Coarse Bayesianism

Alexander M. Jakobsen[*]

February 28, 2019

Preliminary and incomplete.
For the latest version, please visit `www.amjakobsen.com`

Abstract

I introduce a model of belief updating, *Coarse Bayesian Updating*, where a boundedly rational agent, upon receipt of new information, applies a set of subjective criteria to select among competing theories of the world. The agent is characterized by a prior, a partition of the space of all probability distributions, and a representative distribution for each cell of the partition. When new information arrives, the agent computes the Bayesian posterior, determines which cell of the partition it belongs to, and adopts the representative of that cell as his posterior belief. The model includes Bayesian updating as a special case and accommodates many documented violations of Bayesian updating, including both under- and over-reaction to information. I provide behavioral characterizations of this procedure and analyze how it relates to other models and evidence on non-Bayesian updating. I also characterize what it means for an agent to be more sophisticated, and how Coarse Bayesians value information. The model employs standard primitives and, therefore, can be applied in most economic environments.

[*]Department of Economics, University of Calgary; alexander.jakobsen@ucalgary.ca.

1 Introduction

Bayesian updating occupies a central role in economic theory. A wide body of evidence, however, suggests that individual behavior cannot be reconciled with Bayesian updating in a variety of settings. While observed violations of Bayes' rule are rich and varied, a number of patterns typically emerge. For example, individuals often display *conservatism bias*: they under-react to new evidence, possibly ignoring it altogether. In other cases, individuals *over-react* to new information: they falsely extrapolate or, more generally, discern patterns in data that are not actually present. Combinations of these forces can lead to *misattribution*, where individuals under-weight one explanation for the evidence while over-weighting another. In this paper, I introduce and analyze a generalization of standard Bayesian updating—*Coarse Bayesian Updating*—accommodating these (and other) behavioral tendencies.

In my model, an agent considers only some subset of feasible probability distributions over a state space, one of which is his prior. The feasible distributions can be interpreted as competing *theories of the world*. When new evidence arrives (in the form of a noisy signal s), the agent applies a set of subjective criteria for selecting among the competing theories. In particular, the agent is characterized by a partition of the set of all probability distributions, along with a representative distribution for each cell of the partition. After observing s, the agent determines which cell contains the correct Bayesian posterior and adopts the representative of that cell as his new belief. Thus, the representative distributions form the set of competing theories, and the partition captures his criteria, or *standard of proof*, for selecting among them. See Figure 1 below.

I provide two characterization of Coarse Bayesian behavior. The first is a direct characterization, taking signal-contingent beliefs as primitive; I refer to the mapping from signals to beliefs as an *updating rule*. Three axioms are required: Homogeneity, Convexity, and Confirmation. Homogeneity states that beliefs are invariant to scalar transformations of signals. Convexity states that if two signals result in the same belief, then so do convex combinations (randomizations) of those signals. A natural interpretation of Convexity is that the agent understands his own updating rule: if he is unsure as to whether he has observed s or t, but recognizes that s and t would yield the same belief $\hat{\mu}$ according to his updating rule, then he should adopt $\hat{\mu}$. This also implies that the cells in the Coarse Bayesian representation are convex. Finally, Confirmation states that if the Bayesian posterior for some signal t coincides with some feasible belief, then the updating rule associates that belief to signal t. In other words, if the evidence directly confirms a given theory, then that theory should be taken as the new belief.

The second characterization takes as primitive a family of signal-contingent choices from

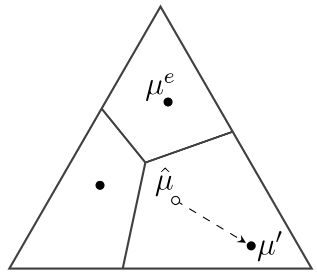

Figure 1: Coarse Bayesian Updating. In this example, the agent entertains three distributions (solid dots). The point μ^e is the prior. After observing a signal s, the agent computes the Bayesian posterior $\hat{\mu}$, determines which cell of the partition contains $\hat{\mu}$, and then adopts the representative of that cell (in this case, μ') as his new belief.

menus of risky actions. In contrast to the first characterization, this also requires the individual to be an expected utility maximizer. My approach is analogous to that of Anscombe and Aumann (1963), with additional axioms translating the Convexity and Confirmation axioms of the first characterization into revealed-preference statements in the second. Both characterizations establish uniqueness of the agent's partition and feasible beliefs. In contrast to Savage (1972) and Anscombe and Aumann (1963), the behavioral characterization establishes not only a subjective prior over the state space, but also subjective criteria (the partition) for updating beliefs. The behavioral setting can also be used to characterize notions of sophistication: a "more sophisticated" agent (ie, one with a finer partition) is more responsive to changes in signals, while a "more Bayesian" agent (ie, one with a larger set of competing theories) behaves like a standard Bayesian for a wider range of signals.

In section 3, I discuss some of the main features of Coarse Bayesian updating, as well as some closely related studies of non-Bayesian updating. I show that Coarse Bayesian updating can be equivalently expressed as a theory of signal distortion, and that it typically implies path dependence (sensitivity to the order in which signals are generated). I also compare and contrast Coarse Bayesianism to an alternative updating model—*Maximum-Likelihood* updating—where an agent holds second-order beliefs over a state space and applies a maximum-likelihood selection criterion after observing signals. I show that standard Bayesian updating is a special case of both Coarse Bayesian and Maximum-Likelihood updating, but that (in general) neither model subsumes the other.

Finally, section 5 analyzes how Coarse Bayesians value information (Blackwell experiments) when faced with menus of risky actions. I show that a Coarse Bayesian's ranking of information structures typically exhibits violations of the Blackwell (1951, 1953) information ordering, and that a Coarse Bayesian adheres to the Blackwell ordering if and only if the menu and his partition satisfy a particular co-measurability condition. The Blackwell ordering can also characterize a notion of sophistication: the agent has a finer partition if and only if his value of information is more responsive to a particular class of *coarse Blackwell garblings*. In fact, the connection runs deeper: both the partition and the set of feasible beliefs can be identified from the agent's ranking of information structures—even if one restricts attention to pairs of experiments that are Blackwell comparable.

1.1 Related Literature

To be written.

