Tests of Bayesian Rationality*

Pooya Molavi†

September 14, 2021

Abstract

What are the testable restrictions imposed on the dynamics of an agent's belief by the hypothesis of Bayesian rationality, which do not rely on the additional assumption that the agent has an objectively correct prior? In this paper, I argue that there are essentially no such restrictions. I consider an agent who chooses a sequence of actions and an econometrician who observes the agent's actions but not her signals and is interested in testing the hypothesis that the agent is Bayesian. I argue that—absent a priori knowledge on the part of the econometrician on the set of models considered by the agent—there are almost no observations that would lead the econometrician to conclude that the agent is not Bayesian. This result holds even if the set of actions is sufficiently rich that the agent's action fully reveals her belief about the payoff-relevant state and even if the econometrician observes a large number of identical agents facing the same sequence of decision problems.

*This paper is based on the second chapter of my PhD dissertation at MIT. I am grateful to Daron Acemoglu, Marios Angeletos, Drew Fudenberg, Arda Gitmez, Parag Pathak, Alireza Tahbaz-Salehi, Juuso Toikka, Olivier Wang, Iván Werning, Alex Wolitzky, Muhammet Yildiz, and participants in the MIT theory lunch and BFI Saieh Family Fellows Seminar for their valuable comments.

†Northwestern University, pmolavi@kellogg.northwestern.edu.

1 Introduction

Following the treatise of Savage (1972), the subjective (or Bayesian) theory of probability has become the dominant paradigm in the economic modeling of decision making under uncertainty.[1] The dominance of the subjective theory of probability in economics is not unwarranted. It allows one to assign probabilities to unique (or rare) events. It has an elegant foundation in the study of rational choice under uncertainty. And it is appealing from a normative point of view—as Epstein and Le Breton (1993) proclaim, "dynamically consistent beliefs must be Bayesian." What is less clear is whether Bayesian rationality is a good positive model of individual behavior. To settle this question requires one to develop formal tests of subjective rationality.[2]

A class of tests commonly used in the literature are those based on the martingale property of beliefs: a Bayesian agent's belief sequence is a martingale with respect to the agent's own prior belief; and any sequence of beliefs that constitutes a martingale with respect to some probability distribution $\mathbb{P}$ can be rationalized as the belief sequence of a Bayesian agent with prior $\mathbb{P}$. This equivalence of Bayesian rationality and the martingale property of beliefs has been known by economists since, at least, Kamenica and Gentzkow (2009).[3]

But although a Bayesian agent's beliefs constitute a martingale with respect to her *subjective belief*, to operationalize the martingale test requires one to use a panel (or repeated cross section) of individuals, in which the sample is drawn from a sampling distribution that depends on the *objective distribution* of the agents' observations. The martingale test is thus a valid test of Bayesian rationality only if agents have objectively correct beliefs about the data-generating process. Said differently, a group of ex-ante-identical agents could fail a martingale-based test of Bayesian rationality for one of two distinct reasons: (i) the agents may not be Bayesian or (ii) their belief about the distribution of the signals they observe may not coincide with the true, objective distribution.

This paper's main contribution is to characterize the empirical content of the theory of subjective rationality without predicating it on the a priori assumption that individuals' beliefs about the data-generating process are accurate. I consider an observer, "the econometrician," who observes a sequence of decisions made by a decision maker, "the agent." The

[1] Savage's theory of subjective probability builds on earlier works by De Finetti (1937), Ramsey (1926), and Von Neumann and Morgenstern (1944).

[2] I use the terms "subjective rationality" and "Bayesian rationality" interchangeably throughout this paper.

[3] The necessity of the martingale property for Bayesian rationality is a trivial consequence of the law of iterated expectation. To the best of my knowledge, Kamenica and Gentzkow (2009) are the first to formalize the sufficiency of the martingale property for Bayesian rationality in the economics literature.

econometrician (he) is wondering whether there are decision sequences that would lead him to conclude, with some certainty, that the agent (she) is not Bayesian. The econometrician observes the decisions made by the agent, but other aspects of the agent's probabilistic model of the world (including her belief about the data-generating process) remain unknown to him.[4] He is free to complete the unobserved parameters of the agent's model in an attempt to rationalize the sequence of decisions made by the agent. The main result of the paper is to establish that almost any sequence of decisions by the agent can be rationalized by the econometrician in the manner just described.

One may object that two obvious obstacles stand in the way of the econometrician's rejecting of the agent's Bayesian rationality. First, the mapping from the agent's beliefs to her actions may not be known to the econometrician or to be invertible. So the econometrician may not be able to separately identify the agent's preferences and her beliefs. Second, any seemingly irrational realization of the agent's belief sequence may be the result of observations by the agent that were unlikely ex ante.

To address these potential objections, I give the econometrician powerful tools to overcome the obstacles outlined in the previous paragraph. In particular, I assume that the mapping from the agent's beliefs to her actions is an invertible mapping that is known to the econometrician. This is equivalent to assuming that the econometrician directly observes the sequence of beliefs held by the agent. This assumption allows the econometrician to separately identify the agent's preferences and her beliefs.

I additionally assume that the econometrician can directly observe the true, objective population distribution from which the agent's belief sequence is drawn. This assumption can be thought of as representing the limit where the econometrician has access to the belief sequence of a large sample of ex ante identical agents who observe independent signals drawn from a common distribution and who face identical decision problems. The observation of the population distribution allows the econometrician to overcome the second obstacle mentioned above.

I show that there are few observations that lead the econometrician to conclude that the agent is not Bayesian—despite the unrealistically powerful tools that are put at his disposal. To be more specific, the only testable implication of the theory of Bayesian rationality is the absolute continuity of the posterior with respect to the prior for any prior-posterior pair that is realized with positive probability. The result holds even if the agent's prior belief about the payoff-relevant state agrees with its objective distribution. And it does not rely on the

[4] I use the term "probabilistic model of the world" (or simply model) to refer to the subjective probability space $(\Omega, \mathcal{F}, \mu)$ (in the sense of Savage) that rationalizes a rational agent's choices. I use the term "theory" to refer to a hypothesis (such as subjective rationality) being tested by the econometrician.

inapplicability of Bayes' rule after contingencies that were assigned zero probability by the agent's prior.

