu$, and thus $\nu^{\gamma\alpha} = \nu^\alpha$. Moreover, since $D^\nu_\alpha = A_1 \cup A_2 \in \Pi_\gamma$, α and γ are ν-orthogonal. Thus, by *Exchangeability*, $\nu^{\alpha\gamma} = \nu^{\gamma\alpha} = \nu^\alpha$. It follows that $\nu^{\alpha\gamma}(S\backslash(A_1 \cup A_2)) < \nu^{\alpha\gamma}(A_1 \cup A_2)$, a contradiction to the definition of updating rules. Note that the case $a = 1$ is, in fact, directly ruled out by *Exchangeability*: If $a = 1$, then $\nu^\alpha(A_1 \cup A_2) = 1$ and $\nu^\alpha(S\backslash(A_1 \cup A_2)) = 0$, which implies that γ is ν^α-noncredible and thus $\nu^{\alpha\gamma}$ is not well-defined, contradicting *Exchangeability*.

Suppose $a < 1/2$. Then we have $\nu^\alpha(A_1 \cup A_2) = g(\nu(A_1 \cup A_2)) = g(1/2) = a < 1/2 = \nu(A_1 \cup A_2)$. By $\nu^\alpha(A_1 \cup A_2) < \nu(A_1 \cup A_2)$, we have

$$\nu^\alpha\big(S\backslash(A_1 \cup A_2)\big) > \nu\big(S\backslash(A_1 \cup A_2)\big).$$

It follows that $\nu^\alpha(S\backslash(A_1 \cup A_2)) > \nu^\alpha(A_1 \cup A_2)$. Consider the standard statement $\delta = (A_1 \cup A_2, S\backslash(A_1 \cup A_2))$. By $\nu(S\backslash(A_1 \cup A_2)) = \nu(A_1 \cup A_2)$, δ is ν-credible. Then, by *Stationarity*, $\nu^\delta = \nu$. It then follows that $\nu^{\delta\alpha} = \nu^\alpha$. Moreover, since $D^\nu_\alpha = A_1 \cup A_2 \in \Pi_\delta$, α and δ are ν-orthogonal. Thus, by *Exchangeability*, $\nu^{\alpha\delta} = \nu^{\delta\alpha} = \nu^\alpha$. It follows that $\nu^{\alpha\delta}(S\backslash(A_1 \cup A_2)) > \nu^{\alpha\delta}(A_1 \cup A_2)$, a contradiction to the definition of updating rules. Note that the case $a = 0$ is, in fact, directly ruled out by *Exchangeability*: If $a = 0$, then $\nu^\alpha(A_1 \cup A_2) = 0$ and $\nu^\alpha(S\backslash(A_1 \cup A_2)) = 1$, which implies that δ is ν^α-noncredible, and thus $\nu^{\alpha\delta}$ is not well-defined, contradicting *Exchangeability*.

Therefore, we must have $a = 1/2$, and thus $g(x) = x$ for all $x \in (0, 1)$. Note that although g depends on the pair (μ, α) that we have fixed in its definition, the preceding arguments show that for any (μ, α) such that α is standard, and $0 < \mu(A_1) < \mu(A_2) < 1 - \mu(A_1)$, the corresponding g always satisfies $g(x) = x$ for all $x \in (0, 1)$, establishing condition (iib) in Theorem 1.

To complete the "only if" part, I now turn to condition (iic). It suffices to show that for any $(\mu, \alpha) \in \mathbb{I}(S, \Sigma)$ such that α is standard and $0 < \mu(A_1) < \mu(A_2)$, and for any $\nu \in \Delta(S, \Sigma)$ that is α-*connected* to μ,

$$\frac{\mu(A_2)}{\mu(A_1)} = \frac{\nu(A_2)}{\nu(A_1)} \quad \Rightarrow \quad \frac{\mu^\alpha(A_2)}{\mu^\alpha(A_1)} = \frac{\nu^\alpha(A_2)}{\nu^\alpha(A_1)}.$$

First, consider the following lemma, which states that the Bayesian conditional distribution given $A_1 \cup A_2$ for the posterior only depends on the Bayesian conditional distribution given $A_1 \cup A_2$ for the prior.

Lemma 5. *Let* $(\mu, \alpha) \in \mathbb{I}(S, \Sigma)$ *be such that* α *is standard and* $0 < \mu(A_1) < \mu(A_2)$. *If an updating rule satisfies Exchangeability and Stationarity, then for* $\nu \in \Delta(S, \Sigma)$ *such that* $\nu_{A_1 \cup A_2} = \mu_{A_1 \cup A_2}$, *we have*

$$\big(\mu^\alpha\big)_{A_1 \cup A_2} = \big(\nu^\alpha\big)_{A_1 \cup A_2}.$$

Proof. First, verify that α is ν-credible, since $\nu_{A_1 \cup A_2} = \mu_{A_1 \cup A_2}$ implies that

$$\frac{\nu(A_1)}{\nu(A_2)} = \frac{\mu(A_1)}{\mu(A_2)} > 0.$$

Now, consider $\gamma = (\emptyset, S\backslash(A_1 \cup A_2))$. Since $\mu_{A_1\cup A_2}$ and $\nu_{A_1\cup A_2}$ are well-defined, γ is both μ-credible and ν-credible. Observe that $D^\mu_\alpha = D^\nu_\alpha = A_1 \cup A_2 \in \Pi_\gamma$, which implies that α and γ are both μ-orthogonal and ν-orthogonal. *Exchangeability* then requires that $\mu^{\gamma\alpha} = \mu^{\alpha\gamma}$, and $\nu^{\gamma\alpha} = \nu^{\alpha\gamma}$. Moreover, by Lemma 2, it is clear that $\mu^\gamma = \mu_{A_1\cup A_2} = \nu_{A_1\cup A_2} = \nu^\gamma$, and thus $\mu^{\gamma\alpha} = \nu^{\gamma\alpha}$. It follows that $\mu^{\alpha\gamma} = \nu^{\alpha\gamma}$, which by Lemma 2, implies that $(\mu^\alpha)_{A_1\cup A_2} = (\nu^\alpha)_{A_1\cup A_2}$. Note that by condition (iib) (which we have established), $\mu^\alpha(A_1 \cup A_2) = \mu(A_1 \cup A_2) > 0$, and $\nu^\alpha(A_1 \cup A_2) = \nu(A_1 \cup A_2) > 0$, which implies that $(\mu^\alpha)_{A_1\cup A_2}$ and $(\nu^\alpha)_{A_1\cup A_2}$ are well-defined. □

