# Foundations for Bayesian Updating

Eran Shmaya *    Leeat Yariv†‡

Current Version: February 26, 2008

Abstract. We provide a simple characterization of updating rules that can be rationalized as Bayesian. Namely, we consider a general setting in which an agent observes finite sequences of signals and reports probabilistic predictions on the underlying state of the world. We study when such predictions are consistent with Bayesian updating, i.e., when does there exist some theory about the signal generation process that would be consistent with the agent behaving as a Bayesian updater. We show that the following condition is necessary and sufficient for the agent to appear Bayesian: the probability distribution that represents the agent's belief after observing any finite sequence of signals is a convex combination of the probability distributions that represent her beliefs conditional on observing sequences of signals that are the possible continuations of the original sequence. This condition cannot be derived from the ones the literature has identified when confounding the problem with maximization of expected utility. Additional restrictions are identified for all histories of signals to be given positive probability under the identified information generation process, and for the agent's theory to entail conditional independence or exchangeability of signals.

**JEL classification:** D01, D80, D83.

**Keywords:** Bayesian Updating, Axiomatic Foundations.

---

*Information, Science, and Technology, Caltech, Pasadena CA 91125, USA. http://www.ist.caltech.edu/~eshmaya/ e-mail: eshmaya@ist.caltech.edu

†Division of the Humanities and Social Sciences, Caltech, Pasadena, CA 91125, USA. http://www.hss.caltech.edu/~lyariv/index.htm e-mail: lyariv@hss.caltech.edu

‡We are grateful to Matt Jackson, Ehud Lehrer, and Marciano Siniscalchi for very useful conversations. We also thank Kim Border, Chris Chambers, Federico Echenique, Andrea Mattozzi, and Preston McAfee for helpful comments.

# 1. Introduction

## 1.1 Overview

The assumption that agents use Bayesian updating to revise beliefs in light of new evidence is ubiquitous in the social sciences. De Finetti (1937) was one of the first to attempt tying underlying statistical rules an agent may be following to their behavior (using some of Ramsey's 1931 interpretations of probabilities as subjective). The 70 years that followed provided an abundance of inquiries focusing mostly on observables tying information to behavior and assessing jointly when these observables may be rationalized as coming from Bayesian updating *and* maximization of some expected utility. Consequently, the extant work confounds conditions for Bayesian behavior with some form of optimization. The goal of the current paper is to identify conditions for a prediction rule, a mapping from information into stated beliefs, to be consistent with Bayesian updating.

To fix ideas, consider a simple example in which an agent is given lists of attributes of a person called Bob and required to make a prediction on the probability Bob is an economist. Suppose the agent is told that Bob is 6 feet tall, or 6 feet tall and wearing glasses, or 6 feet tall and not wearing glasses, etc. If we could elicit responses for all such questions, when could we find some belief tying economists and non-economists to attributes that make the agent's reports look *as if* they were deduced using Bayes rule?[1] Certainly, if the agent reports 80% probability of Bob being an economist when Bob is said to be 6 feet tall and wearing glasses, and 70% if Bob is 6 feet tall and not wearing glasses, then the agent cannot be a Bayesian and report, say, 50% probability of Bob being an economist when the only information she gets is that Bob is 6 feet tall. In fact, if the agent has some joint probability in mind tying attributes and professions, it better be the case that knowing only that Bob is 6 feet tall leads to a prediction that is *in between* 70% and 80%.

[1]While the agent knows Bob's potential attributes (e.g., height, dress, age, etc.), here the agent realizes that the particular attributes that are revealed are not correlated with Bob's profession (for example, the revealed attributes could be determined randomly by a computer).

More generally, consider an agent who, provided any finite sequence of signals, is asked to report a *prediction* – a probability distribution over some underlying state space, which represents her beliefs conditional on observing that sequence of signals. We say the agent behaves as a Bayesian whenever there exists some (probabilistic) theory tying states to signal realizations based on which a Bayesian updater would make identical predictions to our agent. In other words, the agent is indistinguishable from someone using Bayes rule with that theory. The example above then generalizes directly. Whenever the agent behaves as a Bayesian, it must be the case that for any sequence of signals, the corresponding predictions are within the *convex hull* of the predictions corresponding to all continuations of that signal sequence. Our main result illustrates that this condition is, in fact, also sufficient for the agent to behave as a Bayesian.

The proof of our result is simple and constructive in nature. We find a consistent sequence of marginal distributions corresponding to realizations of finite sequences of signals. When the signal sequences can persist indefinitely, we can extend the sequence of distributions (using Kolmogorov's Extension Theorem) to a distribution tying any sequence of signals to a belief on the underlying states of the world (which is an arbitrary Borel space, finite or infinite).

While the type of experiment required to identify Bayesian behavior is gedanken in nature, it is worth noting that our results can easily be restricted to questionnaires pertaining only to finite sequences of signals. Furthermore, even if signal sequences are potentially infinite, non-Bayesian behavior will be identified in finite time using our characterization result.

Even when one has access to partial responses of the sort described above there are still consistency requirements that the available responses must satisfy in order to be indistinguishable from those of a Bayesian updater. Namely, Theorem 2 illustrates the following consistency requirement. Consider the sequence of signals $s_1, s_2, ...$ We could describe the potential realizations as a tree in which the root corresponds to no information, branching out to nodes that correspond to different realizations of $s_1$, each branching out to nodes corresponding to the realizations of $s_2$, and so on. At each node we can put the prediction belief, if it is available. Take any sub-tree of the tree in which the leaves have specified predictions.

Similar to the description above, it is rather straightforward to show that a Bayesian updater must be consistent within the sub-tree. That is, any specified prediction on the root of the sub-tree must be within the convex hall of all predictions in the leaves of the sub-tree. As it turns out, this is also a sufficient condition.

One may also worry about the plausibility given the *deduced theory* of some of the signal sequences that our framework requires posing to agents. In particular, the deduced theory that makes an agent indistinguishable from a Bayesian updater may potentially place zero probability on some signal sequences. We identify additional restrictions assuring that this is not the case. Namely, for any sequence of signals, the corresponding predictions need to be within the *relative interior* of the convex hull of the predictions corresponding to all continuations of that signal sequence.

Another natural restriction on the deduced theory would pertain to the particular signal generation process the agent has in mind. Indeed, in much of economic modelling signals are generated by a repeat and independent sample from some distribution that depends on the underlying state, corresponding to conditionally independent and identically distributed (i.i.d.) signals (in fact, the statistics literature views Bayesian updating as updating over such processes). Requiring predictions to correspond to Bayesian updating with respect to a theory under which signals are conditionally i.i.d. is rather restrictive. To glean some intuition, suppose there are only two states of nature, $\{0,1\}$ , and two potential signals, $\{0,1\}$. Signals being conditionally i.i.d. corresponds to only three parameter: one determining the probability of either state, and two for the probability that the signal 1 gets realized when the state is either 0 or 1. In our penultimate result, we spell out the rather strict test that determines whether predictions match Bayesian updating with a theory entailing conditionally i.i.d. signals.

An alternative candidate for the deduced signal generation process is *conditional exchangeability* of signals. Conditional exchangeability is a generalization of the notion of conditional i.i.d. signals. It requires that given any realization of the state, the conditinoal probability of any finite sequence of signals does not depend on the order in which the signals appear in the sequence. A necessary condition for the agent to be Bayesian according to a theory that exhibits conditional exchangeability is that all elicited predictions are invariant to the order of signals. Our last result shows that this condition is not sufficient and provides a necessary complementary condition.

Utility considerations are (intentionally) absent in our analysis. However, we would like to stress two points. First, the question of when an agent is Bayesian is not subsumed in the analysis pertaining to expected utility by a Bayesian updater (see the literature review below for a more elaborate comparison).[2] Second, elicitation of beliefs could be thought of as elicitation of actions when payoffs are specified in a particular way (see Offerman, Sonnemans, van de Kuilen, and Wakker, 2006, and references therein).

### 1.2 Related Literature

De Finetti's theorem (see de Finetti, 1937, and a recent overview by Cifarelli and Regazzini, 1996) explained why exchangeable observations are conditionally independent given some (usually) unobservable quantity to which an epistemic probability distribution can be assigned. It crystallized the differences between Bayesian and frequentist methods in statistical inference. Indeed, frequentists often treat observations as independent while Bayesians treat them as exchangeable.

Savage (1954) and Anscombe and Aumann (1963) opened the door for assessing when observed behavior may arise from some form of expected utility maximization in which the agent possesses *subjective* probabilities over states. In a way, our approach can be thought of as identifying a *subjective signaling structure* that can explain agents' behavior as arising from Bayesian updating.

The literature that followed suit is immense in scope and we shall therefore not attempt to cover it in full. Broadly speaking, the underlying model that literature focuses on takes mappings from information into actions as the atom of observation and considers the condi-

[2]One could trivially embed our setup in the standard von-Neumann - Morgenstern world to get foundations for Bayesian updating *and* maximization of expected utility.

tions under which these mappings may be consistent with Bayesian updating *together with* maximization of (some) expected utility (see, e.g., Green and Park, 1994).