The result shows that although subjective rationality is a refutable theory, it has weak testable predictions absent additional a priori assumptions on an individual's probabilistic model of the world. The result also illustrates the difficulty of using observational data to distinguish models of non-Bayesian behavior (such as confirmation bias (Rabin and Schrag, 1999) or diagnostic expectations (Bordalo, Gennaioli, and Shleifer, 2018)) from models where agents have internally consistent beliefs that are based on incorrect priors about the data-generating process. It suggests imposing assumptions on what constitutes a reasonable model as a way of obtaining theories with more predictive power. Doing so leads to hybrid theories in which some aspects of agents' models are assumed to conform to objective reasonableness requirements while other aspects of the models are subjective.

I proceed by introducing two such a priori reasonableness assumptions, discussing the contexts in which they are likely to be satisfied, and characterizing their testable implications in conjunction with the assumption of subjective rationality. One such reasonableness assumption is the one discussed at the beginning of the introduction, that the agents' beliefs about the data-generating process conforms to the objective truth. This assumption is likely to be satisfied in lab experiments where the econometrician has full control over the data-generating process and can clearly communicate its parameters to the agents who are the subject of the experiment. As mentioned earlier, the empirical content of this theory is summarized by the martingale property of belief sequences.

The second such assumption is that an agent's decisions fully reflect her belief over the entire state space (and not a section of it). This assumption is likely to be satisfied in lab experiments where the econometrician can (i) plausibly constrain the set of informative signals observed by the agents and (ii) solicit the agents' beliefs about any event in their model of the world, including the probabilities they assign to observing different signals. I show that such a theory imposes sharp restrictions on the agents' prior-posterior pairs. In particular, the set of Bayesian posteriors is fully pinned down by the agents' prior beliefs.

The paper most closely related to this paper is by Shmaya and Yariv (2016). They also provide a characterization of the testable implications of Bayesian rationality and conclude that it is often hard to reject the agents' Bayesian rationality. But beyond a superficial similarity, the two papers are different in their assumptions on what is known a priori by the econometrician and the data available to him and in their precise conclusions. While Shmaya and Yariv focus on experimental settings in which the econometrician has a priori knowledge of the agents' set of signals, in my setting the econometrician cannot rule out the possibility that agents observe private signals in between the two periods. On the other hand, in Shmaya and Yariv the econometrician observes a repeated cross-section of agents, while in my setting the econometrician's observations have a panel structure.

The results of the two papers are thus not directly comparable, being based on distinct assumptions about what is known and observed. The econometrician's a priori knowledge of the set of signals in Shmaya and Yariv (2016) makes it easier for him to reject the agents' Bayesian rationality, whereas his inability to track the agents' beliefs over time makes is harder to do so. The two papers nonetheless reach similarly negative yet distinct conclusions. Shmaya and Yariv (2016) show that everything can be rationalized in their setting; a violation of absolute continuity is the only thing that prevents an observation from being rationalized in mine.

The rest of the paper is organized as follows. Section 2 illustrates the main result of the paper using a simple example. Section 3 formally poses the central question of the paper and presents the main results of the paper. The proofs are relegated to the appendix.

2 A Simple Example

This section uses a simple example to illustrate the main result of the paper and the construction that is used in its proof. There are two time periods: $t = 0, 1$. In each period, an econometrician observes the beliefs of a large number of ex ante identical agents about some payoff-relevant state of the world. The agents may observe informative signals between periods zero and one. But the econometrician does not know the set of signals, the agents' beliefs about the data-generating process, or the realizations of the signals observed by the agents (if any).

The payoff-relevant state takes values in the set $S = \{H, L\}$. In period $t = 0$, the agents' belief about the payoff-relevant state is as follows:

H	0.5
L	0.5

Since agents are assumed to be ex ante identical, they all have the same prior belief: they all believe that the H and L states are equally likely. In period $t = 1$, a quarter of the agents in the population have the following belief about the payoff-relevant state:

H	0.8
L	0.2

The belief of the remaining three quarters of the agents is as follows:

H	1
L	0

Should the observation of this belief sequence lead the econometrician to conclude that the agents are not Bayesian? The answer may seem to be yes at first. This belief sequence does not constitute a martingale with respect to the process that generates the signals observed by the agents: every agent becomes more confidence in period $t = 1$ that H is the true state. But the martingale test of Bayesian rationality requires the agents' beliefs to be a martingale only with respect to their subjective priors (and not the true data-generating process). The agents' beliefs may satisfy the martingale property with respect to their subjective prior about the data-generating process—but not with respect to the (true) process that generates the agents' observations.

The agents' beliefs in this example are in fact consistent with Bayesian updating. This can be shown by specifying the space that captures the underlying uncertainty, the agents' belief about the underlying uncertainty, and the true data-generating process in such a way that the beliefs of Bayesian agents match the observed priors and posteriors about the payoff-relevant state.

The observed belief sequence can be rationalized as follows. The signal observed between periods zero and one is drawn from the set $\{0.8, 1\} \times \{+, -\}$. The underlying uncertainty thus can be represented by the probability space $\Omega = \{H, L\} \times \{0.8, 1\} \times \{+, -\}$. The agents' prior belief $\mu_0 \in \Delta\Omega$ is given by

	0.8^+	1.0^+	0.8^-	1.0^-
H	0.25	0.25	0	0
L	0.0625	0	0.1875	0.25

The agents' information is represented by the sigma-algebra $\mathcal{F}_1$, which is generated by the partition illustrated in red. According to the agents' prior, the probability that the payoff-relevant state is H is 0.5. This is consistent with the observed prior ν_0^*. The agents' posterior belief about the payoff-relevant state equals $\nu_1^*(H) = 0.8$ conditional on the signal being 0.8^+, equals $\nu_1^*(H) = 1$ conditional on the signal being 1.0^+, and equals $\nu_1^*(H) = 0$ otherwise.

The true data-generating process is described by the following distribution:

	0.8^+	1.0^+	0.8^-	1.0^-
H	0.125	0.375	0	0
L	0.125	0.375	0	0

Under this distribution, signal 0.8^+ is realized with probability 0.25 and signal 1.0^+ is realized with probability 0.75. Therefore, a quarter of the agents will have the posterior belief $\nu_1^*(H) = 0.8$ and the remaining three quarters will have the posterior belief $\nu_1^*(H) = 1.0$. This is exactly the posterior distribution observed by the econometrician.