Now we are ready to prove condition (iic). Let $(\mu, \alpha) \in \mathbb{I}(S, \Sigma)$ be such that α is standard and $0 < \mu(A_1) < \mu(A_2)$, and $\nu \in \Delta(S, \Sigma)$ be *α-connected* to μ with $\mu(A_2)/\mu(A_1) = \nu(A_2)/\nu(A_1)$. Suppose that ν is *α-connected* to μ via some standard statement $\beta = (B_1, B_2)$. In other words, α and β are μ-orthogonal, and for $i = 1, 2$,

$$\left(\mu^\beta\right)_{A_i} = \nu_{A_i}. \tag{A.6}$$

First, I show that

$$\frac{\mu^\beta(A_2)}{\mu^\beta(A_1)} = \frac{\mu(A_2)}{\mu(A_1)}. \tag{A.7}$$

If $\mu(B_1) \geq \mu(B_2)$, by *Stationarity*, there is nothing to prove. Thus, I focus on the case in which $\mu(B_1) < \mu(B_2)$. Since α and β are μ-orthogonal, we have the following two cases:

Case 1: $D^\mu_\alpha = A_1 \cup A_2 \subseteq P$ for some $P \in \Pi_\beta$. Since $(\mu^\beta)_{A_i}$ is well-defined for $i = 1, 2$, it is clear that $\mu^\beta(P) \geq \mu^\beta(A_1 \cup A_2) > 0$. Then by Lemma 2, we have

$$\frac{\mu^\beta(A_1)}{\mu^\beta(P)} = \frac{\mu(A_1)}{\mu(P)} \quad \text{and} \quad \frac{\mu^\beta(A_2)}{\mu^\beta(P)} = \frac{\mu(A_2)}{\mu(P)},$$

which implies that $\mu^\beta(A_2)/\mu^\beta(A_1) = \mu(A_2)/\mu(A_1)$.

Case 2: $D^\mu_\beta \subseteq P'$ for some $P' \in \Pi_\alpha$. For this to be true, β cannot be μ-degenerate, and thus $D^\mu_\beta = B_1 \cup B_2$. By $\mu(B_1) < \mu(B_2)$, we have $0 < \mu(B_1) < \mu(B_2)$. By condition (iib) in Theorem 1 (which we have already established), we have $\mu^\beta(B_1 \cup B_2) = \mu(B_1 \cup B_2)$. Then Lemma 2 implies that $\mu^\beta(E) = \mu(E)$ for any $E \subseteq S\backslash(B_1 \cup B_2)$. In other words, μ and μ^β agree within $S\backslash(B_1 \cup B_2)$. Thus,

$$\begin{aligned}\mu^\beta(P') &= \mu^\beta(B_1 \cup B_2) + \mu^\beta\big(P'\backslash(B_1 \cup B_2)\big)\\ &= \mu(B_1 \cup B_2) + \mu\big(P'\backslash(B_1 \cup B_2)\big)\\ &= \mu(P').\end{aligned}$$

Furthermore, any other $P'' \in \Pi_\alpha$ satisfies $P'' \subseteq S\backslash(B_1 \cup B_2)$, which implies $\mu^\beta(P'') = \mu(P'')$. Thus, we always have $\mu^\beta(A_i) = \mu(A_i)$ for $i = 1, 2$, establishing (A.7).

Observe that since $\mu(A_2)/\mu(A_1) = \nu(A_2)/\nu(A_1)$, by (A.6) and (A.7), we have $(\mu^\beta)_{A_1\cup A_2} = \nu_{A_1\cup A_2}$. Applying Lemma 5 with μ replaced by μ^β yields $(\mu^{\beta\alpha})_{A_1\cup A_2} = (\nu^\alpha)_{A_1\cup A_2}$, which implies

$$\frac{\mu^{\beta\alpha}(A_2)}{\mu^{\beta\alpha}(A_1)} = \frac{\nu^\alpha(A_2)}{\nu^\alpha(A_1)}.$$

Note that by the definition of an updating rule, we have $\mu^{\beta\alpha}(A_1) \geq \mu^{\beta\alpha}(A_2)$ and $\nu^{\alpha}(A_1) \geq \nu^{\alpha}(A_2)$. Thus, even if $\nu^{\alpha}(A_1) = 0$ or $\mu^{\beta\alpha}(A_1) = 0$, the ratios are still well-defined due to the convention $0/0 = 0$. Applying *Exchangeability* yields

$$\frac{\mu^{\alpha\beta}(A_2)}{\mu^{\alpha\beta}(A_1)} = \frac{\nu^{\alpha}(A_2)}{\nu^{\alpha}(A_1)}. \tag{A.8}$$

The last step is to show

$$\frac{\mu^{\alpha\beta}(A_2)}{\mu^{\alpha\beta}(A_1)} = \frac{\mu^{\alpha}(A_2)}{\mu^{\alpha}(A_1)}. \tag{A.9}$$

The argument is similar to that for (A.7). If $\mu^{\alpha}(B_1) \geq \mu^{\alpha}(B_2)$, by *Stationarity*, there is nothing to prove. If $\mu(B_1) \geq \mu(B_2)$, by *Stationarity* and *Exchangeability*, $\mu^{\alpha\beta} = \mu^{\beta\alpha} = \mu^{\alpha}$ and we are done. Henceforth, I will assume $\mu(B_1) < \mu(B_2)$ and $\mu^{\alpha}(B_1) < \mu^{\alpha}(B_2)$. Again, we need to consider two cases.

Case 1: $D^{\mu}_{\alpha} = A_1 \cup A_2 \subseteq P$ for some $P \in \Pi_{\beta}$. Since $(\mu^{\beta})_{A_i}$ is well-defined for $i = 1, 2$, it is clear that $\mu^{\beta}(P) \geq \mu^{\beta}(A_1 \cup A_2) > 0$. By Lemma 2, $\mu_P = (\mu^{\beta})_P$, which implies that $\mu_{A_1\cup A_2} = (\mu^{\beta})_{A_1\cup A_2}$. It follows that $0 < \mu^{\beta}(A_1) < \mu^{\beta}(A_2)$. Applying condition (iib) in Theorem 1 (which we have established) yields $\mu^{\beta\alpha}(A_1 \cup A_2) = \mu^{\beta}(A_1 \cup A_2) > 0$. Thus, $\mu^{\alpha\beta}(P) = \mu^{\beta\alpha}(P) \geq \mu^{\beta\alpha}(A_1 \cup A_2) > 0$. By Lemma 2,

$$\mu^{\alpha\beta}(A_1) = \frac{\mu^{\alpha}(A_1)}{\mu^{\alpha}(P)}\mu^{\alpha\beta}(P) \quad \text{and} \quad \mu^{\alpha\beta}(A_2) = \frac{\mu^{\alpha}(A_2)}{\mu^{\alpha}(P)}\mu^{\alpha\beta}(P).$$

If $\mu^{\alpha}(P) = 0$, then Lemma 2 implies that $\mu^{\alpha\beta}(P) = 0$, a contradiction. Hence, $\mu^{\alpha}(P) > 0$. Combining the two equations above, $\mu^{\alpha}(P) > 0$, and $\mu^{\alpha\beta}(P) > 0$ yields (A.9).