Gilboa and Lehrer (1991) consider the problem from a different direction. They find necessary and sufficient conditions for a real valued function on the partitions of a measure space to be the value of information function for a Bayesian decision maker. Their analysis is similar to that corresponding to expected utility elicitation in that underlying the value of information is some action space and mapping from actions and states to payoffs (or utility levels) that warrant the difference in value corresponding to different partitions.

Recently, there have also been several notable attempts to pin down non-standard behavior with regards to information (see, for instance, Billot, Gilboa, Samet, and Schmeidler, 2005 and Epstein, Noor, and Sandroni, 2006).

Experimentally, there are several investigations trying to assess whether laboratory subjects appear Bayesian, notably El-Gamal and Grether (1995), who identify several behavioral strategies subjects use when confronted with incentivized updating tasks.

### 1.3 Structure of the Paper

In the following section we spell out the model. Section 3 describes the main result for finite signal spaces and a complete set of predictions. The section also provides the proof as it is rather simple and instructive. Section 4 extends the result to situations where an arbitrary subset of observations is available, while Section 5 extends the result to a general setting in which signals are taken from an arbitrary set (finite or not). Section 6 considers the case in which the identified Bayesian updating rules are restricted to place positive probability on all signal profiles and Section 7 illustrates the necessary and sufficient conditions for a prediction to be rationalized by a conditionally i.i.d. or exchangeable theory regarding the signal generation process. Section 8 briefly discusses the implications one can draw from our analysis to a set formulation of the problem. Section 9 concludes. Technical proofs are relegated to the Appendix.

## 2. The Model

Let $K$ be a Borel space[3] of *states of nature* and let $S$ be a finite set of *signals*. We denote by $S^*$ the set of all finite sequences of elements of $S$:

$$S^* = \bigcup_{n \geq 0} S^n.$$

For $\alpha = (s_1, \ldots, s_n) \in S^*$ and $s_{n+1} \in S$ we denote by $\alpha * s_{n+1}$ the element of $S^*$ given by $(s_1, \ldots, s_n, s_{n+1})$.

An *updating rule* is a function $\sigma : S^* \to \Delta(K)$. For $\alpha \in S^*$ we denote the image of $\alpha$ under $\sigma$ by $\sigma[\alpha]$. The interpretation is that after seeing a sequence $\alpha$ of signals the agent's prediction regarding the underlying state of nature is given by $\sigma[\alpha]$.

Before defining the concept of Bayesian updating rules, we recall the notion of a *discrete conditional distribution*: Let $P$ be a Borel space, $\xi : P \to X$ a measurable function over $P$ with values in some finite set $X$, and let $\mu$ be a probability measure over $K \times P$. Then there exists probability measures $\mu(\cdot|x)$ over $K$ for every $x \in X$ such that for every Borel subset $B$ of $K$ one has

$$\mu_K(B) = \sum_{x \in X} \mu(B|x)\mu_X(x),$$

where $\mu_K$ is the marginal distribution of $\mu$ over $K$:

$$\mu_K(B) = \mu(B \times P),$$

and $\mu_X$ is the marginal of $\mu$ over $X$:

$$\mu_X(x) = \mu(K \times \xi^{-1}(x))).$$

Slightly abusing notation, we will omit the subscripts $K$ and $X$ corresponding to the marginal

[3] Throughout the paper, all Borel spaces are, by assumption or by construction, standard. Recall that the sigma-algebra of a standard Borel space is generated by some Polish topology, i.e., a topology that is separable and metrizable by a complete metric. The $\sigma$-algebra of a Borel space is assumed to be fixed. Notions like a Borel set, Borel probability measure, and Borel function always correspond to this $\sigma$-algebra.

distributions $\mu_K$ and $\mu_X$ and denote them by $\mu$. Thus we write the last equation as follows:

$$\mu(B) = \sum_{x\in X} \mu(B|x)\mu(x). \tag{1}$$

The probability distribution $\mu(\cdot|x)$ is called *the conditional distribution over $K$ given $x$* and is uniquely determined whenever $\mu(x) > 0$.

In our setup, $P = S^\infty$ is the Borel space of infinite sequences of signals, $X = S^n$ is the finite set of sequences of length $n$, and $\xi : P \to X$ is the natural projection given by

$$\xi(s_1, s_2, \dots) = (s_1, \dots, s_n).$$

**Definition (Bayesian Updating)** *An updating rule $\sigma : S^* \to \Delta(K)$ is Bayesian if there exists a probability measure $\mu$ over $K \times S^\infty$ such that, for every $n$ and every Borel subset $B$ of $K$,*

$$\sigma[s_1, \dots, s_n](B) = \mu(B|s_1, \dots, s_n). \tag{2}$$

## 3. Main Result

Our main result establishes a necessary and sufficient condition for an updating rule to be Bayesian. This condition is rather simple and intuitive to phrase. For any sequence of signals $(s_1, ..., s_n)$, consider all possible continuation signals $(s_1, ..., s_n, \tilde{s}_{n+1})$. If the updating rule is derived from Bayesian updating, it must be the case that there is a certain weight associated to each such continuation and, therefore, that the prediction corresponding to $(s_1, ..., s_n)$ is a weighted average of the predictions corresponding to all continuations $(s_1, ..., s_n, \tilde{s}_{n+1})$. In particular, the prediction corresponding to $(s_1, ..., s_n)$ is in the convex hull of those corresponding to $(s_1, ..., s_n, \tilde{s}_{n+1})$. The theorem's claim is that this condition is, in fact, not only necessary but also sufficient. Formally,

**Theorem 1** *Let $K$ be a Borel set of states of nature, $S$ a finite set of signals and $\sigma : S^* \to \Delta(K)$ an updating rule.*

1. *If $\sigma$ is Bayesian and the probability measure $\mu$ over $K \times S^\infty$ explains $\sigma$ then*

$$\sigma[s_1, \ldots, s_n] \in \mathrm{Conv}\{\sigma[s_1, \ldots, s_n, s_{n+1}] | s_{n+1} \in S\} \tag{3}$$

*for every $s_1, \ldots, s_n \in S$ such that $\mu(s_1, \ldots, s_n) > 0$, where* Conv *stands for the convex hull.*

2. *If (3) is satisfied for every $s_1, \ldots, s_n \in S$ then $\sigma$ is Bayesian.*

We now prove Theorem 1. While necessity follows almost directly, sufficiency is slightly more challenging. The idea is to construct a consistent sequence of marginal distributions which we then extend using Kolmogorov's Extension Theorem. Formally,

**Proof.** Assume first that $\sigma$ is Bayesian. Let $\mu$ be a probability measure over $K \times S^\infty$ such that (2) is satisfied.

For every $n$ and every $s_1, \ldots, s_{n+1}$ we have, by the definition of conditional probability in (1),

$$\mu(B) = \sum_{s_1, \ldots, s_{n+1}} \mu(s_1, \ldots, s_{n+1}) \mu(B | s_1, \ldots, s_{n+1}) =$$
$$\sum_{\substack{s_1, \ldots, s_n \\ \mu(s_1, \ldots, s_n) \neq 0}} \mu(s_1, \ldots, s_n) \sum_{s_{n+1}} \frac{\mu(s_1, \ldots, s_{n+1})}{\mu(s_1, \ldots, s_n)} \mu(B | s_1, \ldots, s_{n+1}).$$

From the uniqueness of conditional probabilities it follows that

$$\mu(B | s_1, \ldots, s_n) = \sum_{s_{n+1}} \frac{\mu(s_1, \ldots, s_{n+1})}{\mu(s_1, \ldots, s_n)} \mu(B | s_1, \ldots, s_{n+1}).$$

for every $s_1, \dots, s_n \in S$ such that $\mu(s_1, \dots, s_n) > 0$. From the last equation and (2) it follows that

$$\sigma[s_1, \dots, s_n] = \sum_{s_{n+1}} \frac{\mu(s_1, \dots, s_{n+1})}{\mu(s_1, \dots, s_n)} \sigma[s_1, \dots, s_{n+1}], \tag{4}$$

In particular,

$$\sigma[s_1, \dots, s_n] \in \mathrm{Conv}\{\sigma[s_1, \dots, s_n, s_{n+1}] | s_{n+1} \in S\},$$

as desired.

Assume now that $\sigma$ satisfies the condition of Theorem 1. Thus, there exist non-negative numbers $\lambda(s_{n+1}; s_1, \dots, s_n)$ such that, for every $s_1, \dots, s_n$,

$$\sum_{s_{n+1}} \lambda(s_{n+1}; s_1, \dots, s_n) = 1,$$

and

$$\sigma[s_1, \dots, s_n] = \sum_{s_{n+1}} \lambda(s_{n+1}; s_1, \dots, s_n) \cdot \sigma[s_1, \dots, s_n, s_{n+1}]. \tag{5}$$

We now define, for every $n$, a probability measure $\mu_n$ over $K \times S^n$ as follows

$$\mu_n \left( B \times \{(s_1, \dots, s_n)\} \right) = \lambda(s_1; ) \cdot \lambda(s_2; s_1) \cdot \dots \cdot \lambda(s_n; s_1, \dots, s_{n-1}) \cdot \sigma[s_1, \dots, s_n](B), \tag{6}$$

for every Borel subset $B$ of $K$ and every $s_1, \dots, s_n \in S$.