2 Model

I consider a single agent who updates beliefs after observing noisy signals. Let Ω denote a finite set of N states and Δ the set of probability distributions over Ω. A distribution $\mu \in \Delta$ assigns probability μ_ω to states $\omega \in \Omega$.

Following Jakobsen (2016), a **signal** is a profile $s = (s_\omega)_{\omega\in\Omega} \in [0,1]^\Omega$ such that $s_\omega \neq 0$ for at least one state ω. Let S denote the set of all signals. Intuitively, a signal represents a message that can be generated, and the entries of s_ω are the likelihoods of the message being generated in different states of the world. As explained in section 5, general information structures (Blackwell experiments) can be represented as collections of signals. I reserve e to denote the **uninformative signal**; that is, $e \in S$ and $e_\omega = 1$ for all $\omega \in \Omega$.

For two profiles $x = (x_\omega)_{\omega\in\Omega}$ and $y = (y_\omega)_{\omega\in\Omega}$ of real numbers, let $xy := (x_\omega y_\omega)_{\omega\in\Omega}$ denote the profile formed by multiplying x and y component-wise. The dot product of x and y is given by $x \cdot y := \sum_{\omega\in\Omega} x_\omega y_\omega$. The notation $x \approx y$ indicates that $x = \lambda y$ for some $\lambda > 0$. Clearly, if $\mu, \mu' \in \Delta$ and $\mu \approx \mu'$, then $\mu = \mu'$.

If $\mu \in \Delta$ and $s \in S$ such that $\mu s \neq 0$, then $B(\mu|s)$ denotes the Bayesian posterior of μ for signal s; that is, the unique $\mu' \in \Delta$ such that $\mu' \approx s\mu$.

Finally, an **updating rule** is a function $\mu : S \to \Delta$ assigning probability distributions $\mu^s \in \Delta$ to signals $s \in S$. For each $s \in S$, μ^s is the agent's posterior belief conditional on observing signal s. I assume μ^e, the **prior**, has full support. Updating rules will often be written as profiles: $\mu = (\mu^s)_{s\in S}$.

2.1 Coarse Bayesian Representations

The primary goal of this paper is to introduce and analyze a model of (typically) non-Bayesian updating, *Coarse Bayesian* updating. Coarse Bayesian updating rules are characterized by the following three axioms on μ.

Homogeneity. If $s \approx t$, then $\mu^s = \mu^t$.

Homogeneity requires that the agent's analysis of a signal only depends on the likelihood ratios $s_\omega / s_{\omega'}$. This is a key feature of standard Bayesian updating: $B(\mu^e|s)$ coincides with $B(\mu^e|\lambda s)$, provided $\lambda > 0$ and $\lambda s \in S$.

Convexity. If $\mu^s = \mu^t$ and $\lambda \in [0,1]$, then $\mu^{\lambda s+(1-\lambda)t} = \mu^s$.

Convexity asserts that if signals s and t result in the same (potentially non-Bayesian) posterior $\hat{\mu}$, then so does $r := \lambda s + (1-\lambda)t$. An interpretation of Convexity is that it describes the behavior of an agent who understands his own updating rule. To see this, note that r represents a garbled signal: observing r indicates that either s was generated (with probability λ) or t was generated (with probability $1-\lambda$). Therefore, an agent who recognizes that both s and t yield $\hat{\mu}$—in other words, an agent who understands his own updating procedure—ought to conclude that r also yields posterior $\hat{\mu}$. In this sense, Convexity is a "sure-thing" principle for updating, requiring the agent to be cognizant of his own behavior.

Confirmation. If $t\mu^e \approx \mu^s$, then $\mu^t = \mu^s$.

Confirmation requires that if the Bayesian posterior at t coincides with some feasible belief μ^s, then the updating rule satisfies Bayes' rule at signal t: $\mu^t = \mu^s$. In other words, if signal t exactly confirms a feasible belief, then that belief is the posterior at t.

Proposition 1. *An updating rule μ is homogeneous, convex, and confirmatory if and only if there is a partition $\mathcal{P}$ of $\Delta\Omega$ and a profile $(\mu^P)_{P\in\mathcal{P}}$ of distributions $\mu^P \in \Delta$ such that*

(i) each cell $P \in \mathcal{P}$ is convex,

(ii) for all $P \in \mathcal{P}$, $\mu^P \in P$, and

(iii) for all $s \in S$, $B(\mu^e|s) \in P$ implies $\mu^s = \mu^P$.

*The pair $\langle \mathcal{P}, (\mu^P)_{P\in\mathcal{P}} \rangle$ is a **Coarse Bayesian Representation** of μ. If $\langle \mathcal{Q}, (\mu^Q)_{Q\in\mathcal{Q}} \rangle$ is another Coarse Bayesian Representation of μ, then $\mathcal{P} = \mathcal{Q}$ and $(\mu^P)_{P\in\mathcal{P}} = (\mu^Q)_{Q\in\mathcal{Q}}$.*

In a Coarse Bayesian representation, the agent is characterized by a partition, $\mathcal{P}$, of Δ as well as a representative distribution $\mu^P \in P$ for each cell P of the partition. Each cell is convex, and one of the distributions is the prior: $\mu^P = \mu^e$. When a signal s arrives, the agent selects the unique μ^P such that the Bayesian posterior $B(\mu^e|s)$ belongs to cell P. The updating rule μ (in particular, the set S) is sufficiently rich to uniquely identify both $\mathcal{P}$ and μ^P for all $P \in \mathcal{P}$.

There are (at least) two interpretations of Coarse Bayesian behavior. In each interpretation, the agent entertains only some restricted set of distributions as possible theories of the world, and the updating rule is essentially a minimal deviation from standard Bayesian updating subject to this restriction. The interpretations, however, differ in terms of whether the agent understands signals and Bayesian updating. In the first interpretation, the agent fully understands the informational content of signals: he computes the Bayesian posterior, but applies his own "standard of proof" for selecting among the competing theories. If the Bayesian posterior exactly coincides with one of the candidate distributions, then he adopts that distribution as his posterior. Hence, the standard of proof—encapsulated by $\mathcal{P}$—is a subjective characteristic of the individual. In the second interpretation, the agent does not fully grasp the informational content of signals. Rather, he only infers that his posterior belief ought to belong to some particular region (the cell containing the Bayesian posterior). Therefore, $\mathcal{P}$ represents the agent's ability to process signals. Effectively, each cell represents some combination of properties that a distribution may have, and the agent correctly processes information only to the extent that it allows him to determine which properties must be satisfied. For each feasible combination of properties, he has in mind a representative theory of the world, which he adopts if he determines that those properties are satisfied.