The agents' belief sequence constitutes a martingale with respect to their subjective prior. The observation of signals 0.8^+ and 1.0^+ raises the agents' belief in the high state—hence the + superscript—while the observation of signals 0.8^- and 1.0^- lowers their belief in the high state. According to the agents' subjective prior, the positive and negative signals are just likely enough to make the agents' belief sequence a martingale. Yet under the true distribution, the negative signals are unlikely.

The main result of the paper generalizes this example. It shows that any distribution of prior-posterior pairs is consistent with Bayesian rationality as long as some weak absolute continuity assumption is satisfied. The proof of the main result follows the construction in the example. The econometrician rationalizes the observed posteriors as arising from Bayesian updating by agents whose prior assigns positive probability to a set of signals that are unlikely given the true data-generating process.

Note that the objective distribution coincides with the agents' subjective prior about the distribution of the payoff-relevant state. In particular, according to both the agents' subjective prior and the objective distribution, the two payoff-relevant states are equally likely ex ante. So the econometrician can rationalize the observed sequence of beliefs without requiring the agents to hold a prior about the payoff-relevant state that disagrees with the objective distribution; the agents only need to be wrong about the data-generating process. This is a general feature of the construction used in the proof the main theorem: (almost) any belief sequence can be rationalized by the econometrician using an objective distribution that respects the agents' prior about the payoff-relevant state.

3 The General Results

In this section, I generalize the insights of the example discussed in the previous section by characterizing the testable implications of Bayesian rationality. The main challenge is to formally express what it means for an agent to be subjectively rational when there is a true underlying data-generating process that determines the signals observed by the agent. I start by introducing a framework that combines elements of subjective and objective probabilities: the econometrician is interested in testing the hypothesis that agent has an internally consistent *subjective* probability systems but can only use samples drawn from an *objective* distribution. I then formally state the question of which belief sequences are consistent with Bayesian rationality.

3.1 Mathematical Preliminaries

I maintain the following standard technical assumptions throughout the remainder of the paper. Every set X is assumed to be a complete separable metric space that is endowed with its corresponding Borel sigma-algebra $\mathcal{X}$. The set of probability distributions over $(X, \mathcal{X})$ is denoted by ΔX and is endowed with the topology of weak convergence and the corresponding Borel sigma-algebra, which I denote by $\mathcal{B}(\Delta X)$.

3.2 The Environment

I consider the following environment. There are two periods indexed by $t = 0, 1$ and two types of actors in the economy: a population of ex ante identical decision-makers, or "agents," indexed by $i \in I$ and an observer, or "the econometrician." Each agent i makes a decision a^*_{it} in period $t = 0, 1$. The econometrician is interested in testing the hypothesis that the agents are Bayesian.

There is a fixed *payoff-relevant state* s that belongs to a measurable space $(S, \mathcal{S})$. The set of payoff-relevant states is fixed and known to the econometrician. The set S can be thought of as the largest slice of the state space capturing the underlying uncertainty over which the agents' beliefs can be solicited by the econometrician (either directly or indirectly). I let $\nu^*_{it} \in \Delta S$ denote the time-t belief of agent i about the value of the payoff-relevant state s.

Agents may observe an informative signal about the value of the payoff-relevant state between periods 0 and 1. The realized signals are independent and identically distributed across the agents. The econometrician knows what has been described so far and observes the decisions made by the agents, but he does not know anything about the signals observed by the agents between the two periods (if any).

I assume that the mapping from an agent's belief ν^*_{it} about the payoff-relevant state to her action a^*_{it} is invertible and known by the econometrician. This assumption is equivalent to assuming that the econometrician directly observes the agents' beliefs about the payoff-relevant state. It allows me to abstract away from the question of whether preferences and beliefs can be separately identified. Moreover, it makes the negative result of the paper only stronger: if the econometrician cannot rule out the possibility that the agents are Bayesian after observing their beliefs, then a fortiori, he will not be able to rule out their Bayesian rationality by observing their actions.

The econometrician is interested in using his sample $\{(\nu^*_{i0}, \nu^*_{i1})\}_{i\in I}$ of prior-posterior pairs to test the hypothesis that the agents are Bayesian. I assume that I is sufficiently large that the empirical distribution of $\{(\nu^*_{i0}, \nu^*_{i1})\}_{i\in I}$ equals the population distribution from which the agents' prior-posterior pairs are drawn. This assumption, too, serves to make the negative results of the paper only stronger: if the econometrician cannot reject the Bayesian rationality of the agents when he has access to the population distribution of their prior-posterior pairs, then a fortiori, he will not be able to do so given the corresponding empirical distribution.

I can use the assumption of large I to replace the population of agents with a single "representative agent" with random beliefs. Since the agents are ex ante identical, the prior of the representative agent is drawn from a degenerate distribution with unit mass at some $\nu^*_0 \in \Delta S$. On the other hand, the agents may observe different realizations of the signal between periods 0 and 1, and so, may end up with different posteriors. I let $P^*_1 \in \Delta(\Delta S)$ denote the distribution of the representative agent's posterior. The question is then which pairs $\left(\nu^*_0, P^*_1\right)$, consisting of the representative agent's prior and the probability distribution of her posterior, are consistent with Bayesian rationality. In the next subsection, I lay the groundwork to formally express what it means for a pair $\left(\nu^*_0, P^*_1\right)$ to be consistent with Bayesian rationality. Whenever there is no risk of confusion, I refer to the representative agent simply as the agent.

3.3 Formalism

The underlying uncertainty and the agent's information can be fully captured by a measure space $(\Omega, \mathcal{F})$ and sigma-algebras $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}$. The pair $(\Omega, \mathcal{F})$ is an abstract measurable space that captures all the uncertainty faced by the agent. Each $\omega \in \Omega$ is a complete description of all the variables that affect the agent's decisions. It includes, at a minimum, the value of the payoff-relevant state s, but it also includes the description of any signals that may be observed by the agent between the two periods. I refer to ω as the *state of the world* to contrast it with s, the payoff-relevant state. I let $\mathbf{S}(\omega) \in S$ denote the value of the payoff-relevant state when the state of the world is given by ω, with $\mathbf{S} : \Omega \to S$ a measurable mapping.[5] The sigma-algebra $\mathcal{F}_t$ captures the agents' information in period t—the agents'

[5] It is without loss of generality to assume that Ω is a superset of S, the restriction of $\mathcal{F}$ to S is equal to $\mathcal{S}$, and $\mathbf{S}$ is the canonical projection.

actions in period t are measurable with respect to $\mathcal{F}_t$. I assume without loss of generality that $\mathcal{F}_0$ is the trivial sigma-algebra. Therefore, $\mathcal{F}_1$ represents the information content of the signal observed by the agent between periods 0 and 1.