It follows that for every $s_1, \dots, s_n \in S$ and every Borel subset $B$ of $K$ one has

$$\begin{aligned}
&\sum_{s_{n+1}} \mu_{n+1} \left( B \times \{(s_1, \dots, s_{n+1})\} \right) = \\
&\sum_{s_{n+1}} \lambda(s_1; ) \cdot \lambda(s_2; s_1) \cdot \dots \cdot \lambda(s_{n+1}; s_1, \dots, s_n) \cdot \sigma[s_1, \dots, s_{n+1}](B) = \\
&\lambda(s_1; ) \cdot \lambda(s_2; s_1) \cdot \dots \cdot \lambda(s_n; s_1, \dots, s_{n-1}) \cdot \sigma[s_1, \dots, s_n](B) = \\
&\mu_n \left( B \times \{(s_1, \dots, s_n)\} \right),
\end{aligned}$$

where the first and third equalities follows from (6) and the second equality follows from (5). Therefore, the marginal distribution of $\mu_{n+1}$ over $K \times S^n$ is $\mu_n$. It follows from Kolmogorov's Extension Theorem that there exists a probability $\mu$ over $K \times S^\infty$ such that the marginal of $\mu$ over $K \times S^n$ is $\mu_n$. In particular, substituting $B = K$ in (6) we get that the marginal of $\mu$ over $S^n$ is given by

$$\mu(s_1, \ldots, s_n) = \lambda(s_1;) \cdot \lambda(s_2; s_1) \cdot \cdots \cdot \lambda(s_n; s_1, \ldots, s_{n-1}). \tag{7}$$

It follows that

$$\mu_K(B) = \mu(B \times S^n) = \sum_{s_1,\ldots,s_n} \mu\left(B \times \{(s_1, \ldots, s_n)\}\right) = \sum_{s_1,\ldots,s_n} \mu(s_1, \ldots, s_n)\sigma[s_1, \ldots, s_n](B),$$

where the second equality follows from (6) and (7), and the fact that $\mu_n$ is the marginal of $\mu$ over $K \times S^n$. By the uniqueness of the conditional probabilities it follows that

$$\sigma[s_1, \ldots, s_n](B) = \mu(B|s_1, \ldots, s_n)$$

as desired. ■

As can be expected, the identified distribution (formally, $\mu$) is not necessarily determined uniquely. Our result addresses the broader question regarding whether, observing agents' predictions, we can reject the hypothesis of them updating using Bayes rule.

We will refer to an updating rule $\sigma : S^* \to \Delta(K)$ as *sound* if it satisfies (3) for all $s_1, \ldots, s_n \in S$. Theorem 1 implies that a sound updating rule is Bayesian.

**Observing Actions and Dynamic Consistency.** An action can be thought of as a bounded Borel function $u : K \to R$.[4] Every probability measure $\pi \in \Delta(K)$ induces a preference relation $\preceq_\pi$ over actions: $u \preceq_\pi v$ when $\int u \, \mathrm{d}\pi \leqslant \int v \, \mathrm{d}\pi$ for every pair $u, v$ of actions. By the separation theorem, the condition in Theorem 1 can equivalently be stated in the following way: For every pair $u, v$ of actions, if $u \preceq_{\sigma[\alpha * s_{n+1}]} v$ for every $s_{n+1} \in S$ then $u \preceq_{\sigma[\alpha]} v$: If, whatever signal will

[4]Thus, we identify actions with von-Neumann Morgenstern utility functions over the set of states of nature.

arrive tomorrow, the agent will prefer action $u$ to $v$ then she prefers action $u$ to $v$ today.

That is, if the decision maker knows that given additional information tomorrow she will prefer action $u$ to $v$ regardless of what that information will be, then she prefers $u$ to $v$ today. This condition is reminiscent of dynamic consistency (see, e.g., Ghirardato, 2002, and references therein). Note, however, that while the soundness condition is expressed solely in terms of conditional beliefs (and, by extension, conditional preferences over actions), dynamic consistency is a property of a preference relation over contingent plans, which is not determined by an updating rule.

The soundness condition is weaker then dynamic consistency: if the order induced by a prediction $\sigma$ over actions can be extended to an order over contingent plans that satisfies dynamic consistency then $\sigma$ must satisfy the soundness condition. By Theorem 1, the converse is also true. However, as we show in Section 8 below, in a more general setup in which predictions are event-based, dynamic consistency is a strictly stronger condition.[5]

**Excess Stickiness.** It is interesting to note that Theorem 1 implies that a large class of updating rules that exhibit stickiness to prior beliefs (i.e., rules in which the reported beliefs are always tilted toward previous reports) is observationally equivalent to (or non-identifiable from) Bayesian updating. Indeed, consider an agent who holds a prior $\mu$ on $K \times S^{\infty}$ and provides reports $\tilde{\sigma}$ as follows:

$$\tilde{\sigma}[s_1, s_2, ..., s_n](k) = \alpha_n(s_1, ..., s_n)\mu(k|s_1, s_2, ..., s_n) + (1 - \alpha_n(s_1, ..., s_n))\,\tilde{\sigma}[s_1, ..., s_{n-1}](k),$$

where $\alpha_n(s_1, ..., s_n) \in [0, 1]$ for all $n$ and $s_1, ..., s_n \in S$.

That is, the agent weighs the correct posterior with her previous prediction. It is straightforward to show that $\tilde{\sigma}$ would, in fact, be sound. In particular, it would be indistinguishable from Bayesian updating (with a prior different than $\mu$).[6]

[5] In order not to distract from the main point of the paper, we do not formulate the notion of dynamic consistency explicitly. Further details can be provided from the authors upon request.

[6] If signals were assumed to be conditionally independent, an equally sensible model for stickiness would be one in which current belief reports are a convex combination of last period's reports and a Bayesian posterior based on these *reported beliefs* and the current signal. Such a model would generate predictions

**Learning.** It follows directly from the definition of Bayesian rules that $\sigma : S^\infty \to \Delta(K)$ is Bayesian if and only if there exists some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and random variables $\kappa : \Omega \to K$ and $\zeta_1, \zeta_2, \dots : \Omega \to S$ such that

$$\sigma[s_1, \dots, s_n](B) = \frac{\mathbb{P}(\kappa = k, \zeta_1 = s_1, \dots, \zeta_n = s_n)}{\mathbb{P}(\zeta_1 = s_1, \dots, \zeta_n = s_n)},$$

for every $s_1, \dots, s_n$ such that $\mathbb{P}(\zeta_1 = s_1, \dots, \zeta_n = s_n) > 0$. The probability measure $\mu$ appearing in our analysis above corresponds to the joint distribution of $\kappa, \zeta_1, \zeta_2, \dots$.

Note that in general we cannot assume that $\kappa$ is measurable with respect to $\zeta_1, \dots, \zeta_n$, i.e., that the agent is asked to provide her prediction about her future observation, since in this case the agent will eventually *learn* the state – that is, her predictions will converge to an extreme point of $\Delta(K)$. However, the soundness condition of Theorem 1 does not guarantee learning. As an example, consider the case in which $K = \{0, 1\}$, so that a prediction can be summarized by a number corresponding to the assessed probability that the state is 1. Assume that $\sigma[s_1, \dots, s_n](1) = 0.5$ for every $s_1, \dots, s_n$. Clearly $\sigma$ is sound, but no learning occurs.

By the Martingale Convergence Theorem, if $\sigma$ is Bayesian then almost surely $\sigma[s_1, \dots, s_n]$ is a convergent sequence – it converges to the agent's prediction "at infinity." However, these predictions need not be deterministic as shown by the above example.

**Experimentation.** If one entertains testing whether an individual is a Bayesian updater experimentally, one may find the prospects of presenting the potential subject with all signal sequences in $S^*$ daunting. An immediate response that follows from our main result is that an agent that is non-Bayesian will violate Theorem 1's condition of soundness for some $n$, and therefore be detected as non-Bayesian in finite time.

Beyond that, the construction in our proof translates directly when considering only finite sequences of signals bounded in length. Specifically, we have the following immediate corollary:

**Corollary (Finite Experimentation)** *Suppose we restrict the domain of $\sigma$ to $\bigcup_{n=0}^{T} S^n$ for some integer $T$. Then $\sigma$ is Bayesian if and only if the (soundness) condition in Theorem*

indistinguishable from Bayesian updating as well.