The proof of Proposition 1 is quite simple, so I only provide the following sketch. Necessity of the axioms is clear. For sufficiency, observe that Homogeneity and Convexity imply the existence of a partition of S (into convex cones) such that $\mu^s = \mu^t$ if and only if s and t belong to the same cell of the partition. Since the cells are convex cones, each cell corresponds to a convex set P of points in Δ (in particular, signal s is associated with the unique $\mu' \in \Delta$ such that $\mu' \approx s\mu^e$). The Confirmation property ensures that if $B(\mu^e|s) \in P$, then $\mu^s \in P$.

3 Models and Evidence

Coarse Bayesian updating is related to a number of other models and concepts of non-Bayesian updating, and accommodates a variety of experimental findings. In this section, I examine these relationships.

3.1 Signal Distortions

In addition to the two main interpretations offered in the previous section, Coarse Bayesian updating can also be interpreted as a model of signal distortion. Formally, a **signal distortion** is a map $d : S \to S$ such that $d(e) = e$. The interpretation is that when signal s is generated, the agent behaves as if $d(s)$ had been generated instead. Hence, a signal distortion captures errors or biases in the agent's perception of information.

Definition 1. An updating rule μ has a **Signal Distortion Representation** if there exists a signal distortion d such that $\mu^s = B(\mu^e|d(s))$ for all $s \in S$.

In a signal distortion representation, an agent who observes signal s misperceives the signal as $d(s)$, and then applies standard Bayesian updating to the distorted signal. Thus, in such representations, the prior μ^e and the distortion function d are the essential behavioral parameters. Naturally, some restrictions on d are required.

Definition 2. A signal distortion d is:

(i) *Convex* if $d(s) \approx d(t)$ implies $d(\lambda s + (1-\lambda)t) \approx d(s)$ for all $\lambda \in [0,1]$, and

(ii) *Idempotent* if $d(d(s)) = d(s)$ for all s.

Convexity requires mixtures of signals to result in the same distortion if the constituent signals have a common distortion (up to scalar transformation). The interpretation is similar to that of convexity for updating rules: if the agent is uncertain of which signal was generated (s or t), but if s and t result in the same distortion, then the mixed signal must also yield that distortion.

Idempotency requires distorted signals to be stable: the distortion of $d(s)$ is $d(s)$. Thus, idempotent distortions effectively categorize signals and assign the same distortion to signals in the same category.

Proposition 2. *An updating rule has a Coarse Bayesian Representation if and only if it has a convex, idempotent Signal Distortion Representation. If d and d' are Signal Distortion Representations for a given updating rule, then $d(s) \approx d'(s)$ for all $s \in S$.*

Proposition 2 establishes an equivalence between Coarse Bayesian and Signal Distortion Representations. Therefore, any updating rule μ satisfying Homogeneity, Convexity, and Confirmation has a signal distortion representation, and the distortion d is unique up to scalar transformation.

There is an important distinction, however, between Coarse Bayesian and Signal Distortion Representations. Suppose the agent observes a sequence of signals (rather than a single s) and applies his updating rule upon each realization (where the posterior resulting from the previous realization becomes the prior for the next round of updating). In this case, the distortion d would have to change upon each realization in order to coincide with the Coarse

Bayesian procedure (effectively, d would have to be scaled by a signal that compensates for the new prior), and an appropriate transformation may not exist if the new prior lacks full support. Thus, although there is a clear connection between the two concepts for a single round of updating, there is a sense in which the Coarse Bayesian representation is more readily adaptable to dynamic settings.

3.2 Maximum Likelihood

Since Coarse Bayesians effectively select among competing models using subjective criteria, it is natural to wonder if Coarse Bayesian updating results from subjective second-order beliefs (that is, priors over priors). In this section, I consider a particular class of such updating rules.

Definition 3. A homogeneous, convex updating rule μ has a **Maximum-Likelihood Representation** if there exists a probability distribution Γ over Δ (with density γ) such that

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

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

In a Maximum Likelihood (ML) representation, the agent has a second-order prior Γ that he updates (using Bayes' rule) upon arrival of signal s. Then, he selects the prior with the highest posterior probability. As is easily verified, this rule selects among beliefs $\hat{\mu}$ that maximize the likelihood function. Notice that L is homogeneous (of degree 0) and convex in s; the restriction to homogeneous convex updating rules, therefore, only takes effect when there are ties (multiple candidate beliefs that maximize L).

Maximum-Likelihood (and, more generally, Coarse Bayesian) updating is reminiscent of the Hypothesis-Testing model of Ortoleva (2012). In the Hypothesis-Testing model, the agent applies Bayesian updating only for signals that have sufficiently high prior probability; for "unexpected" signals, the agent applies a Maximum-Likelihood procedure. Behavior in the Hypothesis-Testing model can be well-approximated by Coarse Bayesian behavior by choosing a partition where each cell is a singleton unless the cell is near the extremes of the probability simplex. This way, the agent applies Bayesian updating for signals that do not take the Bayesian posterior too far away from the prior. At more "extreme" signals, the agent may respond in a non-Bayesian fashion, as in the Hypothesis-Testing model.

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

Example 2. Not every Coarse Bayesian rule can be expressed as a Maximum-Likelihood rule. Suppose $|\Omega| = 3$ and consider a Coarse Bayesian representation where $\mathcal{P}$ has two cells, P and P', with $\mu^P = \mu^e$ and $\mu^{P'} = \mu' \neq \mu^e$. The boundary between P and P' corresponds to a hyperplane, H, in S. We will choose H (hence, $\mathcal{P}$) in such a way that no distribution γ on Δ (with support $\{\mu^e, \mu'\}$) can generate the same updating behavior as $\langle \mathcal{P}, (\mu^P)_{P\in\mathcal{P}} \rangle$ under the Maximum-Likelihood procedure.