A Bayesian agent's subjective prior about the state of the world can be represented by a probability distribution $\mu_0 : \mathcal{F} \to [0,1]$. The probability distribution μ_0 captures both the agent's belief about the payoff-relevant state and her prior about the distribution of the signal. The agent's induced subjective prior about the payoff-relevant state is given by the probability distribution $\nu_0 : \mathcal{S} \to [0,1]$, defined as

$$\nu_0(B) \equiv \mu_0\left(\mathbf{S}^{-1}(B)\right) \tag{1}$$

for any arbitrary event $B \in \mathcal{S}$.[6]

The agent is Bayesian if her posterior about the state of the world is obtained from her prior by conditioning on the sigma-algebra $\mathcal{F}_1$. This requirement can be stated formally using the concept of regular conditional probability. A mapping $\mu_1 : \Omega \times \mathcal{F} \to [0,1]$ is a *(subfield) regular conditional probability* for $(\Omega, \mathcal{F}, \mu_0)$ given $\mathcal{F}_1$ if (i) the mapping $\omega \mapsto \mu_1(\omega, B)$ is $\mathcal{F}_1$-measurable for all $B \in \mathcal{F}$ and (ii)

$$\mu_0(B \cap E) = \int_E \mu_1(\omega, B)\mu_0(d\omega) \tag{2}$$

for any $B \in \mathcal{F}$ and $E \in \mathcal{F}_1$.[7] The notion of regular conditional probability is the natural generalization of the elementary notion of conditional probability to probability distributions with uncountable supports. Condition (i) is the measurability requirement: the posteriors need to be the same conditional on any two states that are indistinguishable given $\mathcal{F}_1$. Condition (ii) is Bayes' rule. It is the internal consistency requirement that underpins any test of Bayesian updating.

The consistency condition (2) is not directly verifiable by the econometrician. The econometrician only observes the agent's prior and posterior *over* S, but Bayesian rationality requires the agent's prior and posterior to satisfy the consistency requirement (2) *over* Ω—and S is in general only a slice of Ω. Yet equation (2) induces a consistency requirement for the agent's belief over S. More specifically, the regular conditional probability μ_1 and the random variable $\mathbf{S} : \Omega \to S$ define a regular conditional probability $\nu_1 : \Omega \times \mathcal{S} \to [0,1]$ as follows: for any $\omega \in \Omega$ and $B \in \mathcal{S}$,

$$\nu_1(\omega, B) \equiv \mu_1\left(\omega, \mathbf{S}^{-1}(B)\right). \tag{3}$$

[6]The probability measure ν_0 is known as the *pushforward measure*.

[7]For a proof of the existence of a regular conditional probability when the underlying space is Polish, see Faden (1985).

Intuitively, $\nu_1(\omega, B)$ is the agent's posterior belief that the payoff-relevant state belongs to set B conditional on the event that the realized state of the world is ω.

While the agent's prior is a *subjective* probability distribution, the distribution of her posterior depends on the *objective* distribution of the signals she observes. Given the set of states of the world $(\Omega, \mathcal{F})$, the agent's information structure $\mathcal{F}_1$, the mapping $\mathbf{S}$ from the state of the world to the payoff-relevant state, and the agent's prior μ_0, equation (3) defines the agent's posterior belief about the value of the payoff-relevant state. The agent's posterior is a random variable whose realization depends on the signal observed by the agent. Therefore, the distribution of the agent's posterior depends on the objective distribution from which her signal is drawn. I let $\mathbb{P} \in \Delta\Omega$ denote the objective probability distribution that determines the distribution of the signal observed by the agent. Given $((\Omega, \mathcal{F}), \mathbf{S}, \mu_0, \mathcal{F}_1, \mathbb{P})$, the agent's posterior about the payoff-relevant state is distributed according to the probability distribution $P_1 \in \Delta(\Delta S)$, defined as

$$P_1(B_1) = \mathbb{P}\left(\{\omega \in \Omega : \nu_1(\omega, \cdot) \in B_1\}\right) \tag{4}$$

for any $B_1 \in \mathcal{B}(\Delta S)$, where ν_1 is the regular conditional probability defined in (3).

3.4 Tests of Bayesian Rationality

I can now use the notation introduced in the previous subsection to formally state what it means for the agent's belief sequence to be consistent with Bayesian rationality.

Definition 1. A pair $\left(\nu_0^*, P_1^*\right)$ of observations consisting of the agent's prior and the distribution of her posterior is *consistent with Bayesian rationality given* $(\Omega, \mathcal{F}, \mathcal{F}_1, \mathbf{S}, \mu_0, \mathbb{P})$ if $\nu_0 = \nu_0^*$ and $P = P^*$, where ν_0 is defined in (1) and P_1 is defined in (4).

The definition considers one extreme case where the econometrician has a priori knowledge about every aspect of the environment: the underlying state space, the agent's prior on the entire state space, the set of signals observed by the agent, and the true data-generating process. Given this knowledge, the only distribution of posteriors that is consistent with the agent's Bayesian rationality and her subjective prior is the one defined in (4). I next consider the other extreme where the econometrician does not have any a priori knowledge about the environment.

Definition 2. A pair $\left(\nu_0^*, P_1^*\right)$ of observations consisting of the agent's prior and the distribution of her posterior is *consistent with Bayesian rationality* if it is consistent with Bayesian rationality given some $(\Omega, \mathcal{F}, \mathcal{F}_1, \mathbf{S}, \mu_0, \mathbb{P})$.