*1 is satisfied for* $0 \le n < T$. *The measure over* $K \times S^T$ *that explains* $\sigma$ *in this case is the measure* $\mu_T$ *constructed in Theorem 1's proof.*

In terms of the number of degrees of freedom pertaining to the objects analyzed in this paper, looking at finite experimentation is particularly useful. Consider an updating rule $\sigma$ that is defined over sequences of signals of length up to $T$, i.e., $\sigma : \bigcup_0^T S^n \to \Delta(K)$. Note that the number of degrees of freedom for specifying such $\sigma$ is $(|K|-1) \cdot \frac{|S|^{T+1}-1}{|S|-1}$. On the other hand, the number of degrees of freedom available for specifying a belief $\mu$ over $K \times S^T$ is $|K| \times |S|^T - 1$. If $|K| = |S|$ then the number of degrees of freedom for specifying $\mu$ equals the number of degrees of freedom for specifying $\sigma$. If $|K| > |S|$ then the set of Bayesian updating rules has smaller dimension than the set of all updating rules, which implies that the soundness condition we propose restricts the dimension of $\sigma$.

In the following section we consider a generalization of Theorem 1 regarding the case in which only partial (and arbitrary) predictions are observed.

## 4. Partial Observations

In many situations it may be difficult to access the full range of predictions corresponding to all signal sequences and one may observe an arbitrary set of predictions. Those are, of course, still subject to some consistency restrictions in order to be explained as the outcome of a Bayesian updating process. The focus of the current section is the identification of these general restrictions.

Formally, we now assume that $\sigma$ is defined over an arbitrary subset $A$ of $S^*$. Again, we ask whether $\sigma$ can be derived from a Bayesian updating rule. In other words, we inspect when $\sigma$ can be extended to $S^*$ in such a way as to satisfy the condition presented in Theorem 1.

It is useful to view $S^*$ as the set of nodes of an infinite tree: The root is the empty sequence, and the immediate successors of a node $(s_1, \ldots, s_n) \in S^*$ are all the elements $(s_1, \ldots, s_n, \tilde{s}_{n+1})$, where $\tilde{s}_{n+1} \in S$. A *sub-tree* is given by a pair $(r, L)$ where $r \in S^*$ and $L \subseteq S^*$ is a set of successors of $r$ such that any branch of $S^*$ that passes through $r$ passes through a unique

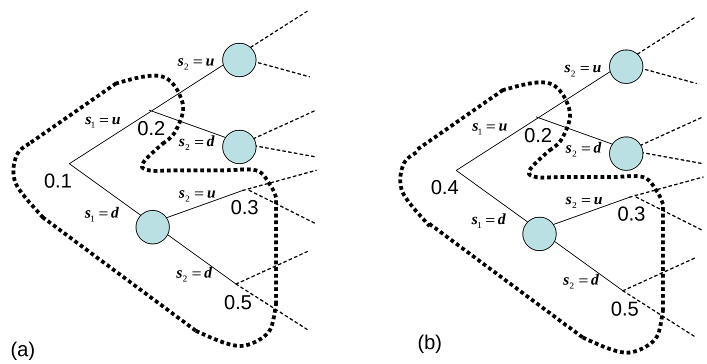


Figure 1: Consistency on Sub-trees

element of $L$. $r$ is the root of the sub-tree and $L$ is the set of its leaves. We say that a sub-tree $(r, L)$ is $A$-*subtree* if $r \in A$ and $L \subseteq A$.

The following simple example illustrates some of the notation and the condition required for consistency with Bayesian updating.

**Example 1** Consider the case in which $S = \{u, d\}$ and $K = \{0, 1\}$ so that a prediction can now be summarized by one number (say, the probability that the state is 1). Suppose further that predictions are available only for

$$A = \{\Phi, (s_1 = u), (s_1 = d, s_2 = u), (s_1 = d, s_2 = d)\}$$

(so, in particular, there is no prediction available for the revelation $s_1 = d$). Figure 1 illustrates two such scenarios, where numbers in each node correspond to the available predictions and shaded circles correspond to missing observations. The two panels differ in the reported prior probability (appearing in the root of the tree). In this example, the only non-trivial $A$-subtree is the one appearing within the dotted thick line.

Consider first panel (a). If we are to complete the observation regarding $s_1 = d$, from Theorem 1, it must be the case that this prediction, call it $x$, is in between 0.3 and 0.5. But then 0.1, the reported prior, cannot be in between 0.2 and $x$, and there would be no way in which to fill out the missing observation matching $s_1 = d$ and satisfy the restriction posited in Theorem 1. Note that this is, in fact, (indirectly) a consequence of 0.1 not being a convex combination of the predictions appearing in the leaves of the sub-tree: 0.2, 0.3, and 0.5.

In contrast, in panel (b) of the Figure, the prediction in the root, 0.4, is certainly a convex combination of those appearing in the leaves, and there are many ways in which to complete the sub-tree (for $s_1 = d$) in a consistent manner. In fact, any prediction within $[0.4, 0.5]$ would satisfy the conditions of Theorem 1.

The example generalizes directly. In fact, Theorem 2 illustrates that the necessary and sufficient condition for $\sigma$ to be extended to a Bayesian updating rule over $S^*$ is that for each $A$-subtree the prediction at the root is within the convex hull of the predictions in the leaves. Formally,

**Theorem 2 (Partial Observations)** *Let $K$ be a Borel set of states of nature, $S$ a finite set of signals, $A \subseteq S^*$ and $\sigma : A \to \Delta(K)$.*

1. *If $\sigma$ can be extended to a Bayesian updating rule over $S^*$ and the probability measure $\mu$ over $K \times S^\infty$ explains $\sigma$ then*

$$\sigma[r] \in \text{Conv}\{\sigma[l] | l \in L\} \tag{8}$$

*for every $A$-subtree $(r, L)$* and such that $\mu(s_1, \ldots, s_n) > 0$ where $r = (s_1, \ldots, s_n)$.

2. *If (8) is satisfied for every sub-tree $(r, L)$ that is contained in $A$ then $\sigma$ can be extended to a Bayesian updating rule over $S^*$.*

The formal proof of Theorem 2 appears in the Appendix. Intuitively, necessity of the condition follows immediately from the type of arguments used to prove Theorem 1. To prove sufficiency, we restrict $S^*$ to finite trees and consider all minimal sub-trees that are contained in $A$. These are sub-trees that contain no other sub-tree contained in $A$. Each such sub-tree can be completed in a consistent manner, as in the example. We then look at the resulting (extended) predictions and repeat the procedure until all finite predictions are completed in a manner consistent with Theorem 1's soundness condition. We can then use an extension to $S^*$ to utilize Theorem 1.

## 5. General Signal Spaces

Our analysis thus far allowed for arbitrary state spaces but restricted the signals to be taken from a finite set. In this section we extend our main result to general signal spaces. The difficulty arising from continuous signal spaces can be easily seen: The condition of Theorem 1 ensured that for any sequence of signals $(s_1, ..., s_n)$, the prediction is a convex combination of the predictions corresponding to all continuation signal sequences $(s_1, ..., s_n, \tilde{s}_{n+1})$. The weights placed on each such prediction, denoted by $\lambda(\tilde{s}_{n+1}; s_1, ..., s_n)$ in the proof of Theorem 1, were not determined uniquely. In order to repeat the construction of the original proof, and end up with an admissible measure, we need to make sure we *select* $\lambda(\tilde{s}_{n+1}; s_1, ..., s_n)$ in a way that makes it *measurable* with respect to all its arguments. When $S$ is discrete this is, of course, immediate as any selection will be measurable.

For a Borel space $X$ we denote by $\Delta(X)$ the Borel space of all Borel probability measures over $X$. For $\mu \in \Delta(\Delta(X))$, the *barycenter* of $\mu$ is the unique element $[\mu]$ of $\Delta(X)$ such that

$$\int h \, \mathrm{d}[\mu] = \int \left( \int h \, \mathrm{d}\omega \right) \mu(\mathrm{d}\omega)$$

for every bounded measurable function $h : X \to \mathbb{R}$. Note that if $\mu$ is a discrete probability, concentrated on a finite set of atoms, then the barycenter of $\mu$ is the weighted average of the atoms of $\mu$. The map $\mu \mapsto [\mu]$ is a Borel map from $\Delta(\Delta(X))$ to $\Delta(X)$. For a universally measurable subset[7] $A$ of $\Delta(X)$, we denote by $\mathrm{BC}(A)$ the set of all barycenters of Borel measures

[7] A subset $A$ of a Borel space $X$ is *universally measurable* if it is measurable with respect to the completions of all Borel probability measures on $X$.

over $\Delta(X)$ that are concentrated on $A$:

$$\mathrm{BC}(A) = \{[\mu] | \mu \in \Delta(\Delta(X)) \text{ and } \mu(A) = 1\}.$$

In the general setting, the barycenter of a set of distributions over $K$ will replace the convex hull in Theorem 1 (note that if $A$ is finite then $\mathrm{BC}(A)$ is the convex hull of $A$). The reason we define barycenters for universally measurable sets (and not just for Borel sets) is that the set $\{\sigma[s_1, \ldots, s_{n+1}] | s_{n+1} \in S\}$ is not necessarily a Borel set, even for a Borel function $\sigma$. However, as the image of a Borel space in another Borel space, this subset is *analytic*, and it is known that every analytic set is universally measurable.