Observe that if γ generates the same updating behavior, then $L(\mu^e|s) = L(\mu'|s)$ for all $s \in H$. In particular, $[\gamma(\mu^e)\mu^e - \gamma(\mu')\mu'] \cdot s = 0$ for all $s \in H$. Thus, the line $\{\lambda[\mu^e + \mu'] - \mu' : \lambda \geq 0\}$ is orthogonal to the hyperplane H. Since $\mu^e \neq \mu'$, we may assume H strictly separates μ^e and μ'. Thus, we may perturb the hyperplane H to ensure it is not orthogonal to the line. Consequently, the resulting Coarse Bayesian Representation cannot be represented by any Maximum-Likelihood rule.

Note that the argument in the example above requires $|\Omega| \geq 3$; if there are only two states, then Coarse Bayesian behavior is a special case of Maximum-Likelihood updating. Even with $|\Omega| \geq 3$, however, the two theories do overlap: standard Bayesian updating is a special case of both Coarse Bayesian updating and Maximum-Likelihood updating. It is easy to see how Coarse Bayesian updating accommodates standard Bayesian updating (choose $\mathcal{P}$ so that each cell is a singleton). The next example shows how to express Bayesian updating as Maximum-Likelihood.

Example 3. To express standard Bayesian updating as a Maximum-Likelihood rule, take

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

where $\sqrt{\mu^e} := (\sqrt{\mu^e_\omega})_{\omega\in\Omega}$ and $\|\cdot\|$ denotes the standard norm on $\mathbb{R}^N$. Notice that $B(\mu^e|s) = \mu'$

if and only if $s \approx \mu'/\mu^e := (\mu'_\omega/\mu^e_\omega)_{\omega\in\Omega}$. Thus, it will suffice to verify that $L(\cdot|s)$ is maximized at μ' for such signals s. This is done as follows:

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

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

3.3 Path Dependence

Suppose a Coarse Bayesian agent observes a sequence of signals $\vec{s} = (s^1, \ldots, s^n)$ and updates beliefs sequentially: first, given prior μ^e and signal s^1, he applies the Coarse Bayesian procedure to arrive at some belief μ^P. This belief acts as his prior when processing s^2, and so on. Does his final belief, denoted $\mu^{\vec{s}}$, depend on the order of the signal realizations?

For a standard Bayesian, the order does not matter: as long as the product $s^1 s^2 \ldots s^n := (s^1_\omega s^2_\omega \ldots s^n_\omega)_{\omega\in\Omega}$ is a well-defined signal (that is, at least one entry is nonzero), then the final belief is simply $B(\mu^e|s^1 s^2 \ldots s^n)$.[1] Therefore, $\mu^{\vec{s}} = \mu^{\pi(\vec{s})}$ for all permutations $\pi(\vec{s})$ of $\vec{s}$; this is referred to as **path-independence**. If there exists an $\vec{s}$ and a permutation $\pi(\vec{s})$ such that $\mu^{\vec{s}} \neq \mu^{\pi(\vec{s})}$, then the agent exhibits **path-dependence**.[2]

Example 4. Not every (non-Bayesian) Coarse Bayesian Representation exhibits path dependence. For example, if $\mathcal{P}$ consists of a single cell (namely, $\{\Delta\}$), then $\mu^s = \mu^e$ for all $s \in S$. Less trivially, suppose $N = 2$ (so that Δ may be represented by the interval $[0,1]$) and consider $\mathcal{P} = \{[0,1), \{1\}\}$ with $\mu^{[0,1)} = 1/2$ and $\mu^{\{1\}} = 1$. It is straightforward to verify that this representation induces path-independent updating.

Although there are cases where Coarse Bayesians do not exhibit path dependence, there are many scenarios in which they do. The next proposition highlights a simple class of Coarse Bayesian Representations that exhibit path dependence.

[1] See Cripps (2018) for a general analysis of updating rules that are invariant to how an agent partitions histories of signals.

[2] Path-dependence is a key feature of Rabin and Schrag (1999); Coarse Bayesians (with appropriately specified partitions) can exhibit similar behavior precisely because of path dependence.

Proposition 3. *Suppose each μ^P ($P \in \mathcal{P}$) has full support. Then the agent exhibits path-dependence.*

4 A Behavioral Characterization

Let $\mathcal{A}$ denote the set of all nonempty, compact subsets of $X := \mathbb{R}^N$. Each $A \in \mathcal{A}$ is a *menu*, and elements $x \in A$ are (risky) *actions* the agent may take. If the agent chooses action $x \in A$, then he attains payoff x_ω in state ω.

I take as primitive a collection $c = (c^s)_{s \in S}$ of *signal-contingent choice correspondences*. For each $A \in \mathcal{A}$ and $s \in S$, $c^s(A) \subseteq A$ is the (nonempty) set of actions chosen by the agent after observing signal s. Formally, c is a function from $S \times \mathcal{A}$ to $\mathcal{P}(X)$, the set of all nonempty subsets of X.

Axiom 1 (Rationality). If $x, y \in A \cap B$, $x \in c^s(A)$, and $y \in c^s(B)$, then $x \in c^s(B)$.

Axiom 1 states that each c^s satisfies the Weak Axiom of Revealed Preference. On the domain $\mathcal{A}$, this is necessary and sufficient for the existence of a rationalizing preference relation $\succsim^s$.

Some additional notation is required for the next axiom. If $A, B \in \mathcal{P}(\mathbb{R}^N)$ and $\alpha \in [0, 1]$, let $\alpha A + (1 - \alpha)B := \{\alpha x + (1 - \alpha y) : x \in A,\ y \in B\}$.

Axiom 2 (Independence). $c^s(\alpha A + (1 - \alpha)B) = \alpha c^s(A) + (1 - \alpha)c^s(B)$.

Axiom 2 ensures that the derived relation $\succsim^s$ satisfies the standard (von Neumann-Morgenstern) Independence axiom. Next, endow $\mathcal{P}(\mathbb{R}^N)$ (and $\mathcal{A}$) with the Hausdorff metric.

Axiom 3 (Continuity). Each c^s is closed-valued and upper hemicontinuous.

Axiom 3 is needed to establish that $\succsim^s$ has closed contour sets and, hence, satisfies Archimedean continuity. Thus, Axioms 1–3 ensure the existence of a linear utility function representing $\succsim^s$ and, hence, rationalizing c^s.