Case 2: $D^{\mu}_{\beta} \subseteq P'$ for some $P' \in \Pi_{\alpha}$. For this to be true, β cannot be μ-degenerate, and thus $D^{\mu}_{\beta} = B_1 \cup B_2$. By $\mu(B_1) < \mu(B_2)$, it then follows that $0 < \mu(B_1) < \mu(B_2)$. Furthermore, since $\mu^{\alpha}(B_1) < \mu^{\alpha}(B_2)$, it must be the case that $\mu^{\alpha}(P') > 0$. Then by Lemma 2,

$$\frac{\mu^{\alpha}(B_1)}{\mu^{\alpha}(B_2)} = \frac{\mu(B_1)}{\mu(B_2)},$$

which implies that $0 < \mu^{\alpha}(B_1) < \mu^{\alpha}(B_2)$. Applying condition (iib) in Theorem 1 (which we have established) yields $\mu^{\alpha\beta}(B_1 \cup B_2) = \mu^{\alpha}(B_1 \cup B_2)$. Then, Lemma 2 implies that $\mu^{\alpha\beta}(E) = \mu^{\alpha}(E)$ for any $E \subseteq S\backslash(B_1 \cup B_2)$. In other words, $\mu^{\alpha\beta}$ and μ^{α} agree within $S\backslash(B_1 \cup B_2)$. Thus,

$$\begin{aligned}
\mu^{\alpha\beta}(P') &= \mu^{\alpha\beta}(B_1 \cup B_2) + \mu^{\alpha\beta}\big(P'\backslash(B_1 \cup B_2)\big) \\
&= \mu^{\alpha}(B_1 \cup B_2) + \mu^{\alpha}\big(P'\backslash(B_1 \cup B_2)\big) \\
&= \mu^{\alpha}(P').
\end{aligned}$$

Furthermore, any other $P'' \in \Pi_{\alpha}$ satisfies $P'' \subseteq S\backslash(B_1 \cup B_2)$, which implies $\mu^{\alpha\beta}(P'') = \mu^{\alpha}(P'')$. Thus, we always have $\mu^{\alpha\beta}(A_i) = \mu^{\alpha}(A_i)$ for $i = 1, 2$, establishing (A.9).

Combining (A.8) and (A.9), we have proved that for any $(\mu, \alpha) \in \mathbb{I}(S, \Sigma)$ such that α is standard and $0 < \mu(A_1) < \mu(A_2)$, and $\nu \in \Delta(S, \Sigma)$ be α-*connected* to μ, we have

$$\frac{\mu(A_2)}{\mu(A_1)} = \frac{\nu(A_2)}{\nu(A_1)} \quad \Rightarrow \quad \frac{\mu^\alpha(A_2)}{\mu^\alpha(A_1)} = \frac{\nu^\alpha(A_2)}{\nu^\alpha(A_1)},$$

which concludes the proof of the "only if" part of Theorem 1.

Next, I show the "if" part. Clearly, condition (i) implies *Stationarity*. In addition, since $L_\alpha = L_{(L_\alpha, R_\alpha)}$, $R_\alpha = R_{(L_\alpha, R_\alpha)}$, and $C_\alpha = C_{(L_\alpha, R_\alpha)}$, *Reduction* holds. It suffices to show *Exchangeability*.

Let $(\mu, \alpha), (\mu, \beta) \in \mathbb{I}(S, \Sigma)$ be such that α and β are μ-orthogonal. We want to show that $\mu^{\alpha\beta} = \mu^{\beta\alpha}$. WLOG, assume that $D^\mu_\alpha \subseteq P$ for some $P \in \Pi_\beta$. There are three cases:

Case 1: α is μ-degenerate. Then $D^\mu_\alpha = S \in \Pi_\beta$. Therefore, β must be $(\emptyset, \emptyset)$ or $(S, \emptyset)$, since $(\emptyset, S)$ is μ-noncredible (and thus α and $(\emptyset, S)$ cannot be μ-orthogonal). Whether $\beta = (\emptyset, \emptyset)$ or $\beta = (S, \emptyset)$, condition (i) in Theorem 1 implies $\mu^{\alpha\beta} = \mu^{\beta\alpha}$.

Case 2: α is not μ-degenerate but β is μ-degenerate. Thus, $D^\mu_\alpha = A_1 \cup A_2$, and $0 = \mu(B_1) < \mu(B_2)$. For α and β to be μ-orthogonal, both statements need to be μ-credible, which implies $\mu(B_2) < 1$. By condition (iii), we have $\mu^\beta = \mu_{S\setminus B_2}$.

If $P = B_1$, then $\mu(A_1) = \mu(A_2) = \mu(B_1) = 0$. By condition (i), $\mu^\alpha = \mu$. Thus, since β is μ-credible, it is also μ^α-credible. It follows that $\mu^{\alpha\beta} = \mu^\beta$. Since $0 = \mu(B_1) < \mu(B_2)$, by condition (iii), we have $\mu^\beta(B_1) = 0$, which implies that $\mu^\beta(A_1) = \mu^\beta(A_2) = 0$, and thus α is μ^β-credible. Again by condition (i), we have $\mu^{\beta\alpha} = \mu^\beta = \mu^{\alpha\beta}$.

If $P = B_2$, then by $\mu^\beta = \mu_{S\setminus B_2}$, $\mu^\beta(A_1) = \mu^\beta(A_2) = \mu^\beta(B_2) = 0$, which implies that α is μ^β-credible. By condition (i), $\mu^{\beta\alpha} = \mu^\beta = \mu_{S\setminus B_2}$. Since α is not μ-degenerate, either $\mu(A_1) \geq \mu(A_2)$ or $0 < \mu(A_1) < \mu(A_2)$. If $\mu(A_1) \geq \mu(A_2)$, by condition (i), $\mu^\alpha = \mu$. Since β is μ-credible, it is also μ^α-credible. Thus, $\mu^{\alpha\beta} = \mu^\beta = \mu^{\beta\alpha}$ and we are done. Now suppose $0 < \mu(A_1) < \mu(A_2)$. By condition (iib), we always have $\mu^\alpha(A_1 \cup A_2) = \mu(A_1 \cup A_2)$. Then condition (iia) implies that $\mu^\alpha(E) = \mu(E)$ for any $E \subseteq S\setminus(A_1 \cup A_2)$. Hence,

$$\begin{aligned} \mu^\alpha(B_2) &= \mu^\alpha(A_1 \cup A_2) + \mu^\alpha\big(B_2\setminus(A_1 \cup A_2)\big) \\ &= \mu(A_1 \cup A_2) + \mu\big(B_2\setminus(A_1 \cup A_2)\big) \\ &= \mu(B_2) \in (0, 1), \end{aligned} \tag{A.10}$$

which implies that β is μ^α-credible. Furthermore, by $B_1 \subseteq S\setminus(A_1 \cup A_2)$, $\mu^\alpha(B_1) = \mu(B_1) = 0$. Then, by condition (iii), $\mu^{\alpha\beta} = (\mu^\alpha)_{S\setminus B_2}$. Then since $S\setminus B_2 \subseteq S\setminus(A_1 \cup A_2)$, we have $(\mu^\alpha)_{S\setminus B_2} = \mu_{S\setminus B_2}$, which implies that $\mu^{\alpha\beta} = \mu_{S\setminus B_2} = \mu^{\beta\alpha}$.