We now turn to the formulation of the theorem for general signal spaces. An updating rule $\sigma : S^* \to \Delta(K)$ is called Borel if the restriction of $\sigma$ to $S^n$ is a Borel function for every $n$. It is *Bayesian* if there exists a probability measure $\mu$ over $K \times S^\infty$ such that $\sigma[s_1, \ldots, s_n]$ is the conditional distribution of $\mu$ given $s_1, \ldots, s_n$ for $\mu$-almost every $s_1, \ldots, s_n$. As in Theorem 1, for a Bayesian updating rule, it is immediate to show that the predictions at stage $n$ must be in the barycenter of those at stage $n+1$. The theorem illustrates the converse as well.

**Theorem 3 (General Spaces)** *Let $S, K$ be Borel spaces, and let $\sigma : S^* \to \Delta(K)$ be a Borel updating rule.*

1. *If $\sigma$ is Bayesian and $\mu$ explains $\sigma$ then*

$$\sigma[s_1, \ldots, s_n] \in \mathrm{BC}\left(\{\sigma[s_1, \ldots, s_n, s_{n+1}] | s_{n+1} \in S\}\right) \text{ almost-surely.}$$

2. *If, for every $s_1, \ldots, s_n \in S$,*

$$\sigma[s_1, \ldots, s_n] \in \mathrm{BC}\left(\{\sigma[s_1, \ldots, s_n, s_{n+1}] | s_{n+1} \in S\}\right)$$

*then $\sigma$ is Bayesian.*

The proof (in the Appendix) relies on results on selection of measurable functions.

## 6. Plausibility of Signal Sequences

In principle, our main result does not rule out situations in which the derived distribution $\mu$ places zero probability on some sequences of signals (that our hypothetical subject may face in the experimental questionnaire). Indeed, consider the following example:

**Example 2** Assume that $K = S = \{0, 1\}$. Then $\Delta(K)$ can be identified with the interval $[0, 1]$, where an element $p \in [0, 1]$ stands for the probability measure over $K$ that assigns probability $1 - p$ to 0 and probability $p$ to 1. Consider the updating rule $\sigma$ given by

$$\sigma[s_1, \ldots, s_n](1) = \frac{s_1 + \cdots + s_n}{n}, \text{ and } \sigma[](1) = \frac{1}{2}.$$

It is easy to verify that it satisfies the condition in Theorem 1. The corresponding probability measure $\mu$ over $K \times S^\infty$ is an atomic probability with two mass-$\frac{1}{2}$ atoms, one at $(0, 0, 0, \ldots)$ and one at $(1, 1, 1, \ldots)$: If the state of nature is 0 the agent expects to receive with probability 1 the signal 0 at every stage, and if the state of nature is 1 the agent expects to receive with probability 1 the signal 1 at every stage. The values of $\sigma[s_1, \ldots, s_n]$ for a sequence of non-identical signals $s_1, \ldots, s_n$ is irrelevant, because according to $\mu$ such a sequence has probability 0.

In order to guarantee that every finite sequence $\alpha \in S^*$ has a strictly positive probability, we need to ensure that the numbers $\lambda(s_{n+1}|s_1, \ldots, s_n)$ are strictly positive (recall (7)). To achieve this, we have to replace the Convex Hull in Theorem 1 with its relative interior. Recall that the *relative interior* ri $X$ of a convex set $X$ is its interior within its affine hull (the minimal set containing all affine combinations of points in $X$). We will need the following lemma (the short and technical proof of which appears in the Appendix):

**Lemma (Relative Interior)** *Let $V$ be some real vector space, $S$ be finite set and, for every $s \in S$, let $x_s \in V$. Then*

$$\operatorname{ri}\operatorname{Conv}(\{x_s | s \in S\}) = \left\{ \sum_{s \in S} \lambda_s x_s | \lambda_s > 0 \textit{ for every } s \in S \textit{ and } \sum_{s \in S} \lambda_s = 1 \right\}.$$

The Lemma combined with the construction of Theorem 1's proof assure that when the convex hull in the soundness condition is replaced with its relative interior, the corresponding updating rule is Bayesian and can be explained as arising from Bayesian updating of a belief placing strictly positive probabilities for any finite signal sequence (indeed, the coefficients $\lambda(s_{n+1}; s_1, \ldots, s_n)$ in the proof of Theorem 1 are strictly positive). Formally,

**Corollary** *Let $K$ be a Borel set of states of nature, $S$ a finite set of signals and $\sigma : S^* \to \Delta(K)$ an updating rule. Then the following conditions are equivalent:*

1. *There exists a probability measure $\mu$ over $K \times S^\infty$ that explains $\sigma$ and such that $\mu(s_1, \ldots, s_n) > 0$ for every $s_1, \ldots, s_n \in S$.*
2. $$\sigma[s_1, \ldots, s_n] \in \text{ri Conv}\{\sigma[s_1, \ldots, s_n, s_{n+1}] | s_{n+1} \in S\} \tag{9}$$

## 7. Independence and Exchangeability

So far, we have placed no restrictions on the admissibility of beliefs that make an agent appear Bayesian. In particular, elicited beliefs may exhibit correlations between signals (even conditional on states), may depend on the order of signals, etc. In this section, we identify additional restrictions on updating rules that guarantee some natural attributes of the elicited belief over $K \times S^\infty$.

We start with the notion of conditionally independent and identically distributed (i.i.d.) signals. In a wide array of economic models, ranging from strategic voting to private value auctions, private signals are often assumed to be conditionally independent. Furthermore, in the statistics literature, the term Bayesian is frequently used to indicate conditional independence, the idea being that the sequence of signals $s_1, s_2, \ldots$ are samples from some distribution that depends on the state, or parameter, $k$. The statistician has some prior belief over the true parameter that governs the distribution of the sample, and she updates her belief given the observations (much like questions leading to maximum likelihood methods in econometrics, see, e.g., Greene, 1993).

In analogy to our original definitions, we say that an updating rule $\sigma : S^* \to \Delta(K)$ is *conditionally i.i.d. Bayesian* if there exists a probability measure $\mu$ over $K \times S^\infty$ such that, for every $n$, signals $s_1, s_2, ...$ are conditionally i.i.d. given the state of nature $k$ and for every Borel subset $B$ of $K$,

$$\sigma[s_1, \ldots, s_n](B) = \mu(B|s_1, \ldots, s_n).$$

As the following example illustrates, the restrictions of Theorem 1 do not generally assure that an updating rule that is Bayesian is conditionally i.i.d.

**Example** Assume that $K = S = \{0, 1\}$. Then $\Delta(K)$ can be identified with the interval $[0, 1]$, where an element $p \in [0, 1]$ stands for the probability measure over $K$ that assigns probability $1 - p$ to 0 and probability $p$ to 1. Consider the updating rule $\sigma$ given by

$$\sigma\left[s_1, \ldots, s_n\right](1) = \begin{cases} 1/(n+2), & \text{if } s_1 + \cdots + s_n \text{ is even} \\ 1 - 1/(n+2), & \text{if } s_1 + \cdots + s_n \text{ is odd.} \end{cases}$$

In particular, $\sigma\left[\right](1) = 1/2$. This updating rule is such that a signal '0' strengthens the agent's previous opinion about the state of nature, whereas a signal '1' changes it drastically in the other direction. A Bayesian statistician, who believes the signals to be i.i.d. cannot exhibit such behavior. However, $\sigma$ clearly satisfies the condition of Theorem 1 and is therefore Bayesian according to our definition.

Naturally, the number of degrees of freedom allowed by a conditionally i.i.d. signal generating process is lower than the number of degrees of freedom corresponding to an arbitrary signal generation process. Indeed, recall that the number of degrees of freedom available for specifying a belief $\mu$ over $K \times S^T$ is $|K| \times |S|^T - 1$. However, the number of (non-linear) degrees of freedom corresponding to a measure $\mu$ that corresponds to i.i.d. signals given the state of nature (according to the statistics paradigm) is $(|K| - 1) + |K| \times (|S| - 1)$, which is much smaller.

Requiring an updating rule to be derived from a belief that signals are conditionally i.i.d. translates directly into a formula derived from Bayes rule. For the sake of illustration, suppose $K = S = \{0, 1\}$. A conditional i.i.d. signal generation process is tantamount to the underlying belief $\mu$ over $K \times S^\infty$ being characterized by three parameters: $q, p_0$, and $p_1$, where $\mu_K(1) = q$, and the probability of $s_i = 1$ conditional on $k = 1$ (or $k = 0$) is $p_1$ (or $p_0$), for every $i$. An updating rule $\sigma : S^* \to \Delta(K)$ is conditionally i.i.d. Bayesian with belief $\mu$ if and only if:

$$\sigma\left[s_1, ..., s_n\right](1) = \frac{qp_1^{\bar{s}}\left(1 - p_1\right)^{n-\bar{s}}}{qp_1^{\bar{s}}\left(1 - p_1\right)^{n-\bar{s}} + (1 - q)\, p_0^{\bar{s}}\left(1 - p_0\right)^{n-\bar{s}}}, \tag{10}$$

where $\bar{s} = \sum_{i=1}^n s_i$, for some $q, p_0$, and $p_1$.