The definition does not confound Bayesian rationality with other restrictions on what constitutes an objectively reasonable model of the world (such as having a correct prior about the data-generating process). The econometrician rejects the agent's Bayesian rationality in the sense of Definition 2 only if there is no internally consistent subjective probability that rationalizes the observed prior and posterior distribution. It is this definition that is used to state the main question of the paper:

Question 1. Which pairs (ν_0^*, P_1^*) of priors and posterior distributions are consistent with the agent's Bayesian rationality?

The question formalizes an intuitive scenario. The econometrician observes the prior ν_0^* and distribution P_1^* of posteriors but has no a priori knowledge of other aspects of the environment. He chooses the tuple $(\Omega, \mathcal{F}, \mathcal{F}_1, \mathbf{S}, \mu_0, \mathbb{P})$ in an attempt to explain his observation as resulting from Bayesian updating by agents with prior μ_0 in a world described by $(\Omega, \mathcal{F}, \mathcal{F}_1, \mathbf{S}, \mathbb{P})$. If he is unable to find a tuple $(\Omega, \mathcal{F}, \mathcal{F}_1, \mathbf{S}, \mu_0, \mathbb{P})$ that explains his observation, he concludes that the agent is not Bayesian. The main result of the paper establishes that he can always do so as long as the following condition is satisfied:

Condition 1. For any ν_1^* in the support of P_1^*, the probability distribution ν_1^* is absolutely continuous with respect to ν_0^* with an essentially bounded Radon–Nikodym derivative.[8]

This is a weak condition. In particular, it reduces to absolute continuity if S is a finite set. It is easy to see that absolute continuity is necessary for (ν_0^*, P_1^*) to be consistent with Bayesian rationality: if the prior of a Bayesian agent assigns zero probability to an event, her posterior must also assign zero probability to the event—regardless of the set of signals, the agent's belief about the distribution of signals, and the true data-generating process. What is more surprising is that Condition 1 is sufficient for the observed pair (ν_0^*, P_1^*) to be consistent with Bayesian rationality.

Theorem 1. *Suppose the pair (ν_0^*, P_1^*) of observations, consisting of the agent's prior and the distribution of her posterior, satisfies Condition 1. Then it is consistent with Bayesian rationality.*

The proof is constructive. The construction generalizes the example of Section 2. The econometrician constructs a large enough state space Ω, an objective distribution $\mathbb{P}$, and a subjective belief μ_0 for the agent under which the agent's prior about the probability of the payoff-relevant state s coincides with ν_0^* and the distribution of her Bayesian posterior about

[8]The Radon–Nikodym derivative $f \equiv d\nu_1^*/d\nu_0^*$ is *essentially bounded* if there exists a constant $c < \infty$ and a set $\widehat{S} \in \mathcal{S}$ with $\nu_0^*(\widehat{S}) = 1$ such that $f(s) \leq c$ for all $s \in \widehat{S}$.

s coincides with P_1^*. The construction requires the econometrician to postulate an objective probability distribution that is in general different from the subjective prior held by the agent.

But the econometrician observes the agent's prior belief about the payoff-relevant state. Can requiring the objective distribution to respect the observed prior of the agent restrict the set of observations that are consistent with Bayesian rationality? The next theorem shows that the answer is in fact no. The econometrician can rationalize almost any distribution of posteriors using an objective distribution that agrees with the agent's prior about the distribution of the payoff-relevant state.

Before stating the theorem, I formally define what it means for an objective distribution $\mathbb{P}$ to agree with the agent's prior about the payoff-relevant state. Recall that $(\Omega, \mathcal{F})$ denotes the underlying state space and $\mathbf{S}$ is the random variable that determines the value of the payoff-relevant state as a function of the underlying state. Given $(\Omega, \mathcal{F})$ and $\mathbf{S}$, the distribution of the payoff-relevant state implied by the objective distribution $\mathbb{P}$ is given by

$$\eta_0(B) \equiv \mathbb{P}\left(\mathbf{S}^{-1}(B)\right) \tag{5}$$

for any arbitrary event $B \in \mathcal{S}$.

Definition 3. Given the underlying probability space $(\Omega, \mathcal{F})$ and the random variable $\mathbf{S}$, the objective probability $\mathbb{P}$ *agrees with the subjective prior* ν_0^* about the distribution of the payoff-relevant state if $\eta_0 = \nu_0^*$, where η_0 is defined in (5).

The next theorem is a generalization of Theorem 1. It establishes that requiring agreement with the subjective prior does not put any restrictions, above and beyond Condition 1, on the set of observations that are consistent with Bayesian rationality.

Theorem 2. *Suppose the pair (ν_0^*, P_1^*) of observations, consisting of the agent's prior and the distribution of her posterior, satisfies Condition 1. Then it is consistent with Bayesian rationality given an objective probability $\mathbb{P}$ that agrees with the subjective prior ν_0^* about the distribution of the payoff-relevant state.*

The theorem has a striking consequence. Bayesian rationality does not impose any meaningful restriction on the distribution of posteriors even if the agent's observed prior agrees with the objective distribution of the payoff-relevant state. Even if the econometrician observes the agent's prior belief over the set S and even under the assumption that the agent has a correct prior over S, there is no restriction on the agent's Bayesian posterior other than the absolute continuity Condition 1. Note that the set S can be arbitrarily large—the econometrician may solicit the agent's beliefs about the probabilities of an arbitrarily large set of events. And yet, there is almost no distribution of posteriors that cannot be made a martingale with respect to the agent's prior over S by choosing a sufficiently large Ω.

Intuitively, if the econometrician only observes the dynamics of an agent's belief about the variables that belong to S, then there is always a richer state space Ω that encodes more complex models of the world over which variables in S are defined such that the observed belief dynamic about S is rational given some subjective prior over Ω. The next result further refines Theorem 2 by establishing that the probability space Ω can be chosen independently of the observation (ν_0^*, P_1^*) that the econometrician is attempting to rationalize.