If $P = S\setminus(B_1 \cup B_2)$, since α is not μ-degenerate, either $\mu(A_1) \geq \mu(A_2)$ or $0 < \mu(A_1) < \mu(A_2)$. If $\mu(A_1) \geq \mu(A_2)$, then by condition (i), $\mu^\alpha = \mu$. It follows that β is μ^α-credible and $\mu^{\alpha\beta} = (\mu^\alpha)_{S\setminus B_2} = \mu_{S\setminus B_2}$. By the properties of Bayes' rule, since $A_1 \cup A_2 \subseteq P \subseteq S\setminus B_2$, we have $\mu_{S\setminus B_2}(A_1) \geq \mu_{S\setminus B_2}(A_2)$, which implies that $\mu^\beta(A_1) \geq \mu^\beta(A_2)$. By condition (i), it follows that α is μ^β-credible, and $\mu^{\beta\alpha} = \mu^\beta = \mu_{S\setminus B_2} = \mu^{\alpha\beta}$.

Now suppose $0 < \mu(A_1) < \mu(A_2)$. Clearly, μ^β is α-connected to μ (via β). Moreover, by the properties of Bayes' rule, $(\mu^\beta)_{A_1\cup A_2} = (\mu_{S\setminus B_2})_{A_1\cup A_2} = \mu_{A_1\cup A_2}$, which implies that

$\mu^{\beta}(A_2)/\mu^{\beta}(A_1) = \mu(A_2)/\mu(A_1) > 1$ (and thus α is μ^{β}-credible). Thus, by condition (iic),

$$\frac{\mu^{\beta\alpha}(A_2)}{\mu^{\beta\alpha}(A_1)} = \frac{\mu^{\alpha}(A_2)}{\mu^{\alpha}(A_1)}.$$

Since $\mu^{\beta}(A_2)/\mu^{\beta}(A_1) > 1$, by condition (iib), $\mu^{\beta\alpha}(A_1 \cup A_2) = \mu^{\beta}(A_1 \cup A_2)$.

By conditions (iia) and (iib), we have $\mu^{\alpha}(A_1 \cup A_2) = \mu(A_1 \cup A_2)$ and $\mu^{\alpha}(E) = \mu(E)$ for any $E \subseteq S\backslash(A_1 \cup A_2)$. It follows that $\mu^{\alpha}(B_i) = \mu(B_i)$ for $i = 1, 2$, and thus β is μ^{α}-credible. By condition (iii), $\mu^{\alpha\beta} = (\mu^{\alpha})_{S\backslash B_2}$. Hence,

$$\frac{\mu^{\beta\alpha}(A_2)}{\mu^{\beta\alpha}(A_1)} = \frac{\mu^{\alpha}(A_2)}{\mu^{\alpha}(A_1)} = \frac{\mu^{\alpha\beta}(A_2)}{\mu^{\alpha\beta}(A_1)}.$$

Note that

$$\mu^{\beta\alpha}(A_1 \cup A_2) = \mu^{\beta}(A_1 \cup A_2) = \frac{\mu(A_1 \cup A_2)}{1 - \mu(B_2)} = \frac{\mu^{\alpha}(A_1 \cup A_2)}{1 - \mu^{\alpha}(B_2)} = \mu^{\alpha\beta}(A_1 \cup A_2).$$

Combining the two display equations above yields $\mu^{\beta\alpha}(A_i) = \mu^{\alpha\beta}(A_i)$ for $i = 1, 2$. Since $\mu^{\beta}(B_1) = \mu^{\beta}(B_2) = 0$, by condition (iia), $\mu^{\beta\alpha}(B_1) = \mu^{\beta\alpha}(B_2) = 0$. In addition, since $\mu^{\alpha}(B_i) = \mu(B_i)$ for $i = 1, 2$, and $\mu^{\alpha\beta} = (\mu^{\alpha})_{S\backslash B_2}$, we have $\mu^{\alpha\beta}(B_1) = \mu^{\alpha\beta}(B_2) = 0$. Thus, $\mu^{\beta\alpha}(B_i) = \mu^{\alpha\beta}(B_i)$ for $i = 1, 2$. Then, by conditions (iib) and (iii), the probability distribution over each element of $\{A_1, A_2, B_1, B_2, S\backslash(A_1 \cup A_2 \cup B_1 \cup B_2)\}$ is updated in proportion to the prior upon receiving α or β. Thus, we conclude that $\mu^{\alpha\beta} = \mu^{\beta\alpha}$.

Case 3: Neither α nor β is μ-degenerate. Suppose $\mu(B_1) \geq \mu(B_2)$. Condition (i) implies that $\mu^{\beta} = \mu$, and thus α is μ^{β}-credible and $\mu^{\beta\alpha} = \mu^{\alpha}$. If $\mu(A_1) \geq \mu(A_2)$, condition (i) ensures that β is μ^{α}-credible, and $\mu^{\alpha\beta} = \mu^{\beta\alpha} = \mu$. If $0 < \mu(A_1) < \mu(A_2)$, by condition (iia) and (iib), $\mu^{\alpha}(A_1 \cup A_2) = \mu(A_1 \cup A_2)$ and $\mu^{\alpha}(E) = \mu(E)$ for any $E \subseteq S\backslash(A_1 \cup A_2)$. Using the argument along the lines of (A.10), it follows that if $A_1 \cup A_2 \subseteq P$ for some $P \in \Pi_{\beta}$, then $\mu^{\alpha}(B_1) = \mu(B_1)$ and $\mu^{\alpha}(B_2) = \mu(B_2)$. Thus, β is μ^{α}-credible and, by condition (i), $\mu^{\alpha\beta} = \mu^{\alpha} = \mu^{\beta\alpha}$.

Suppose $0 < \mu(B_1) < \mu(B_2)$. If $\mu(A_1) \geq \mu(A_2)$, then condition (i) implies $\mu^{\alpha} = \mu$, and thus β is μ^{α}-credible and $\mu^{\alpha\beta} = \mu^{\beta}$. For any $P \in \Pi_{\beta}$, $A_1 \cup A_2 \subseteq P$ implies, by condition (iia), that

$$\mu^{\beta}(A_1) = \frac{\mu(A_1)}{\mu(P)}\mu^{\beta}(P) \quad \text{and} \quad \mu^{\beta}(A_2) = \frac{\mu(A_2)}{\mu(P)}\mu^{\beta}(P).$$

If $\mu^{\beta}(P) = 0$, then $\mu^{\beta}(A_1) = \mu^{\beta}(A_2) = 0$. If $\mu^{\beta}(P) > 0$ (which implies $\mu(P) > 0$), then $\mu(A_1) \geq \mu(A_2)$ implies $\mu^{\beta}(A_1) \geq \mu^{\beta}(A_2)$. In both cases, α is μ^{β}-credible. Moreover, in both cases, condition (i) implies $\mu^{\beta\alpha} = \mu^{\beta}$.