For any updating rule $\sigma : S^* \to \Delta(K)$, denote by $L(s_1, ..., s_n) = \ln \frac{\sigma[s_1,...,s_n](1)}{\sigma[s_1,...,s_n](0)}$ the log-likelihood corresponding to reported beliefs.

It follows from (10) that conditional independence with parameters $q, p_0$, and $p_1$ translates into:

$$L(s_1, ...s_n) = \ln \frac{q}{1-q} + \left(\sum_{i=1}^n s_i\right) \ln \frac{p_1}{p_0} + \left(n - \sum_{i=1}^n s_i\right) \ln \frac{1 - p_1}{1 - p_0}. \tag{11}$$

This allows us to derive a necessary and sufficient condition for an updating rule to be conditionally i.i.d. Bayesian.

**Corollary (Conditionally i.i.d. Beliefs)** *Suppose $K = S = \{0, 1\}$. An updating rule $\sigma : S^* \to \Delta(K)$ is conditionally i.i.d. Bayesian if and only if there exist $a, b \in R$, $sign(a)sign(b) = -1$, such that*

$$\begin{aligned} L(s_1, ..., s_{n+1}) - L(r_1, ..., r_n) &= a \quad \textit{whenever} \quad \sum_{i=1}^{n+1} s_i = \sum_{i=1}^n r_i, \ \textit{and} \\ L(s_1, ..., s_n) - L(r_1, ..., r_n) &= b - a \quad \textit{whenever} \quad \sum_{i=1}^{n} s_i = \sum_{i=1}^n r_i + 1. \end{aligned} \tag{12}$$

Indeed, (11) and (12) are equivalent with

$$L() = \ln \frac{q}{1-q}, \quad a = \ln \frac{1 - p_1}{1 - p_0}, \quad b = \ln \frac{p_1}{p_0}.$$

The class of i.i.d. processes is clearly rather restrictive. One natural extension is to the class of exchangeable processes. These are processes in which the joint distribution of each set of signals does not depend on the order at which they arrive (see Feller, 1966). Formally,

**Definition (Exchangeability)** *A measure $\mu$ over $K \times S^{\infty}$ is called* conditionally exchangeable *if for any $k \in K$, for any $\acute{\imath}_1, .., i_m$ and permutation $\pi : \{i_1, ...i_m\} \to \{i_1, ...i_m\}$, and any $s, r \in S^{\infty}$ such that for all $i \notin \{i_1, ...i_m\}$, $s_i = r_i$ and for all $j = 1, ..., m$, $r_{i_j} = s_{\pi(i_j)}$, $\mu(k, s) = \mu(k, r)$.*

In keeping with terminology, we say that an updating rule $\sigma : S^* \to \Delta(K)$ is *conditionally exchangeable Bayesian* if there exists a conditionally exchangeable probability measure $\mu$ over $K \times S^{\infty}$ such that, for every $n$ and every Borel subset $B$ of $K$,

$$\sigma[s_1, \ldots, s_n](B) = \mu(B|s_1, \ldots, s_n).$$

In the remains of the section, we concentrate on the case of $K = S = \{0, 1\}$ and identify the restrictions required of updating rules to be exchangeable Bayesian.

For any updating rule $\sigma : S^* \to \Delta(K)$, denote by $\tau(s_1, ..., s_n) = \sigma[s_1, ..., s_n](1)$. For simplicity, we will make the additional assumption that $\tau(s_1, \ldots, s_n, s_{n+1}) \neq \tau(s_1, \ldots, s_n, 1 - s_{n+1})$ for every $s_1, \ldots, s_{n+1} \in S$. This assumption, implying that every signal is informative, simplifies the computations, as it assures that the coefficients appearing in the proof of Theorem 1 are determined uniquely. That is, for a convex hull condition of the form:

$$\tau(s_1, ..., s_n) = \lambda(s_{n+1}; s_1, ..., s_n)\tau(s_1, ..., s_n, s_{n+1}) + \lambda(1 - s_{n+1}; s_1, ..., s_n)\tau(s_1, ..., s_n, 1 - s_{n+1}),$$

where $\lambda(s_{n+1}; s_1, ..., s_n) + \lambda(1 - s_{n+1}; s_1, ..., s_n) = 1$, it follows that

$$\lambda(s_{n+1}; s_1, ..., s_n) = \frac{\tau(s_1, ..., s_n) - \tau(s_1, ..., s_n, 1 - s_{n+1})}{\tau(s_1, ..., s_n, s_{n+1}) - \tau(s_1, ..., s_n, 1 - s_{n+1})}.$$

Consider two signals $s_1, s_2$. We can use the construction of Theorem 1 (as shown, uniquely defined in the binary case) to derive conditions assuring that the underlying belief $\mu$ satisfies

$\mu(k, s_1, s_2) = \mu(k, s_2, s_1)$ for $k = 0, 1$. These translate into:

$$\frac{\tau(\cdot) - \tau(1)}{\tau(0) - \tau(1)} \cdot \frac{\tau(0) - \tau(0,0)}{\tau(0,1) - \tau(0,0)} = \frac{\tau(\cdot) - \tau(0)}{\tau(1) - \tau(0)} \cdot \frac{\tau(1) - \tau(1,1)}{\tau(1,0) - \tau(1,1)}.$$

Note that while exchangeability certainly requires that $\tau(s_1, s_2) = \tau(s_2, s_1)$, the above conditions illustrates that this is not sufficient. For example, suppose $\tau(\cdot) = 0.5$, $\tau(0) = 0.4$, $\tau(1) = 0.6$, $\tau(0,1) = \tau(1,0) = 0.5$, $\tau(0,0) = 0.3$, and $\tau(1,1) = 0.9$. While these predictions are consistent with Bayesian updating, and satisfy $\tau(s_1, s_2) = \tau(s_2, s_1)$ for all $s_1, s_2$, they are not consistent with exchangeable Bayesian updating.

The condition above can directly be generalized to any sequence of signals, and we get the following corollary.

**Corollary (Exchangeable Beliefs)** *Suppose $K = S = \{0, 1\}$. An updating rule $\sigma : S^* \to \Delta(K)$ is exchangeable Bayesian if and only if for every sequences $(s_1, \ldots, s_n)$ and $(r_1, \ldots, r_n)$ of signals such that $\sum_{i=1}^n s_i = \sum_{i=1}^n r_i$ one has*

$\tau(s_1, ..., s_n) = \tau(r_1, ..., r_n)$, *and*

$$\prod_{i=1}^{n} \frac{\tau(s_1, ..., s_{i-1}) - \tau(s_1, ..., s_{i-1}, 1 - s_i)}{\tau(s_1, ..., s_{i-1}, s_i) - \tau(s_1, ..., s_{i-1}, 1 - s_i)} = \prod_{i=1}^{n} \frac{\tau(r_1, ..., r_{i-1}) - \tau(r_1, ..., r_{i-1}, 1 - r_i)}{\tau(r_1, ..., r_{i-1}, r_i) - \tau(r_1, ..., r_{i-1}, 1 - r_i)}.$$

## 8. Event-Based Predictions

One natural direction for generalization of our analysis could come from discarding the sequencing aspect inherent in our model of signal observations. That is, one could contemplate predictions that are based on arbitrary subsets of some underlying space representing the information the agent can receive. As we show in this section, a condition analogous to the soundness conditions we identified in the signaling model is necessary but not sufficient in such a set formulation of the problem.

Formally, let $P$ be a Borel space of *outcomes* and let $\mathcal{I}$ be a family of *events*, i.e., Borel subsets of $P$. Assume that we are given a function $\sigma : \mathcal{I} \to \Delta(K)$: Fore every event $\alpha \in \mathcal{I}$,

$\sigma(\alpha)$ is the agent's updated belief about the state of nature after he learns that the event $\alpha$ was obtained. As before, for $\alpha \in \mathcal{I}$, we denote by $\sigma[\alpha]$ the image of $\alpha$ under $\sigma$. We say that $\sigma$ is *Bayesian* if there exists some probability distribution $\mu$ over $K \times P$ such that

$$\sigma[\alpha](C) = \frac{\mu(C \times \alpha)}{\mu_P(\alpha)}$$

for every $\alpha \in \mathcal{I}$ and every Borel subset $C$ of $K$, where $\mu_P$ is the marginal of $\mu$ over $P$. Note that the right hand side of the last equation is the conditional probability (under $\mu$) of $C$ given $\alpha$. The model of Section 2 is a special case, where $P = S^\infty$ and $\mathcal{I}$ is the set of all cylinders of the form $(s_1, \ldots, s_n) \times S^\infty$ for some $s_1, \ldots, s_n \in S^\infty$. In the general setting of this section we do not assume that the information is provided through a sequence of signals.