For actions $x, y \in X$, write $x \geq y$ to indicate that $x_\omega \geq y_\omega$ for all $\omega \in \Omega$, and $x > y$ to indicate $x_\omega > y_\omega$ for all $\omega \in \Omega$. Combined with Axioms 1–3, the next axiom ensures that each $\succsim^s$ can be represented by expected utility (with prior μ^s), where μ^e has full support. Thus, the agent's behavior can be summarized by an updating rule $\mu = (\mu^s)_{s \in S}$.

Axiom 4 (Monotonicity). Let $A, B \in \mathcal{A}$.

(i) If $x > y$ for all $x \in A$ and $y \in B$, then $c^s(A \cup B) \subseteq B$ for all $s \in S$.

(ii) If $x \geq y \neq x$ for all $x \in A$ and $y \in B$, then $c^e(A \cup B) \subseteq B$.

The final two axioms apply to a special class of menus. If $A \in \mathcal{A}$ and $x \in A$, a *supporting half-space* for (x, A) is a closed half-space H such that $x \in \partial H \supseteq A$. A menu $A \in \mathcal{A}$ is *generic* if each $x \in A$ has at most one supporting hyperplane for (x, A). Intuitively, generic menus have full dimension in $\mathbb{R}^N$ and have smooth boundaries. The term "generic" is appropriate because any menu can be arbitrarily well-approximated by such menus. Let $\mathcal{A}^*$ denote the set of all generic menus.

Axiom 5 (Convexity). If $A \in \mathcal{A}^*$ and $c^s(A) = c^t(A)$, then $c^{\alpha s + \beta t}(A) = c^s(A)$.

Note that Axiom 5 only applies for $\alpha, \beta \geq 0$ such that $\alpha s + \beta t \in S$. Intuitively, Axiom 5 ensures that the updating rule μ is homogeneous and convex because in generic menus A, $c^s(A) = c^t(A)$ if and only if $\mu^s = \mu^t$.

Finally, for any nonempty $A \subseteq X$ and $s \in S$, let $sA := \{sx : x \in A\}$. Effectively, sA perturbs the payoffs of actions in A by scaling down the payoff in state ω by a factor s_ω.

Axiom 6 (Confirmation). If $A \in \mathcal{A}^*$ and $sc^t(A) \subseteq c^e(sA)$, then $c^s(A) = c^t(A)$.

Axiom 6 asserts that $c^s(A) = c^t(A)$ (hence, $\mu^s = \mu^t$) if, after observing signal t, an optimal action from A remains optimal in sA under e if its payoffs are scaled by s. Roughly speaking, Bayesian updating requires an equivalence between scaling payoffs and scaling state likelihoods: choices from A after Bayesian updating of signal s should be proportional (by factor s) to those chosen under the prior μ^e from sA. Thus, Axiom 6 essentially states that if behavior at t is consistent with Bayesian updating of signal s, then $\mu^s = \mu^t$. Consequently, the updating rule μ is confirmatory.

Theorem 1. *The family c satisfies Axioms 1–6 if and only if there is an updating rule μ with a Coarse Bayesian Representation $\langle \mathcal{P}, (\mu^P)_{P \in \mathcal{P}} \rangle$ such that, for all $s \in S$ and $A \in \mathcal{A}$,*

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

If μ' and $\langle \mathcal{P}', (\mu^{P'})_{P' \in \mathcal{P}'} \rangle$ also represent c, then $\mu' = \mu$, $\mathcal{P}' = \mathcal{P}$, and $(\mu^{P'})_{P' \in \mathcal{P}'} = (\mu^P)_{P \in \mathcal{P}}$.

This theorem provides a complete characterization of Coarse Bayesian behavior in the context of expected utility maximization. By analyzing the agent's signal-contingent choices from menus A, one can uniquely determine his prior as well as $\langle \mathcal{P}, (\mu^P)_{P \in \mathcal{P}} \rangle$.

4.1 Sophistication

The concept of Coarse Bayesian updating gives rise to two natural notions of sophistication: finer partitions, or richer sets of "theories" (feasible beliefs). In this section, I characterize when one agent (represented by the family $\dot{c}$) is more sophisticated than another (represented by the family c).

The analysis in this section works best under the assumption of a common prior: $\mu = \dot{\mu}$. This assumption can be expressed behaviorally as: $c^e(A) = \dot{c}^e(A)$ for all $A \in \mathcal{A}^*$ (in fact, existence of a single $A \in \mathcal{A}^*$ such that $c^e(A) = \dot{c}^e(A)$ is sufficient to guarantee $\mu = \dot{\mu}$). Throughout, I assume c and $\dot{c}$ have Coarse Bayesian Representations $\langle \mathcal{P}, (\mu^P)_{P\in\mathcal{P}} \rangle$ and $\langle \mathcal{Q}, (\dot{\mu}^Q)_{Q\in\mathcal{Q}} \rangle$, respectively. The objective is to perform comparative statics on these parameters in terms of c and $\dot{c}$ without having to fully identify them.

Proposition 4. *Suppose $\mu^e = \dot{\mu}^e$. The following are equivalent:*

(i) $\mathcal{Q}$ refines $\mathcal{P}$ (that is, each cell of $\mathcal{P}$ is a union of cells of $\mathcal{Q}$).

(ii) For all $A \in \mathcal{A}$, $\dot{c}^s(A) = \dot{c}^t(A)$ implies $c^s(A) = c^t(A)$.

Proposition 4 establishes that a Coarse Bayesian has a finer partition if his choices are "more responsive" changes to the signal (there are fewer scenarios where two signals result in the same choices). Since each cell of the partition is associated with a unique theory of the world, there is a sense in which the more sophisticated agent entertains more theories. However, Proposition 4 does not guarantee that $\mu(S) \subseteq \dot{\mu}(S)$, where $\mu(S) := \{\mu^s : s \in S\}$ (ie, if $P \in \mathcal{P} \cap \mathcal{Q}$, it is not necessarily the case that $\dot{\mu}^P = \mu^P$).

Proposition 5. *Suppose $\mu^e = \dot{\mu}^e$. The following are equivalent:*

(i) $\mu(S) \subseteq \dot{\mu}(S)$.

(ii) For all $A \in \mathcal{A}$, $sc^t(A) \subseteq c^e(sA)$ implies $s\dot{c}^t(A) \subseteq \dot{c}^e(sA)$.