The final remaining case is when $0 < \mu(B_1) < \mu(B_2)$ and $0 < \mu(A_1) < \mu(A_2)$. Recall that $A_1 \cup A_2 \subseteq P \in \Pi_{\beta}$. By conditions (iia) and (iib), we have $\mu^{\beta}(B_1 \cup B_2) = \mu(B_1 \cup B_2)$ and $\mu^{\beta}(E) = \mu(E)$ for any $E \subseteq S\backslash(B_1 \cup B_2)$; $\mu^{\alpha}(A_1 \cup A_2) = \mu(A_1 \cup A_2)$ and $\mu^{\alpha}(F) = \mu(F)$ for any $F \subseteq S\backslash(A_1 \cup A_2)$.

First, I show that $\mu^{\alpha\beta}(B_1) = \mu^{\beta\alpha}(B_1)$ and $\mu^{\alpha\beta}(B_2) = \mu^{\beta\alpha}(B_2)$. Observe that regardless of which element P is, we always have $\mu^{\alpha}(B_1) = \mu(B_1)$ and $\mu^{\alpha}(B_2) = \mu(B_2)$, which implies that β is μ^{α}-credible, and, by condition (iib),

$$\mu^{\alpha\beta}(B_1 \cup B_2) = \mu^{\alpha}(B_1 \cup B_2) = \mu(B_1 \cup B_2). \tag{A.11}$$

Note that μ^{α} is β-connected to μ (via α), and $\mu^{\alpha}(B_2)/\mu^{\alpha}(B_1) = \mu(B_2)/\mu(B_1)$. Then it follows from condition (iic) that

$$\frac{\mu^{\alpha\beta}(B_2)}{\mu^{\alpha\beta}(B_1)} = \frac{\mu^{\beta}(B_2)}{\mu^{\beta}(B_1)}. \tag{A.12}$$

To characterize the relation between $\mu^{\beta\alpha}$ and μ^{β}, note that if $\mu^{\beta}(P) = 0$, then $\mu^{\beta}(A_1) = \mu^{\beta}(A_2) = 0$. It follows that α is μ^{β}-credible and by condition (i), $\mu^{\beta\alpha} = \mu^{\beta}$. If $\mu^{\beta}(P) > 0$, then by condition (iia), $0 < \mu^{\beta}(A_1) < \mu^{\beta}(A_2)$, which implies that α is μ^{β}-credible. In addition, by conditions (iia) and (iib), we have $\mu^{\beta\alpha}(A_1 \cup A_2) = \mu^{\beta}(A_1 \cup A_2)$, and $\mu^{\beta\alpha}(E) = \mu^{\beta}(E)$ for any $E \subseteq S\backslash(A_1 \cup A_2)$. Regardless of which element P is, we always have $\mu^{\beta\alpha}(B_1) = \mu^{\beta}(B_1)$ and $\mu^{\beta\alpha}(B_2) = \mu^{\beta}(B_2)$. Therefore, whether $\mu^{\beta}(P) > 0$ or not, we have

$$\mu^{\beta\alpha}(B_1 \cup B_2) = \mu^{\beta}(B_1 \cup B_2) = \mu(B_1 \cup B_2) \tag{A.13}$$

and

$$\frac{\mu^{\beta}(B_2)}{\mu^{\beta}(B_1)} = \frac{\mu^{\beta\alpha}(B_2)}{\mu^{\beta\alpha}(B_1)}. \tag{A.14}$$

Combining (A.11), (A.12), (A.13), and (A.14), we have $\mu^{\alpha\beta}(B_1) = \mu^{\beta\alpha}(B_1)$ and $\mu^{\alpha\beta}(B_2) = \mu^{\beta\alpha}(B_2)$.

Now I show that $\mu^{\alpha\beta}(A_1) = \mu^{\beta\alpha}(A_1)$ and $\mu^{\alpha\beta}(A_2) = \mu^{\beta\alpha}(A_2)$. If $\mu^{\beta}(P) = 0$, then it must be the case that $P = B_i$ for some $i \in \{1, 2\}$, since $\mu(P) > 0$ and $\mu^{\beta}(B_1 \cup B_2) = \mu(B_1 \cup B_2)$. Then (A.12) implies that $\mu^{\alpha\beta}(P) = 0$, and thus $\mu^{\alpha\beta}(A_1) = \mu^{\alpha\beta}(A_2) = 0$. Moreover, since $\mu^{\beta}(P) = 0$ and $A_1 \cup A_2 \subseteq P$ together imply $\mu^{\beta}(A_1) = \mu^{\beta}(A_2) = 0$, by condition (i), $\mu^{\beta\alpha} = \mu^{\beta}$. Therefore, $\mu^{\alpha\beta}(A_1) = \mu^{\alpha\beta}(A_2) = \mu^{\beta\alpha}(A_1) = \mu^{\beta\alpha}(A_2) = 0$ and we are done.

If $\mu^{\beta}(P) > 0$, then by condition (iia), we have $\mu^{\beta}(A_2)/\mu^{\beta}(A_1) = \mu(A_2)/\mu(A_1)$. It follows from condition (iib) that

$$\mu^{\beta\alpha}(A_1 \cup A_2) = \mu^{\beta}(A_1 \cup A_2). \tag{A.15}$$

Furthermore, μ^{β} is α-connected to μ (via β). Then by condition (iic),

$$\frac{\mu^{\beta\alpha}(A_2)}{\mu^{\beta\alpha}(A_1)} = \frac{\mu^{\alpha}(A_2)}{\mu^{\alpha}(A_1)}. \tag{A.16}$$

To characterize how $\mu^{\alpha\beta}$ is related to μ^{β} and μ^{α}, we first show $\mu^{\alpha\beta}(P) > 0$. By way of contradiction, assume $\mu^{\alpha\beta}(P) = 0$. Clearly, P cannot be B_1 or B_2: If so, then (A.12) implies that $\mu^{\beta}(P) = 0$, a contradiction. If $P = S\backslash(B_1 \cup B_2)$, then by condition (iib), it must be that $\mu^{\alpha}(P) = 0$ (recall that $\mu^{\alpha}(B_i) = \mu(B_i)$ for $i = 1, 2$). Then $\mu^{\alpha}(A_1 \cup A_2) \leq \mu^{\alpha}(P) = 0$, a contradiction.