Let $\alpha \in \mathcal{I}$. By a *partition of* $\alpha$ we mean mutually disjoint subsets $(\alpha_1, \ldots, \alpha_n)$ of $\mathcal{I}$ such that $\alpha = \bigcup_i \alpha_i$. $\alpha_i$ are called *atoms* of the partition. The soundness condition in Theorem 1 can be adapted to a necessary condition for $\sigma$ to be rationalizable by $\mu$ in the general model: If $(\alpha_1, \ldots, \alpha_n)$ is a partition of $\alpha \in \mathcal{I}$ then $\sigma[\alpha] \in \mathrm{Conv}\{\sigma[\alpha_i] | 1 \leq i \leq n\}$ (provided that $\mu_P(\alpha) > 0$).
However, without further assumptions on $\mathcal{I}$ the condition is not sufficient, as shown by the following example:

**Example** Let $K = \{0, 1\}$ so that a prediction can be summarized by a number (the probability that the state is 1) and let $P = \{a, b, c\}$. Assume that

$$\begin{aligned} \sigma[\{a\}] &= 0.1, \sigma[\{b\}] = 0.3, \sigma[\{c\}] = 0.5, \text{ and} \\ \sigma[\{a,b\}] &= 0.2, \sigma[\{b,c\}] = 0.4, \sigma[\{a,c\}] = 0.2, \sigma[\{a,b,c\}] = 0.25. \end{aligned}$$

The reader can verify that these predictions satisfy the above conditions. However, they cannot be rationalized by any $\mu$ as the following argument suggests:

$$\begin{array}{llll} \sigma[\{a\}] = 0.1, & \sigma[\{b\}] = 0.3, & \sigma[\{a,b\}] = 0.2 & \Rightarrow \mu_P(a) = \mu_P(b). \\ \sigma[\{b\}] = 0.3, & \sigma[\{c\}] = 0.5, & \sigma[\{b,c\}] = 0.4 & \Rightarrow \mu_P(b) = \mu_P(c). \\ \sigma[\{a\}] = 0.1, & \sigma[\{c\}] = 0.5 & \sigma[\{a,c\}] = 0.2 & \Rightarrow \mu_P(a) = 3\mu_P(c). \end{array}$$

As can be seen from the example, for any partition $(\alpha_1, \ldots, \alpha_n)$ of $\alpha$, $\sigma[\alpha], \sigma[\alpha_1], \ldots, \sigma[\alpha_n]$ restricts the set of possible quotients between the probabilities $\mu_P(\alpha_i)$ of the atoms, and restrictions that are derived from different partitions can contradict one another.

Consider the directed graph whose nodes are the elements of $\mathcal{I}$ with an arrow from $\beta \in \mathcal{I}$ to $\alpha \in \mathcal{I}$ if $\alpha \subsetneq \beta$ and there exists no $\gamma \in \mathcal{I}$ such that $\alpha \subsetneq \gamma \subsetneq \beta$. If the graph is a *forest* (i.e., has no cycles) then it is a union of *trees*, and we can carry out the proof of Theorem 1 going from the root to the leaves on each tree. One can verify that the graph being a forest is equivalent to the condition that for every $\alpha_1, \alpha_2 \in \mathcal{I}$, either $\alpha_1 \subseteq \alpha_2$ or $\alpha_2 \subseteq \alpha_1$ or $\alpha_1 \cap \alpha_2 = \varnothing$. Note that in the special case of Section 2 the graph is a tree and the condition is satisfied.

## 9. Conclusions

The paper identified a simple condition for an updating rule to be indistinguishable from Bayesian updating. The essence of the condition pertains to the effects of additional information. It is simple enough to identify a non-Bayesian updater in finite time. It is general enough to encompass environments with arbitrary state and signal spaces. It also extends nicely to situations in which only partial responses to information are observable.

We also inspected properties of the statistical theories that explain certain updating rules as Bayesian. In particular, we identified additional restrictions on updating rules that assure such theories place strictly positive probability on any finite signal sequence, or that the signal generation process is conditionally i.i.d. or exchangeable.

## REFERENCES

[1] Charalambos D. Aliprantis and Kim C. Border (2006), *Infinite Dimensional Analysis A Hitchhiker's Guide*, Springer-Verlag, Third Edition.

[2] Francis J. Anscombe and Robert .J. Aumann (1963), "A Definition of Subjective Probability," *Annals of Mathematical Statistics*, Volume 34, pages 199-205.

[3] Antoine Billot, Itzhak Gilboa, Dov Samet, and David Schmeidler (2005), "Probabilities as Similarity-Weighted Frequencies," *Econometrica*, Volume 73(4), pages 1125-1136.

[4] Bruno de Finetti (1937), "La Prevision: Ses Lois Logiques, ses Sources Subjectives," *Annales de l'Institut Henri Poincare*, Volume 7, pages 1-68. Translated by Kyburg, H. E. Jr., "Foresight: Its Logical Laws, its Subjective Sources," in Kyburg, H. E. Jr. and Smokler, H. E., eds. (1964), *Studies in Subjective Probability*, Wiley: New York, 53-118.

[5] Donato M. Cifarelli and Eugenio Regazzini (1996), "De Finetti's Contribution to Probability and Statistics," *Statistical Science*, Volume 11(4), pages 253-282.

[6] Mahmoud A. El-Gamal and David M. Grether (1995), "Are People Bayesian? Uncovering Behavioral Strategies," *Journal of the American Statistical Association*, Volume 90(432), pages 1137-1145.

[7] Larry Epstein, Jawaad Noor, and Alvaro Sandroni (2006), "Non-Bayesian Updating: A Theoretical Framework," mimeo.

[8] William Feller (1966), *An Introduction to Probability Theory and Its Applications*, Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons.

[9] Itzhak Gilboa and Ehud Lehrer (1991), "The Value of Information - An Axiomatic Approach," *Journal of Mathematical Economics*, Volume 20(5), pages 443-459.

[10] Paolo Ghirardato (2002), "Revisiting Savage in a Conditional World ," *Economic Theory*, Volume 20(1), pages 83-92.

[11] Edward J. Green and In-Uck Park (1994), "Bayes Contingent Plans," http://ideas.repec.org/p/wpa/wuwpga/9307002.html

[12] William H. Greene (1993), *Econometric Analysis*, Prentice Hall International Editions.

[13] Erwin Klein and Anthony C. Thompson (1984), *Theory of Correspondences: including applications to mathematical economics.*

[14] Offerman, Theo, Joep Sonnemans, Gijs van de Kuilen, and Peter Wakker (2006), "Correcting Proper Scoring Rules for Risk Attitudes," mimeo.

[15] Frank P. Ramsey (1931), *The Foundations of Mathematics*, Routledge and Kegan Paul Ltd.

[16] Ralph T. Rockafellar (1996), *Convex Analysis*, Princeton University Press.

[17] Leonard J. Savage (1954), *The Foundations of Statistics*, Wiley.

Appendix

**Proof of Theorem 2.**

The *depth* of a node $(s_1, \dots, s_n) \in S^*$ is given by $d(s_1, \dots, s_n) = n$. The *diameter* of a subtree $(r, L)$ is given by $\mathrm{diam}(r, L) = \max\{d(l) - d(r) | l \in L\}$. For nodes $u, v \in S^*$ we denote $u \prec v$ if $u$ is a predecessor of $v$. If $A \subseteq S^*$ then an $A$-subtree $(r, L)$ is $A$-*minimal* if $\mathrm{diam}(r, L) > 0$ and there exists no $l \in L$ and $u \in A$ such that $r \prec u \prec l$. Note that every node can be the root of at most one $A$-minimal subtree and a leaf of at most one $A$-minimal subtree.

Assume first that $\sigma : S^* \to \Delta(K)$ is a Bayesian updating rule and let $(r, L)$ be an $A$-subtree and such that $\mu(s_1, \dots, s_n) > 0$ where $r = (s_1, \dots, s_n)$. We prove by indunction over $\mathrm{diam}(r, L)$ that (8) is satisfied. If $\mathrm{diam}(r, L) = 0$ then $L = \{r\}$ and (8) holds. Assume that $\mathrm{diam}(r, L) > 0$. Since $\mu$ explains $\sigma$ and $\mu(s_1, \dots, s_n) > 0$ it follows from (4) in the proof of Theorem 1 that

$$\sigma[s_1, \dots, s_n] \in \mathrm{Conv}\{\sigma[s_1, \dots, s_{n+1}] | s_{n+1} \in S \text{ and } \mu(s_1, \dots, s_{n+1}) > 0\}. \tag{13}$$

Fix $s_{n+1} \in S$ such that $\mu(s_1, \dots, s_{n+1}) > 0$, let $r' = (s_1, \dots, s_{n+1})$ and let $L' \subseteq L$ be the set of successors of $r'$ in $L$. By the induction hypothesis over $(r', L')$ it follows that

$$\sigma[s_1, \dots, s_{n+1}] \in \mathrm{Conv}\{\sigma[l] | l \in L'\} \subseteq \mathrm{Conv}\{\sigma[l] | l \in L\}.$$

Since this equation holds for every $s_{n+1} \in S$ such that $\mu(s_1, \dots, s_{n+1}) > 0$ it follows from (13) that (8) holds.