Proposition 5 establishes that the agent entertains more theories ($\mu(S) \subseteq \dot{\mu}(S)$) if and only if he is "more Bayesian": there is a larger set of signals s such that $\mu^s = B(\mu^e|s)$. Note that this does not guarantee that $\mathcal{Q}$ refines $\mathcal{P}$. By combining Propositions 4 and 5, a more sophisticated agent (in the sense of holding a finer partition as well as entertaining more theories) is one who is both more responsive to changes in signals and more Bayesian in his processing of signals.

5 Valuing Information

In this section, I examine how a Coarse Bayesian agent values information (Blackwell experiments). I employ the framework of actions and menus from section 4.

An **experiment** is a matrix σ with entries in $[0,1]$ and $|\Omega|$ rows where each row is a probability distribution and each column has at least one nonzero entry. Thus, each column is a signal $s \in S$, and each row represents a probability distribution over signals. The requirement that each row constitutes a probability distribution can be re-expressed as $\sum_{s\in\sigma} s = e$, where the notation $s \in \sigma$ indicates that s is a column of σ. Let $\mathcal{E}$ denote the set of all experiments. For experiments σ, σ', the relation $\sigma \sqsupseteq \sigma'$ indicates that σ is more informative than σ' in the sense of Blackwell (1951, 1953). This is a partial order on $\mathcal{E}$.

Given a menu $A \in \mathcal{A}$, the agent's **value of information** is the function $V^A : \mathcal{E} \to \mathbb{R}$ given by

$$V^A(\sigma) := \sum_{\omega\in\Omega} \mu^e_\omega \sum_{s\in\sigma} x^s_\omega \quad \text{subject to} \quad x^s \in \operatorname*{argmax}_{x\in A} x \cdot \mu^s$$

for all $\sigma \in \mathcal{E}$. This is the ex-ante expected utility for an agent who correctly anticipates his future behavior; in particular, he is cognizant of his own updating procedure. Alternatively, if one assumes that the prior μ^e is (objectively) the correct distribution over states, then $V^A(\sigma)$ represents the actual average utility experienced by the agent, regardless of whether he correctly anticipates his own behavior.

5.1 The Blackwell Ordering

The function V^A **satisfies the Blackwell ordering** if $\sigma \sqsupseteq \sigma'$ implies $V^A(\sigma) \geq V^A(\sigma')$. As is well-known, a standard Bayesian agent's value of information satisfies the Blackwell ordering in all menus A. The objective of this section is to characterize if and when a Coarse Bayesian satisfies the Blackwell ordering. Throughout, I consider Coarse Bayesian representations $\langle \mathcal{P}, (\mu^P)_{P\in\mathcal{P}} \rangle$ that are **nontrivial**: $\mathcal{P}$ contains at least two cells.

Proposition 6. *Suppose $\mathcal{P}$ is nontrivial. If each V^A ($A \in \mathcal{A}$) satisfies the Blackwell ordering, then the agent is Bayesian: $\mu^s = B(\mu^e|s)$ for all $s \in S$.*

This result establishes that, conditional on being Coarse Bayesian, the agent is actually Bayesian if his value of information adheres to the Blackwell ordering in all menus. In particular, a Coarse Bayesian who is not Bayesian must exhibit violations of the Blackwell ordering in some menus. The next definition provides a condition characterizing exactly when a Coarse Bayesian satisfies the Blackwell ordering.

Definition 4. Let $A \in \mathcal{A}$ and $\overline{\mathcal{P}} := \{\text{cl}P : P \in \mathcal{P}\}$.

(i) The **support** of $x \in A$ is the set $\Delta^A(x) := \{\hat{\mu} \in \Delta : x \cdot \hat{\mu} \geq y \cdot \hat{\mu} \ \forall y \in A\}$. Let $\Delta^A := \{\Delta^A(x) : x \in A\}$.

(ii) A set $\tilde{\Delta} \in \Delta^A$ is $\overline{\mathcal{P}}$-measurable if it is a union of members of $\overline{\mathcal{P}}$.

(iii) A set $\overline{P} \in \overline{\mathcal{P}}$ is Δ^A-measurable if it is a union of members of Δ^A.

(iv) A and $\mathcal{P}$ are **co-measurable** if the union of all $\overline{P}$-measurable and Δ^A-measurable sets is Δ.

Intuitively, the support of $x \in A$ is the set of beliefs $\hat{\mu}$ for which x maximizes expected utility in A. The concept of co-measurability can be understood by considering two extreme cases. In on case, each support set $\tilde{\Delta}$ can be expressed as a union of members of $\overline{\mathcal{P}}$. Thus, the (Coarse Bayesian) agent chooses x from A at a signal s if and only if a standard Bayesian would also choose x at s. At the other extreme, each $\overline{P} \in \overline{\mathcal{P}}$ can be expressed as a union of members of Δ^A. In this case, one can remove appropriate elements x from the menu in order to make the Coarse Bayesian's choices at all signals s coincide with those of the standard Bayesian from the original menu. The general definition allows these two cases to intermingle: some support sets can be expressed as unions of members of $\overline{\mathcal{P}}$, while some members of $\overline{\mathcal{P}}$ can be expressed as unions of support sets. The fact that co-measurability ensures that the Coarse Bayesian's behavior can be expressed as that of a standard Bayesian suggests that the Coarse Bayesian must adhere to the Blackwell ordering in such cases.

Proposition 7. *Suppose $\mathcal{P}$ is nontrivial and $A \in \mathcal{A}$. The following are equivalent:*

(i) A and $\mathcal{P}$ are co-measurable.

(ii) V^A satisfies the Blackwell ordering.

An immediate consequence of this result is that $\mathcal{P}$ can be deduce by examining when (for which menus) the agent satisfies the Blackwell ordering. The next section establishes even stronger identification results.

5.2 Sophistication and Identification

This section explores identification issues (and the characterization of finer partitions) from the perspective of the value of information. I consider two (nontrivial) Coarse Bayesians, $\langle \mathcal{P}, (\mu^P)_{P \in \mathcal{P}} \rangle$ and $\langle \mathcal{Q}, (\dot{\mu}^Q)_{Q \in \mathcal{Q}} \rangle$. As in section 4.1, I restrict attention to agents with a common prior: $\mu^e = \dot{\mu}^e$. Let V^A and $\dot{V}^A$ associate the value of information in menu A for the two agents.