Theorem 3. *Given any space of payoff-relevant states $(S, \mathcal{S})$, there exists a probability space $(\Omega, \mathcal{F})$ and a random variable $\mathbf{S}$ such that any pair (ν_0^*, P_1^*) satisfying Condition 1 is consistent with Bayesian rationality given $(\Omega, \mathcal{F}, \mathcal{F}_1, \mathbf{S}, \mu_0, \mathbb{P})$ for some $\mathcal{F}_1$, μ_0, and $\mathbb{P}$.*

The results cast doubts on the possibility of deciding whether agents are Bayesian in non-experimental settings. In such settings, an econometrician may be able to impute the agents' beliefs over some set S. But there is no guarantee that any such set S captures *all* the uncertainty that the agents believe to be relevant to what they think about $s \in S$. In particular, given any set S such that the agents' beliefs on S can be solicited by the econometrician, it may be the case that the relevant uncertainty (from the point of view of the agents) is captured by a larger set Ω. The theorems then show that, for any set such S (capturing all that the econometrician can learn about the agents' beliefs), he can postulate a larger set Ω (capturing all the uncertainty that is relevant to the agents) such that almost any belief sequence over S can be rationalized by fine-tuning the agents' beliefs over Ω.

In experimental settings where the econometrician controls the agents' observations, on the other hand, he can solicit the agents' belief over a set S that can be plausibly assumed to be large enough to capture all the uncertainty that is relevant to the agents' decisions. In other words, in experimental settings the set Ω can be taken to be fixed and known to the econometrician. Moreover, the econometrician can solicit the beliefs over the entire set Ω (so that S coincides with Ω). Then there are relatively tight restrictions on the belief sequences that are consistent with Bayesian rationality. These restrictions are spelled out in the following proposition for the case where S is finite and ν_0^* has full support.[9]

Proposition 1. *Suppose $(\Omega, \mathcal{F}) = (S, \mathcal{S})$ and $\mathbf{S} = id_\Omega$, where S is a finite set. Given a prior ν_0^* with full support and a distribution of posteriors P_1^*, there exists some $\mathcal{F}_1$, μ_0, and $\mathbb{P}$ such that (ν_0^*, P_1^*) is consistent with Bayesian rationality given $(\Omega, \mathcal{F}, \mathcal{F}_1, \mathbf{S}, \mu_0, \mathbb{P})$ if and only if*

[9] The extensions to the cases where S is not finite or ν_0^* does not have full support are straightforward. But such extensions require additional technical assumptions. I do not pursue those extensions here for the sake of exposition.

(i) $\operatorname{supp} \nu \cap \operatorname{supp} \widehat{\nu} = \emptyset$ *for any distinct* $\nu, \widehat{\nu} \in \operatorname{supp} P_1^*$*;*

(ii) $\nu(\widehat{S}) = \nu_0^*(\widehat{S} | \operatorname{supp} \nu)$ *for any* $\widehat{S} \in \mathcal{S}$ *and any* $\nu \in \operatorname{supp} P_1^*$.

The proposition clarifies the scope and logic of Theorems 1–3. The theorems rely on the assumption that the econometrician does not have any a priori knowledge about what constitutes a complete description of the uncertainty that is relevant to the agents' decisions and so is unable to observe the agents' beliefs over the underlying state space. He is thus free to reverse-engineer the agents' beliefs over the underlying state space in order to rationalize his observations. Proposition 1 shows how observing what the agents believe about the distribution of the underlying state ties the econometrician's hand.

A Proofs

Proof of Theorems 1–3

Let $\Omega = S \times \Delta S \times \{+,-\}$, and let $\mathcal{F}$ denote the product sigma-algebra. A generic element of Ω is denoted by $(s, \nu^\diamond)$, where s is an element of S, ν is a probability distribution over S, and $\diamond \in \{+,-\}$. Let $\mathbf{S} : \Omega \to S$ be the canonical projection onto S, that is, the mapping that maps $(s, \nu^\diamond)$ to s.

The objects defined above have the following intuitive interpretation. The first element of $(s, \nu^\diamond)$ is the value of the payoff-relevant state while the second element indexes the signal observed by the agents in between periods 0 and 1. The states in $S \times \Delta S \times \{+\}$ are the only ones that are realized with positive probability under the objective probability distribution, whereas states in $S \times \Delta S \times \{-\}$ are needed for the agents' subjective belief to satisfy the martingale property.

While Ω, $\mathcal{F}$, and $\mathbf{S}$ are independent of the observed prior and posterior distribution of the agent, $\mathcal{F}_1$, μ_0, and $\mathbb{P}$ do depend on them. Let $\mathcal{F}_1$ be the smallest sigma-algebra that makes all sets of the form $S \times \{\nu\} \times \{\diamond\}$ for $\nu \in \operatorname{supp} P_1^*$ and $\diamond \in \{+,-\}$ measurable. So $\nu^\diamond$ is the signal observed by the agent in between periods 0 and 1 given $\mathcal{F}_1$. I specify μ_0 such that the agent's posterior belief over S conditional on observing signal ν^+ is given by ν.

I start my construction of μ_0 by fixing some $\nu \in \operatorname{supp} P_1^*$ and specifying $\mu_0(\cdot|\nu^+)$ and $\mu_0(\cdot|\nu^-)$. Since ν is absolutely continuous with respect to ν_0^* by Condition 1, there exists a Radon–Nikodym derivative $f_\nu \equiv d\nu/d\nu_0^* : S \to \mathbb{R}_+$ such that

$$\nu(\widehat{S}) = \int_{\widehat{S}} f_\nu(s)\nu_0^*(ds)$$

for any $\widehat{S} \in \mathcal{S}$. Let $\epsilon_\nu = 1/\operatorname{ess\,sup}_{s\in S} f_\nu(s)$. Since ν_0^* and ν are both probability measures, $\epsilon_\nu \le 1$. Moreover, by Condition 1, $\epsilon_\nu > 0$.