Since $\mu^{\alpha\beta}(P)$ and $\mu^{\alpha}(P)$ are both positive, by condition (iia), we have

$$\frac{\mu^{\alpha\beta}(A_2)}{\mu^{\alpha\beta}(A_1)} = \frac{\mu^{\alpha}(A_2)}{\mu^{\alpha}(A_1)}. \tag{A.17}$$

Furthermore,

$$\mu^{\alpha\beta}(A_1 \cup A_2) = \frac{\mu^{\alpha}(A_1 \cup A_2)}{\mu^{\alpha}(P)}\mu^{\alpha\beta}(P) = \frac{\mu(A_1 \cup A_2)}{\mu^{\alpha}(P)}\mu^{\alpha\beta}(P).$$

Note that

$$\mu^{\alpha}(P) = \mu^{\alpha}(A_1 \cup A_2) + \mu^{\alpha}\big(P\backslash(A_1 \cup A_2)\big) = \mu(A_1 \cup A_2) + \mu\big(P\backslash(A_1 \cup A_2)\big) = \mu(P).$$

In addition,

$$\mu^{\alpha\beta}(P) = \mu^{\beta\alpha}(P) = \mu^{\beta}(P),$$

in which the first equality is due to the fact that $\mu^{\alpha\beta}(B_i) = \mu^{\beta\alpha}(B_i)$ for $i = 1, 2$, and the second equality is because $\mu^{\beta\alpha}(A_1 \cup A_2) = \mu^{\beta}(A_1 \cup A_2)$ and $\mu^{\beta\alpha}(E) = \mu^{\beta}(E)$ for any $E \subseteq S\backslash(A_1 \cup A_2)$. Thus,

$$\mu^{\alpha\beta}(A_1 \cup A_2) = \frac{\mu(A_1 \cup A_2)}{\mu(P)}\mu^{\beta}(P) = \mu^{\beta}(A_1 \cup A_2). \tag{A.18}$$

Combining (A.15), (A.16), (A.17), and (A.18) yields the desired result: $\mu^{\alpha\beta}(A_1) = \mu^{\beta\alpha}(A_1)$ and $\mu^{\alpha\beta}(A_2) = \mu^{\beta\alpha}(A_2)$.

Finally, let Σ' be the smallest σ-algebra that contains A_1, A_2, B_1, B_2. It is clear that by conditions (iia) and (iii), each element of Σ' is updated in proportion to the prior as the DM receives α or β. Thus, since $\mu^{\alpha\beta}(B_i) = \mu^{\beta\alpha}(B_i)$ and $\mu^{\alpha\beta}(A_i) = \mu^{\beta\alpha}(A_i)$ for $i = 1, 2$, we have $\mu^{\alpha\beta} = \mu^{\beta\alpha}$. □

Appendix B: Proof of Theorem 2

Proof. By Theorem 1, it is clear that (i) and (ii) are equivalent. Hence, it suffices to prove that optimization (P) in Theorem 2 has the conservative rule as the unique solution. Recall optimization (P):

$$\min_{\nu \ll \mu} -\int_S \ln\left(\frac{d\nu}{d\mu}\right) d\mu$$
$$\text{subject to} \quad \nu(A_1) \geq \nu(A_2).$$

First of all, note that since $\nu \ll \mu$, the Radon–Nikodym derivative is well-defined. Thus, the constraint optimization problem is equivalent to the following problem:

$$\max_{f:S\to[0,\infty) \text{ measurable}} \int_S \ln f \, d\mu \tag{P*}$$
$$\text{subject to} \int_{A_1} f \, d\mu \geq \int_{A_2} f \, d\mu,$$

$$\int_S f \, d\mu = 1.$$

The next lemma reduces (P*) to a finite dimensional problem.

LEMMA 6. *Let $\mu(C) > 0$ and $p \in [0, 1]$. The following optimization problem has solution $f^* := p/\mu(C)$ and the solution is unique up to μ-almost everywhere equality:*

$$\max_{f:C\to[0,\infty)\ \textit{measurable}} \int_C \ln f \, d\mu$$
$$\textit{subject to} \quad \int_C f \, d\mu = p.$$

PROOF. Let ν be such that $\nu(A) = \mu(A \cap C)/\mu(C)$ for all $A \in \Sigma$. For any measurable f,

$$\int_C \ln f \, d\mu = \mu(C)\mathbb{E}_\nu[\ln f] \le \mu(C) \ln \mathbb{E}_\nu[f] = \mu(C) \ln \frac{p}{\mu(C)} = \int_C \ln f^* \, d\mu$$

by Jensen's inequality. Equality is attained if and only if f is a constant μ-almost everywhere. Then the constraint demands that $f = p/\mu(C)\mu$-almost everywhere. □

With Lemma 6, (P*) is reduced to a finite-dimensional optimization problem. Let $\mu(A_1 \backslash A_2) = p_1$, $\mu(A_2 \backslash A_1) = p_2$, and $p_3 = 1 - p_1 - p_2$. Consider the following optimization problem:

$$\max_{q_i \ge 0} \sum_{i=1}^{3} p_i \ln \frac{q_i}{p_i} \tag{P**}$$
$$\text{subject to} \quad q_1 \ge q_2$$
$$q_i = 0 \quad \text{if } p_i = 0$$
$$q_1 + q_2 + q_3 = 1.$$

Suppose $p_1 > 0$ and ignore the absolute continuity constraint for the time being. Since our objective function is strictly concave, Kuhn–Tucker conditions are necessary and sufficient. It is easy to verify that the Kuhn–Tucker conditions imply $q_3 = p_3$. Moreover, if $p_2 > p_1$, then $q_1 = q_2 = (p_1 + p_2)/2$; if $p_1 \ge p_2$, then $q_i = p_i$ for all i. Clearly, in all circumstances the absolute continuity constraint is not violated.

Suppose $p_1 = 0$. By absolute continuity $q_1 = 0$ and then the inequality constraint demands $q_2 = 0$, which implies that $q_3 = 1$.

Let the solution to (P**) be (q_1^*, q_2^*, q_3^*). Let $h : [0, 1] \to \mathbb{R}$ be such that $h(0) = 0$ and $h(x) = 1/x$ if $x > 0$. By Lemma 6 and the analysis of (P**), (P*) has a solution f that is unique up to μ-everywhere equality, given by

$$f(s) = \begin{cases} q_1^* \cdot h(p_1), & \text{if } s \in A_1, \\ q_2^* \cdot h(p_2), & \text{if } s \in A_2, \\ q_3^* \cdot h(p_3), & \text{otherwise.} \end{cases}$$

Since $d(\mu\|\nu)$ is strictly convex in ν, the solution to (P) (and thus (P*)) is unique. It is easy to verify that the solution above characterizes exactly the conservative rule. □

Appendix C: Proof of Proposition 1

Proof. The "only if" direction of the first statement is trivially implied by the absolute continuity property of pseudo-Bayesian updating implied by conditions (i), (iia), and (iii) of Theorem 1. I now show the "if" direction. Consider a pseudo-Bayesian updating rule that is not the radical rule. Then by definition, there exists $(\mu, \alpha) \in \mathbb{I}(S, \Sigma)$ with $A_1 \cap A_2 = \emptyset$ such that $\mu(A_1) < \mu(A_2)$ and $\mu^\alpha(A_2) > 0$. By Theorem 1, we know that it must be the case that $\mu(A_1) > 0$. Let $B_2 = A_1$ and $B_1 \subseteq A_2$ with $\mu(B_1) = \mu(A_2)/2$. Such B_1 exists, since μ is nonatomic. By Theorem 1, we have $\mu^\alpha(B_1) < \mu^\alpha(A_2) \leq \mu^\alpha(A_1) = \mu^\alpha(B_2)$. Then it follows that $\mu^{\alpha\beta}(A_1) = \mu^{\alpha\beta}(B_2) \leq \mu^{\alpha\beta}(B_1) < \mu^{\alpha\beta}(A_2)$, in which the last inequality follows from the fact that $A_2 \backslash B_1 \subseteq S \backslash (B_1 \cup B_2)$, and thus $\mu^{\alpha\beta}(A_2 \backslash B_1) = \mu^\alpha(A_2 \backslash B_1) > 0$.