By Proposition 7, it follows that $\mathcal{P} = \mathcal{Q}$ (up to closures of cells) if and only if V^A and $\dot{V}^A$ satisfy the Blackwell ordering on the same subset of nontrivial menus A. While somewhat illuminating, this characterization is not testable. Fortunately, a different condition on the functions V^A and $\dot{V}^A$ can be used to identify (and compare) the partitions.

An experiment σ' is a **garbling** of σ if there is a stochastic matrix M such that $\sigma' = \sigma M$. It is a **coarse garbling** if each entry of M belongs to the set $\{0, 1\}$ (each row of M is a degenerate probability distribution). As is well-known, $\sigma \sqsupseteq \sigma'$ if and only if σ' is a garbling of σ.

Proposition 8. *Suppose $\mu^e = \dot{\mu}^e$. The following are equivalent:*

(i) $\mathcal{Q}$ refines $\mathcal{P}$.

(ii) If σ' is a coarse garbling of σ and $\dot{V}^A(\sigma) = \dot{V}^A(\sigma')$, then $V^A(\sigma) = V^A(\sigma')$.

This proposition states that the agent has a finer partition if and only if his value of information is more responsive to coarse garblings: if $V^A(\sigma) \neq V^A(\sigma')$, then $\dot{V}^A(\sigma) \neq \dot{V}^A(\sigma')$. This is analogous to the previous characterization (utilizing choices c and $\dot{c}$) in terms of responsiveness to signals.

In general, indifference to a coarse garbling of σ (in sufficiently rich menus such as generic menus) indicates that there are signals $s, t \in \sigma$ such that the agent's choice from A coincides at s, t, and $s+t$, forcing μ^s and μ^t to belong to the same cell $P \in \mathcal{P}$. Thus, non-indifference indicates that the signals belong to different cells. Greater responsiveness to coarse garblings, then, indicates that fewer signals are absorbed in this fashion, forcing $\mathcal{Q}$ to be finer than $\mathcal{P}$.

The final result shows that a Coarse Bayesian's value of information (hence, his preferences for information) fully reveal $\langle \mathcal{P}, (\mu^P)_{P \in \mathcal{P}} \rangle$. In fact, the identification can be achieved even if one restricts attention to pairs of experiments that are Blackwell comparable:

Proposition 9. *The following are equivalent:*

(i) $\mathcal{P} = \mathcal{Q}$ and $(\mu^P)_{P \in \mathcal{P}} = (\dot{\mu}^Q)_{Q \in \mathcal{Q}}$.

(ii) $V^A = \dot{V}^A$ for all $A \in \mathcal{A}$.

(iii) For all $A \in \mathcal{A}$ and $\sigma \sqsupseteq \sigma'$, $V^A(\sigma) \geq V^A(\sigma') \Leftrightarrow \dot{V}^A(\sigma) \geq \dot{V}^A(\sigma')$.

6 Conclusion

In this paper, I have introduced a new model of non-Bayesian updating, *Coarse Bayesian Updating*, and characterized it both directly (taking signal-contingent beliefs as given) and behaviorally (taking signal-contingent choices as primitive). The agent is summarized by a partition of the set of all probability distributions over a state space, together with a representative belief for each cell of the partition. This allows several interpretations of the procedure, and the rich domain of noisy signals enables unique identification of these parameters. There is also a rich connection between the parameters and the agent's value of information. Coarse Bayesian updating is related to several other theories and concepts of non-Bayesian updating, and accommodates a number of experimental findings.

An advantage of my framework is that it employs standard primitives that frequently appear in applications. The use of noisy signals (relative to the state space), for example, allows one to directly import Coarse Bayesian updating into familiar settings in economics and game theory. Exploring the implications of Coarse Bayesian updating in such settings may be a fruitful avenue for future research.

A Proof of Theorem 1

I prove that if c satisfies Axioms 1–6, then c has the desired representation (the converse is straightforward). For each $s \in S$, define the binary relation $\succsim^s$ on X by:

$$x \succsim^s y \Leftrightarrow \exists A \in \mathcal{A} \text{ such that } x, y \in A \text{ and } x \in c^s(y).$$

Lemma 1. *For each $s \in S$, $\succsim^s$ is complete, transitive, and satisfies $c^s(A) = \{x \in A : x \succsim^s y \ \forall y \in A\}$ for all $A \in \mathcal{A}$.*

Proof. First, $\succsim^s$ is complete because $c^s(\{x, y\}) \neq \emptyset$ for all $x, y \in X$. For transitivity, suppose $x \succsim^s y \succsim^s z$. Then there exists $A \in \mathcal{A}$ such that $x, y \in A$ and $x \in c^s(A)$, and there exists $A' \in \mathcal{A}$ such that $y, z \in A'$ and $y \in c^s(A')$. Let $B = A \cup A'$, and observe that $B \in \mathcal{A}$. Since $c^s(B) \neq \emptyset$, there exists $w \in c^s(B)$. If $w \in A$, then WARP (Axiom 1) implies $x \in c^s(B)$. If $w \in A'$, then WARP implies $y \in c^s(B)$; applying WARP again yields $x \in c^s(B)$. Thus, in all cases, $x \in c^s(B)$, so that $x \succsim^s z$.

We now prove that $c^s = c_{\succsim^s}$, where $c_{\succsim^s}(A) := \{x \in A : x \succsim^s y \ \forall y \in A\}$. Let $A \in \mathcal{A}$. To see that $c^s(A) \subseteq c_{\succsim^s}(A)$, suppose $x \in c^s(A)$. Then $x \succsim^s y$ for all $y \in A$, so that $x \in c_{\succsim^s}(A)$. For the converse inclusion, suppose $x \in c_{\succsim^s}(A)$. Let $y \in c^s(A) \neq \emptyset$. Since $y \in A$ and $x \in c_{\succsim^s}(A)$, we have $x \succsim^s y$. Hence, there exists $A' \in \mathcal{A}$ such that $x, y \in A'$ and $x \in c^s(A')$. Since $x \in c^s(A')$ and $y \in c^s(A)$, WARP implies $x \in c^s(A)$, as desired. $\square$

Lemma 2. *For each $s \in S$, the relation $\succsim^s$ is continuous: for all $x \in X$, the sets $U(x) := \{y \in X : y \succsim^s x\}$ and $L(x) := \{y \in X : x \succsim^s y\}$ are closed.*

Proof. By Axiom 3, c^s is closed-valued and upper hemicontinuous; therefore, c^s has the closed-graph property: if $A^n \to A$, $x^n \to x$, and $x^n \in c^s(A^n)$ for all n, then $x \in c^s(A)$.