With the definition of ϵ_ν in hand, I can specify $\mu_0(\cdot|\nu^\diamond)$ for $\diamond \in \{+,-\}$. For any $\widehat{S} \in \mathcal{S}$, let $\mu_0(\widehat{S}|\nu^+) = \nu(\widehat{S})$. If $\epsilon_\nu = 1$, pick $\mu_0(\cdot|\nu^-)$ to be an arbitrary probability measure over S; otherwise, let

$$\mu_0(\widehat{S}|\nu^-) = \frac{1}{1-\epsilon_\nu}\nu_0^*(\widehat{S}) - \frac{\epsilon_\nu}{1-\epsilon_\nu}\nu(\widehat{S}). \tag{6}$$

I still have to verify that $\mu_0(\cdot|\nu^-)$ as defined above is indeed a probability measure over S. In order to see this, first note that $\mu_0(\cdot|\nu^-)$ is countably additive since both ν_0^* and ν are probability measures and thus countably additive. Moreover,

$$\mu_0(S|\nu^-) = \frac{1}{1-\epsilon_\nu}\nu_0^*(S) - \frac{\epsilon_\nu}{1-\epsilon_\nu}\nu(S) = \frac{1}{1-\epsilon_\nu} - \frac{\epsilon_\nu}{1-\epsilon_\nu} = 1,$$

and

$$\mu_0(\emptyset|\nu^-) = \frac{1}{1-\epsilon_\nu}\nu_0^*(\emptyset) - \frac{\epsilon_\nu}{1-\epsilon_\nu}\nu(\emptyset) = 0,$$

where both equalities are due to the fact that ν_0^* and ν are probability measures over S. Finally, for any set $\widehat{S} \in \mathcal{S}$,

$$\begin{aligned}\mu_0(\widehat{S}|\nu^-) &= \frac{1}{1-\epsilon_\nu}\left(\int_{\widehat{S}} \nu_0^*(ds) - \epsilon_\nu \int_{\widehat{S}} f(s)\nu_0^*(ds)\right)\\ &\geq \frac{1}{1-\epsilon_\nu}\left(\int_{\widehat{S}} \nu_0^*(ds) - \int_{\widehat{S}} \nu_0^*(ds)\right) = 0.\end{aligned}$$

This proves that $\mu_0(\cdot|\nu^-)$ is a probability distribution over S.

To complete the description of μ_0, let λ denote an arbitrary probability distribution over ΔS with support $\operatorname{supp} P_1^*$, and for any measurable sets $\widehat{S} \subseteq S$ and $\widehat{\Delta S} \subseteq \Delta S$, let

$$\begin{aligned}\mu_0(\widehat{S} \times \widehat{\Delta S} \times \{+\}) &= \int_{\widehat{\Delta S}} \mu_0(\widehat{S}|\nu^+)\epsilon_\nu \lambda(d\nu),\\ \mu_0(\widehat{S} \times \widehat{\Delta S} \times \{-\}) &= \int_{\widehat{\Delta S}} \mu_0(\widehat{S}|\nu^-)(1-\epsilon_\nu) \lambda(d\nu).\end{aligned}$$

Finally, let $\mathbb{P}$ be the probability measure supported on $S \times \operatorname{supp} P_1^* \times \{+\}$ defined as $\mathbb{P}(\widehat{S} \times \widehat{\Delta S} \times \{+\}) = \nu_0^*(\widehat{S})P_1^*(\widehat{\Delta S})$ for any measurable sets $\widehat{S} \subseteq S$ and $\widehat{\Delta S} \subseteq \Delta S$. That is, according to $\mathbb{P}$, only states in $S \times \operatorname{supp} P_1^* \times \{+\}$ have positive probability, the probability that the agent observes signal ν^+ is given by the frequency of posterior ν in the target distribution P_1^*, and the objective probability distribution $\mathbb{P}$ agrees with the agent's observed prior ν_1^* over the set of observables S. This completes the construction.

To complete the proof, I have to show that $\nu_0 = \nu_0^*$ and that ν_1 is distributed according to P_1^* given the objective prior $\mathbb{P}$, where ν_0 and ν_1 are defined in equations (1) and (3), respectively. By the law of total probability, for any set $\widehat{S} \in \mathcal{S}$,

$$\begin{aligned}\nu_0(\widehat{S}) &= \mu_0(\mathbf{S}^{-1}(\widehat{S}))\\ &= \mu_0(\widehat{S} \times \Delta S \times \{+\}) + \mu_0(\widehat{S} \times \Delta S \times \{-\})\\ &= \int_{\Delta S}\Big(\mu_0(\widehat{S}|\nu^+)\epsilon_\nu + \mu_0(\widehat{S}|\nu^-)(1-\epsilon_\nu)\Big)\lambda(d\nu)\\ &= \int_{\Delta S} \nu_0^*(\widehat{S})\lambda(d\nu) = \nu_0^*(\widehat{S}),\end{aligned}$$

where in the fourth equality I am using (6) and the fact that $\mu_0(\widehat{S}|\nu^+) = \nu(\widehat{S}) = \nu_0^*(\widehat{S})$ for any set $\widehat{S} \in \mathcal{S}$ whenever $\epsilon_\nu = 1$.

On the other hand, for any $\nu \in \operatorname{supp} P_1^*$, by construction,

$$\nu_1(\widehat{S}|\nu^+) = \mu_0(\mathbf{S}^{-1}(\widehat{S})|\nu^+) = \nu(\widehat{S}),$$

and under the objective prior $\mathbb{P}$, slices $S \times \{\nu^+\}$ are distributed according to $P_1^*(\nu)$ while $S \times \Delta S \times \{-\}$ has zero probability. Therefore, ν_1 as defined in (3) is distributed according to P_1^* under the objective prior $\mathbb{P}$.

The proof is complete once I argue that $\mathbb{P}(\mathbf{S}^{-1}(\widehat{S})) = \nu_0^*(\widehat{S})$ for all sets $\widehat{S} \in \mathcal{S}$. But this is trivially true by construction. $\square$

Proof of Proposition 1

The "if" direction. Let $S(\nu) \in \mathcal{S} = \mathcal{F}$ denote the support of $\nu \in \operatorname{supp} P_1^*$, and let $S^c(\nu)$ denote its complement. Define

$$\mathcal{P} = \left\{S(\nu) : \nu \in \operatorname{supp} P_1^*\right\} \cup \left\{ \bigcap_{\nu \in \operatorname{supp} P_1^*} S^c(\nu) \right\}.$$

Since $\operatorname{supp} \nu \cap \operatorname{supp} \widehat{\nu} = \emptyset$ for any distinct $\nu, \widehat{\nu} \in \operatorname{supp} P_1^*$ by condition (i) of the proposition, $\mathcal{P}$ is a partition of Ω. Let $\mathcal{F}_1 \subseteq \mathcal{F}$ denote the sigma-algebra over $\Omega = S$ generated by $\mathcal{P}$, let $\mu_0 = \nu_0^* \in \Delta S = \Delta\Omega$, and let $\mathbb{P}$ be any probability measure over $\Omega = S$ that satisfies $\mathbb{P}(S(\nu)) = P_1^*(\nu)$ for all $\nu \in \operatorname{supp} P_1^*$. By construction, $\nu_0 = \mu_0 \circ \mathbf{S}^{-1} = \mu_0 = \nu_0^*$.