Next, I prove the second statement. Consider a never-radical pseudo-Bayesian updating rule. For any μ, pick standard statements α, β such that $0 < \mu(A_1) < \mu(A_2)$, $B_2 = A_1$, $B_1 \subseteq A_2$, and $\mu(B_1) = \mu(A_2)/2$. A decision maker equipped with a never-radical pseudo-Bayesian updating rule creates new null events only upon encountering statements with a degenerate reduction. Therefore, as she learns α and β alternately, both statements remain nondegenerate along the path of her beliefs. Hence, $A_2 \backslash B_1$ will remain nonnull along the path. It follows that whenever the DM learns α so that A_1 is at least as likely as A_2, $B_2 = A_1$ will be strictly more likely than B_1; whenever the DM incoporates β so that B_1 is at least as likely as B_2, A_2 will be strictly more likely than $B_2 = A_1$. Thus, (α, β) induces persistent recency bias on μ. □

Appendix D: Proof of Theorem 3

Proof. Let $\alpha_i = (A_{i1}, A_{i2})$ for all i. WLOG assume that each α_i is standard. Let Π be the coarsest common refinement of $\Pi_{\alpha_1}, \ldots, \Pi_{\alpha_n}$. Let ν be a μ-solution to $\{\alpha_i\}_{i=1}^n$ such that $\nu(C \cap P)/\nu(P) = \mu(C \cap P)/\mu(P)$ for any $C \in \Sigma$ and $P \in \Pi$ such that $\nu(P) > 0$. Since $\{\alpha_i\}_{i=1}^n$ is μ-compatible, such ν always exists.

Without loss of generality, assume that α_i is not μ-degenerate for any i. If α_i is μ-degenerate, since the posterior has to be absolutely continuous with respect to the prior, α_i will remain degenerate before it is learned the first time, and once α_i is learned the first time, A_{i1}, A_{i2} will remain null ever after. Since ν is a solution to $\{\alpha_i\}_{i=1}^n$, it must be the case that $\nu(A_{i1}) = \nu(A_{i2}) = 0$. Hence, it suffices to show Theorem 3 for statements that are not μ-degenerate.

Now, for each k, let $\mu_k = \mu^{\alpha_{i_1}\alpha_{i_2}\cdots\alpha_{i_k}}$. First, we show that

$$d(\mu_{k+1}\|\mu_k) \cdot \min_i \nu\big(D_{\alpha_i}^{\mu_k}\big) \leq d(\nu\|\mu_k) - d(\nu\|\mu_{k+1}). \tag{D.1}$$

Since the conservative rule is never-radical, it is clear that $\nu \ll \mu \ll \mu_k \ll \mu_{k+1}$ for all k. Therefore, the KL-divergences in (D.1) are all finite. Let $\alpha_{i_{k+1}} = \alpha$, $\mu_k(A_1) = p_1$, $\mu_k(A_2) =$

p_2, $\nu(A_1) = q_1$, and $\nu(A_2) = q_2$. If $p_1 \geq p_2$ there is nothing to prove, since $\mu_k = \mu_{k+1}$. Hereafter, we assume $p_1 < p_2$. With the conservative rule, inequality (D.1) reads

$$\left(\frac{p_1+p_2}{2}\ln\frac{2p_1}{p_1+p_2}+\frac{p_1+p_2}{2}\ln\frac{2p_2}{p_1+p_2}\right)\cdot\min_i \nu\big(D_{\alpha_i}^{\mu_k}\big)$$
$$\geq q_1\ln\frac{2p_1}{p_1+p_2}+q_2\ln\frac{2p_2}{p_1+p_2}.$$

Since $0 \leq (p_1+p_2)\cdot\min_i \nu(D_{\alpha_i}^{\mu_k}) \leq \nu(D_{\alpha}^{\mu_k}) = q_1+q_2$ and $d(\mu_{k+1}\|\mu_k) \geq 0$, to show (D.1) it suffices to prove that

$$\frac{q_1+q_2}{2}\ln\frac{2p_1}{p_1+p_2}+\frac{q_1+q_2}{2}\ln\frac{2p_2}{p_1+p_2}\geq q_1\ln\frac{2p_1}{p_1+p_2}+q_2\ln\frac{2p_2}{p_1+p_2},$$

which reads

$$\frac{q_1-q_2}{2}\ln\frac{p_2}{p_1}\geq 0,$$

which clearly holds, since $p_1 < p_2$ and $q_1 \geq q_2$.

Thus, by inequality (D.1), $0 \leq d(\nu\|\mu_{k+1}) \leq d(\nu\|\mu_k)$, and thus $\lim_{k\to\infty} d(\nu\|\mu_k)$ exists and $\lim_{k\to\infty} d(\mu_{k+1}\|\mu_k) = 0$.

Let d_{TV} denote the total variation distance, that is,

$$d_{\text{TV}}\big(\rho,\rho'\big)=\sup_{A\in\Sigma}\big|\rho(A)-\rho'(A)\big|$$

for any $\rho, \rho' \in \Delta(S,\Sigma)$. By Pinsker's inequality,

$$\sqrt{\frac{d(\mu_{k+1}\|\mu_k)}{2}}\geq d_{\text{TV}}(\mu_{k+1},\mu_k).$$

See Tsybakov (2009), page 88, for a formal proof of Pinsker's inequalilty.

Therefore, by $\lim_{k\to\infty} d(\mu_{k+1}\|\mu_k) = 0$, $d_{\text{TV}}(\mu_{k+1},\mu_k)$ also converges to 0. Therefore, $\max_{1\leq j,l\leq N} d_{\text{TV}}(\mu_{kN+j},\mu_{kN+l})$ converges to 0 as $k\to\infty$, in which N is the constant given in the definition of comprehensiveness. Since all μ_k's share the same Bayesian conditional distribution given each element in Π, it is easy to see that $\{\mu_k\}$ has a least one limit point, denoted as μ^*, under d_{TV}. Since the learning sequence is comprehensive, μ^* of $\{\mu_k\}$ must be consistent with α_j for all j. Note that since $\mu_k \ll \mu$, it must be the case that $\mu^* \ll \mu$, so μ^* is indeed a μ-solution to $\{\alpha_i\}_{i=1}^n$. Moreover, it is clear that $\mu^*(C\cap P)/\mu^*(P) = \mu(C\cap P)/\mu(P)$ for any $C\in\Sigma$ and $P\in\Pi$ such that $\mu^*(P) > 0$.