Assume now that $A \subseteq S^\infty$ and that $\sigma : A \to \Delta(K)$ satisfies (8) for every $A$-subtree $(r, L)$, in particular (8) is satisfied for every $A$-minimal subtree. Let $u \in S^*$ be a node such that $u \notin A$, and let $A' = A \cup \{u\}$. We claim that there exists and extension of $\sigma$ to $A'$ such that (8) is still satisfied for every $A'$-minimal subtree. Note that every $A'$-subtree that is not an $A$-subtree must contain $u$ either as a node or as a leaf. We distinguish between four cases

*Case 1: $u$ is the root of some $A'$-subtree and the leaf of some $A'$-subtree.* Let $L \subseteq A$ be such that $(u, L)$ is the $A'$-minimal subtree with root $u$ and let $r \in A$ and $M \subseteq A$ be such that $(r, M \cup \{u\})$ is the $A'$-minimal subtree with leaf $u$. In particular $(r, L \cup M)$ is $A$-minimal and therefore

$$\sigma[r] \in \mathrm{Conv}\left(\{\sigma[l] | l \in L\} \cup \{\sigma[m] | m \in M\}\right),$$

so that $\sigma[r] = \sum_{l \in L} \lambda_l \sigma[l] + \sum_{m \in M} \mu_m \sigma[m]$ for some $\lambda_l, \mu_m \geq 0$ such that $\sum_{l \in L} \lambda_l + \sum_{m \in M} \mu_m = 1$. Let $\sigma[u] = \left(\sum_{l \in L} \lambda_l \sigma[l]\right) / \left(\sum_{l \in L} \lambda_l\right)$ if $\sum_{l \in L} \lambda_l > 0$ and choose $\sigma[u] \in \mathrm{Conv}\left(\{\sigma[l] | l \in L\}\right)$ arbitrarily if $\sum_{l \in L} \lambda_l = 0$. Then (8) is satisfied over both $(u, L)$ and $(r, M \cup \{u\})$, as desired.

*Case 2: $u$ is the root of some $A'$-subtree and the leaf of no $A'$-subtree.* Let $L \subseteq A$ be such that $(u, L)$ is the $A'$-minimal subtree with root $u$. Define $\sigma[u] \in \mathrm{Conv}\left(\{\sigma[l] | l \in L\}\right)$ arbitrarily. Then (8) is satisfied over $(u, L)$.

*Case 3: $u$ is the root of no $A'$-subtree and the leaf of some $A'$-subtree.* Let $r \in A$ and $M \subseteq A$ be such that $(r, M \cup \{u\})$ is the $A'$-minimal subtree with leaf $u$. Choose $\sigma[u] = \sigma[r]$. Then (8) is satisfied over $(r, M \cup \{u\})$, as desired.

*Case 4: $u$ is the root of no $A'$-subtree and the leaf of no $A'$-subtree* Define $\sigma[u] \in \Delta(K)$ arbitrarily.

Adding nodes one after another, we arrive at an updating rule $\sigma : S^* \to \Delta(K)$ that satisfies (8) over all $S^*$-minimal subtrees. But since $S^*$-minimal subtrees are of the form $(r, L)$ for $r = (s_1, \ldots, s_n)$ and $L = \{(s_1, \ldots, s_n, s_{n+1}) | s_{n+1} \in S\}$ it follows that $\sigma$ satisfies (3) and therefore, by Theorem 1, Bayesian. ■

**Proof of Theorem 3.**

The proof requires the following measurable selection result (first proven by Robert J. Aumann, appearing as Theorem 14.3.2 in Klein, 1984). As usual, we use the term *graph* of a correspondence $F$ from $X$ to $Y$ to denote the set $\{(x, y) | y \in F(x)\}$.

**Proposition (Measurable Selection)** *Let $X, Y$ be Borel spaces, and let $F$ be a correspondence from $X$ to $Y$ such that the graph of $F$ is a Borel subset of $X \times Y$, and let*

$\lambda \in \Delta(X)$. *Then there exists a* $\lambda$*-measurable map* $f : X \to Y$ *such that* $f(x) \in F(x)$ *for* $\lambda$*-almost* $x \in X$.

As in Theorem 1, the argument for necessity is similar to that of the finite case. Indeed, fix $s_1, \ldots, s_n$. After observing a sequence $s_1, \ldots, s_n$ of signals, a Bayesian agent has a probability distribution $\lambda$ over $S$ which reflects her prediction regarding the next signal $s_{n+1}$ she will receive. Given $s_{n+1}$, the agent will update her prediction over $K$ to $\sigma[s_1, \ldots, s_{n+1}]$. Therefore $\lambda$ induces a distribution $\lambda'$ over distributions over $K$: The agent's belief about what her posterior predictions corresponding to the additional observation of $s_{n+1}$ are. Thus $\lambda' \in \Delta(\Delta(K))$. Naturally, $\lambda'$ is concentrated on the set $\{\sigma[s_1, \ldots, s_n, s_{n+1}] | s_{n+1} \in S\}$, and, moreover, the Bayesian updating rule dictates that the barycenter of $\lambda'$ be equal to $\sigma[s_1, \ldots, s_n]$.

We now prove sufficiency. For every $s_1, \ldots, s_n$, by the assumption of the theorem there exists a probability measure $\lambda'$ over $\Delta(\Delta(K))$ which is concentrated on $\{\sigma[s_1, \ldots, s_n, s_{n+1}] | s_{n+1} \in S\}$ such that the barycenter of $\lambda'$ is $\sigma[s_1, \ldots, s_n]$. We will use the following lemma (appearing as Corollary 18.24 in Aliprantis and Border, 2006. We provide a short proof here for the sake of completeness).

**Lemma** *Let* $X, Y$ *be Borel spaces and let* $g : Y \to X$ *be a surjective measurable map, and let* $\lambda' \in \Delta(X)$. *Then there exists* $\lambda \in \Delta(Y)$ *such that* $g(\lambda) = \lambda'$.

**Proof of Lemma.** Consider the correspondence $F$ from $X$ to $Y$ such that $F(x) = \{y \in Y | g(y) = x\}$. Then the graph of $F$ and the graph of $g$ are the same set. Thus the graph of $F$ is a Borel set (as the graph of the Borel function $g$). It follows from the measurable selection proposition that there exists a $\lambda'$-measurable function $f : X \to Y$ such that $g \circ f = \mathrm{id}$ $\lambda'$-almost surely. Then $\lambda = f(\lambda')$ satisfies $g(\lambda) = \lambda'$. ■

By the Lemma, $\lambda'$ can be pulled back to a probability measure over $S$. Denote this probability measure by $\lambda[s_1, \ldots, s_n]$. By the measurable selection proposition $\lambda[s_1, \ldots, s_n]$

can be selected in a way that is measurable with respect to $s_1, \dots, s_n$. The distribution $\mu_n$ over $K \times S^n$ is define as follows: $s_1$ is randomized according to $\lambda[]$, $s_2$ given $s_1$ is randomized according to $\lambda[s_1]$, ..., $s_n$ given $s_1, \dots, s_{n-1}$ is randomized according to $\lambda[s_1, \dots, s_{n-1}]$ and $k$ is randomized according to $\sigma[s_1, \dots, s_n]$. ■

**Proof of Lemma (Relative Interior).**

Let $\Delta(S) = \{\lambda \in \mathbb{R}^S | \lambda_s \geq 0$ for every $s \in S$ and $\sum_{s \in S} \lambda_s = 1\}$. Then the affine hull of $\Delta(S)$ is of dimension $|S| - 1$ and ri $\Delta(S) = \{\lambda \in \mathbb{R}^S | \lambda_s > 0$ for every $s \in S$ and $\sum_{s \in S} \lambda_s = 1\}$. Let $A : \mathbb{R}^S \to V$ be the linear map that is given by $A\lambda = \sum_s \lambda_s x_s$. Note that $\mathrm{Conv}(\{x_s | s \in S\}) = A(\Delta(S))$. Therefore

$$\text{ri Conv}(\{x_s | s \in S\}) = \text{ri } A(\Delta(S)) = A(\text{ri } \Delta(S)) = \{\sum_{s \in S} \lambda_s x_s | 0 < \lambda_s \text{ for every } s \in S \text{ and } \sum_{s \in S} \lambda_s = 1\},$$

where the second equality is Theorem 6.6 in Rockafellar (1996). ■