To see that upper contour sets are closed, fix x and suppose $y^n \to y$ where $y^n \in U(x)$ for all n. Then there exist $A^n \in \mathcal{A}$ such that, for all n, $x, y^n \in A^n$ and $y^n \in c^s(A^n)$. By WARP (Axiom 1), we have $y^n \in c^s(\{x, y^n\})$ for all n. Clearly, $\{x, y^n\} \to \{x, y\}$. Thus, by the closed-graph property of c^s, we have $y \in c^s(\{x, y\})$ and, hence, $y \succsim^s x$.

For the lower contour sets, fix x and suppose $y^n \to y$ where $y^n \in L(x)$ for all n. Then there exist $A^n \in \mathcal{A}$ such that, for all n, $x, y^n \in A^n$ and $x \in c^s(A^n)$. By WARP, $x \in c^s(\{x, y^n\})$ for all n. Clearly, $\{x, y^n\} \to \{x, y\}$. Letting $x^n = x$ for all n, we have $x^n \to x$. Thus, by the closed-graph property, $x \in c^s(\{x, y\})$, so that $x \succsim^s y$. $\square$

Lemma 3. *For each $s \in S$, there exists a unique $\mu^s \in \Delta\Omega$ such that $x \succsim^s y$ if and only if, for all $x, y \in X$, $x \cdot \mu^s \geq y \cdot \mu^s$. The prior μ^e has full support.*

Proof. Let $s \in S$. By Lemma 1, $\succsim^s$ is complete and transitive. By Lemma 2 and a standard argument, $\succsim^s$ satisfies Archimedean continuity: if $x \succ^s y \succ^s z$, then there exists $\alpha, \beta \in (0,1)$ such that $\alpha x + (1-\alpha)z \succ^s y \succ^s \beta x + (1-\beta)z$. Finally, Axiom 2 implies that $\succsim^s$ satisfies the Independence axiom: if $x \succ^s y$ and $\alpha \in (0,1)$, then $\alpha x + (1-\alpha)z \succ^s \alpha y + (1-\alpha)z$ for all $z \in X$. Thus, by the Mixture Space Theorem, $\succsim^s$ has a representation $U^s : X \to \mathbb{R}$ such that $U^s(\alpha x + (1-\alpha)y) = \alpha U^s(x) + (1-\alpha)U^s(y)$ for all $x, y \in X$ and $\alpha \in [0,1]$ (that is, U^s is *linear*).

By part (i) of Axiom 4, each U^s ranks constant actions $\overline{x}, \overline{y}$ the same way: for all $s, t \in S$, $U^s(\overline{x}) \geq U^s(\overline{y})$ if and only if $U^t(\overline{x}) \geq U^t(\overline{y})$. Thus, each U^s is of the form $U^s(x) = a^s \cdot x$, where $a^s \in \mathbb{R}^N \backslash \{0\}$ and $a^s \geq 0$. By part (ii) of Axiom 4, we have $a^e > 0$. The result follows by taking $\mu^s := \frac{1}{a^s \cdot e} a^s$. $\square$

By Lemma 3, c is represented by an updating rule $\mu = (\mu^s)_{s \in S}$. Observe that if $A \in \mathcal{A}^*$ and A is strongly convex, then $c^s(A)$ is a singleton for all $s \in S$, and $c^s(A) = c^t(A)$ implies $\mu^s = \mu^t$ (the unique supporting hyperplane for $x \in c^s(A) = c^t(A)$ has normal given by μ^s). Thus, by Axiom 5, the updating rule μ is homogeneous and convex.

Pick any $t \in S$ and let $s \in S$ such that $s\mu^e \approx \mu^t$ (that is, $\mu^t = B(\mu^e|s)$). Such an s exists because μ^e has full support. We want to show that $\mu^s = \mu^t$ (so that the updating rule μ is confirmatory). Let $x \in c^t(A)$, where $A \in \mathcal{A}^*$ is strongly convex. Then $x \cdot \mu^t \geq y \cdot \mu^t$ for all $y \in A$. Since $s\mu^e \approx t$, it follows that

$$
\begin{aligned}
& x \cdot (s\mu^e) \geq y \cdot (s\mu^e) \ \forall y \in A \\
\Rightarrow \ & (sx) \cdot \mu^e \geq (sy) \cdot \mu^e \ \forall y \in A \\
\Rightarrow \ & (sx) \cdot \mu^e \geq x' \cdot \mu^e \ \forall x' \in sA
\end{aligned}
$$

and therefore $sx \in c^e(sA)$. Thus, $sc^t(A) \subseteq c^e(sA)$, so that $c^s(A) = c^t(A)$ by Axiom 6. As noted above, $\mu^s = \mu^t$ then follows from the fact that A is generic and strongly convex. To conclude the proof, apply Proposition 1 to get that μ has a Coarse Bayesian Representation.

B Proofs for Section 3

To be written.

C Proofs for Section 5

To be written.

References

Anscombe, F. J. and R. J. Aumann (1963). A definition of subjective probability. *The annals of mathematical statistics 34*(1), 199–205.

Blackwell, D. (1951). Comparison of experiments. In *Proceedings of the second Berkeley symposium on mathematical statistics and probability*, Volume 1, pp. 93–102.

Blackwell, D. (1953). Equivalent comparisons of experiments. *The annals of mathematical statistics 24*(2), 265–272.

Cripps, M. W. (2018). Divisible updating. *Working Paper*.

Jakobsen, A. (2016). Dynamic (in)consistency and the value of information. *Working Paper*.

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

Rabin, M. and J. L. Schrag (1999). First impressions matter: A model of confirmatory bias. *The Quarterly Journal of Economics 114*(1), 37–82.

Savage, L. J. (1972). *The foundations of statistics*. Courier Corporation.