Setting $\nu = \mu^*$, we have that $\lim_{k\to\infty} d(\mu^*\|\mu_k)$ exists. Take a subsequence $\{\mu_{k_j}\}$ that converges to μ^*. By continuity,

$$\lim_{k\to\infty} d\big(\mu^*\|\mu_k\big)=\lim_{j\to\infty} d\big(\mu^*\|\mu_{k_j}\big)=d\big(\mu^*\|\mu^*\big)=0.$$

Then Pinsker's inequality finishes our proof. □

References

Ahn, D. S. (2008), "Ambiguity without a state space." *The Review of Economic Studies*, 75, 3–28. [266]

Alchourrón, C. E., P. Gärdenfors, and D. Makinson (1985), "On the logic of theory change: Partial meet contraction and revision functions." *The Journal of Symbolic Logic*, 50, 510–530. [263, 266]

Basu, P. (2019), "Bayesian updating rules and AGM belief revision." *Journal of Economic Theory*, 179, 455–475. [266]

Berk, R. H. (1966), "Limiting behavior of posterior distributions when the model is incorrect." *The Annals of Mathematical Statistics*, 37, 51–58. [265]

Berk, R. H. (1970), "Consistency a posteriori." *The Annals of Mathematical Statistics*, 41, 894–906. [265]

Bonanno, G. (2009), "Rational choice and AGM belief revision." *Artificial Intelligence*, 173, 1194–1203. [266]

Bregman, L. M. (1967), "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming." *USSR Computational Mathematics and Mathematical Physics*, 7, 200–217. [269]

Bunke, O. and X. Milhaud (1998), "Asymptotic behavior of Bayes estimates under possibly incorrect models." *The Annals of Statistics*, 26, 617–644. [265]

Caticha, A. (2004), "Relative entropy and inductive inference." In *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, volume 707 (G. Erickson and Y. Zhai, eds.), 75–96. [265, 266]

Caticha, A. and A. Giffin (2006), "Updating probabilities." In *Bayesian Inference and Maximum Entropy Methods in Science and Engineering* (A. Mohammad-Djafari, ed.), volume 872 of American Institute of Physics Conference Series, 31–42. [266]

Chambers, C. P. and T. Hayashi (2010), "Bayesian consistent belief selection." *Journal of Economic Theory*, 145, 432–439. [266]

Costa, H. and A. P. Pedersen (2011), "Belief revision." In *The Continuum Companion to Philosophical Logic*, 450–502. [266]

Damiano, E. (2006), "Choice under limited uncertainty." *The B.E. Journal of Theoretical Economics*, 6, 1–35. [266]

de Finetti, B. (1937), "La prévision: ses lois logiques, ses sources subjectives." *Annales de l'Institut Henri Poincaré*, 7, 1–68. [254]

DeMarzo, P., D. Vayanos, and J. Zwiebel (2003), "Persuasion bias, social infuence, and unidimensional opinions." *Quarterly Journal of Economics*, 118, 909–968. [268]

Epstein, L. G. (2006), "An axiomatic model of non-Bayesian updating." *The Review of Economic Studies*, 73, 413–436. [266]

Epstein, L. G., J. Noor, and A. Sandroni (2008), "Non-Bayesian updating: A theoretical framework." *Theoretical Economics*, 3, 193–229. [266]

Fermé, E. and S. O. Hansson (2011), "AGM 25 years: Twenty-five years of research in belief change." *Journal of Philosophical Logic*, 40, 295–331. [266]

Fishburn, P. C. (1986), "The axioms of subjective probability." *Statistical Science*, 1, 335–345. [254]

Gajdos, T., T. Hayashi, J.-M. Tallon, and J.-C. Vergnaud (2008), "Attitude toward imprecise information." *Journal of Economic Theory*, 140, 27–65. [266]

Gärdenfors, P. (1984), "Epistemic importance and minimal changes of belief." *Australasian Journal of Philosophy*, 62, 136–157. [263]

Gennaioli, N. and A. Shleifer (2010), "What comes to mind." *Quarterly Journal of Economics*, 125, 1399–1433. [266]

Jaynes, E. T. (1957), "Information theory and statistical mechanics." *The Physical Review*, 106, 620–630. [266]

Kullback, S. (1959), *Information Theory and Statistics*. Wiley, New York. [265]

Kullback, S. and R. A. Leibler (1951), "On information and sufficiency." *The Annals of Mathematical Statistics*, 22, 79–86. [265]

Majumdar, D. (2004), "An axiomatic characterization of Bayes' rule." *Mathematical Social Sciences*, 47, 261–273. [259]

Mullainathan, S., J. Schwartzstein, and A. Shleifer (2008), "Coarse thinking and persuasion." *The Quarterly Journal of Economics*, 123, 577–619. [266]

Ortoleva, P. (2012), "Modeling the change of paradigm: Non-Bayesian reactions to unexpected news." *American Economic Review*, 102, 2410–2436. [266]

Perea, A. (2009), "A model of minimal probabilistic belief revision." *Theory and Decision*, 67, 163–222. [259]

Rabin, M. (2002), "Inference by believers in the law of small numbers." *The Quarterly Journal of Economics*, 117, 775–816. [266]

Rabin, M. and J. L. Schrag (1999), "First impressions matter: A model of confirmatory bias." *The Quarterly Journal of Economics*, 114, 37–82. [266]

Rott, H. (2000), "Two dogmas of belief revision." *The Journal of Philosophy*, 97, 503–522. [263]

Shore, J. E. and R. W. Johnson (1980), "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy." *IEEE Transactions on Information Theory*, IT-26, 26–27. [266]

Skilling, J. (1988), "The axioms of maximum entropy." *Maximum-Entropy and Bayesian Methods in Science and Engineering*, 1, 173–187. [266]

Spiegler, R. (2017), ““Data monkeys”: A procedural model of extrapolation from partial statistics.” *The Review of Economic Studies*, 84, 1818–1841. [266]

Spiegler, R. (2020), “Modeling players with random “data access”.” Working Paper. [266]

Tsybakov, A. (2009), *Introduction to Nonparametric Estimation*. Springer. [286]

Villegas, C. (1964), “On qualitative probability σ-algebras.” *Ann. Math. Statist.*, 35, 1787–1796. [254, 266]

Zhao, C. (2020), “Representativeness and similarity.” Working Paper. [266]

Co-editor Ran Spiegler handled this manuscript.

Manuscript received 1 October, 2020; final version accepted 17 April, 2021; available online 20 April, 2021.