I next show that $P_1 = P_1^*$, where P_1 is defined in equation (4). Note that $\mu_1(\omega, \cdot)$ is a probability distribution over $(\Omega, \mathcal{F})$ for any $\omega \in \Omega$; it takes value $\mu_0(\cdot|\widehat{\Omega})$ whenever the state of the world ω belongs to cell $\widehat{\Omega}$ of partition $\mathcal{P}$. Therefore, since $\mathbf{S}$ is the identity mapping, ν_1 as defined in equation (3) takes value $\mu_0(\cdot|\widehat{\Omega}) = \nu_0(\cdot|\widehat{\Omega}) = \nu_0^*(\cdot|\widehat{\Omega})$ whenever ω belongs to $\widehat{\Omega}$. Recall that $S(\nu) = \operatorname{supp} \nu$ and $\nu_0^*(\cdot|\operatorname{supp} \nu) = \nu(\cdot)$ by condition (ii) of the proposition. Hence, ν_1 takes value ν whenever $\omega \in S(\nu)$. Finally, note that by construction $S(\nu)$ has probability $P_1^*(\nu)$ according to the objective probability measure $\mathbb{P}$. Therefore, ν_1 is equal to ν with probability $P_1^*(\nu)$ and so $P_1 = P_1^*$ in distribution.

The "only if" direction. I first show (ii). Since $\Omega = S$ is a finite set and $\mu_0 = \mu_0 \circ \mathbf{S}^{-1} = \nu_0 = \nu_0^*$ has full support over Ω, for any $\mathcal{F}_1$ there exists a partition $\mathcal{P} = \{\Omega_k : k \in K\}$ of Ω such that $\mu_1(\omega, \cdot)$ takes value $\mu_0(\cdot|\widehat{\Omega}) \in \Delta S = \Delta\Omega$ whenever the state of the world ω belongs to cell $\widehat{\Omega}$ of $\mathcal{P}$. Therefore, $\nu_1(\omega, \cdot) = \mu_1(\omega, \mathbf{S}^{-1}(\cdot)) = \mu_1(\omega, \cdot)$ only takes values in set $\{\mu_0(\cdot|\Omega_k) : k \in K\}$. Moreover, using the assumption that μ_0 has full support over Ω and Bayes' rule, I can conclude that $\operatorname{supp} \mu_0(\cdot|\Omega_k) = \Omega_k$ for all $k \in K$. On the other hand, $\mu_0 = \mu_0 \circ \mathbf{S}^{-1} = \nu_0 = \nu_0^*$. Therefore, for any realization ν of ν_1, as defined in (3), and any set $\widehat{S} \in \mathcal{S}$, we have that $\nu(\widehat{S}) = \nu_0^*(\widehat{S}|\operatorname{supp} \nu)$. Noting that, when $\mathbb{P}$ is the objective prior, ν_1 is distributed according to P_1, defined in (4), and that $P_1^* = P_1$ completes the proof of (ii).

I show (i) by contradiction. Toward a contradiction, suppose that there exist distinct ν, $\widehat{\nu} \in \operatorname{supp} P_1^*$ such that $\operatorname{supp} \nu \cap \operatorname{supp} \widehat{\nu} = S_0 \neq \emptyset$. Note that the supports of ν and $\widehat{\nu}$ cannot be equal because otherwise by (ii) ν and $\widehat{\nu}$ are not distinct. Let $S(\nu) = \operatorname{supp}\nu \setminus S_0$ and let $S(\widehat{\nu}) = \operatorname{supp}\widehat{\nu} \setminus S_0$. Since $\operatorname{supp}\nu \neq \operatorname{supp}\widehat{\nu}$, at least one of $S(\nu)$ and $S(\widehat{\nu})$ is non-empty. Without loss of generality, assume that $S(\nu) \neq \emptyset$. Since $P_1^* = P_1$, where P_1 is defined in (4), ν and $\widehat{\nu}$ are also realizations of $\mu_0(\mathbf{S}^{-1}(\cdot)|\mathcal{F}_1) = \mu_0(\cdot|\mathcal{F}_1)$. Therefore, $\operatorname{supp}\nu$ and $\operatorname{supp}\widehat{\nu}$ are measurable with respect to $\mathcal{F}_1$, and so are S_0, $S(\nu)$, and $S(\widehat{\nu})$ since they are intersections of measurable sets. Therefore,

$$\mu_0(\mathbf{S}^{-1}(S_0)|\mathcal{F}_1) = \mu_0(S_0|\mathcal{F}_1) = \mathbb{1}\{\omega \in S_0\} \in \{0,1\}. \tag{7}$$

In other words, the agent learns whether the state belongs to set S_0 given the information revealed by $\mathcal{F}_1$ in between periods 0 and 1. On the other hand, since (i) S is a finite set, (ii) neither of S_0 and $S(\nu)$ is empty, and (iii) $\operatorname{supp}\nu = S_0 \cup S(\nu)$, we have that $\nu(S_0) \in (0,1)$. Thus, equation (7) implies that ν cannot be a realization of P_1, contradicting the assumption that $\nu \in \operatorname{supp} P_1^* = P_1$. □

References

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

De Finetti, Bruno (1937), "La Prévision: ses Lois Logiques, ses Sources Subjectives." In *Annales de l'Institut Henri Poincaré*, volume 7, 1–68.

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

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

Kamenica, Emir and Matthew Gentzkow (2009), "Bayesian Persuasion." *American Economic Review*, 101, 2590–2615.

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

Ramsey, Frank P. (1926), "Truth and Probability." In *The Foundations of Mathematics and other Logical Essays* (R.B. Braithwaite, ed.), 156–198, London: Kegan, Paul, Trench, Trubner & Co., New York: Harcourt, Brace and Company. Ch. VII.

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

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

Von Neumann, John and Oskar Morgenstern (1944), *Theory of Games and Economic Behavior*. Princeton University Press.