# Fragility of Asymptotic Agreement under Bayesian Learning*

Daron Acemoglu, Victor Chernozhukov, and Muhamet Yildiz†

March 2008.

**Abstract**

Under the assumption that individuals know the conditional distributions of signals given the payoff-relevant parameters, existing results conclude that, as individuals observe infinitely many signals, their beliefs about the parameters will eventually merge. We first show that these results are fragile when individuals are uncertain about the signal distributions: given any such model, a vanishingly small individual uncertainty about the signal distributions can lead to a substantial (non-vanishing) amount of differences between the asymptotic beliefs. We then characterize the conditions under which a small amount of uncertainty leads only to a small amount of asymptotic disagreement. According to our characterization, this is the case if the uncertainty about the signal distributions is generated by a family with "rapidly-varying tails" (such as the normal or the exponential distributions). However, when this family has "regularly-varying tails" (such as the Pareto, the log-normal, and the t-distributions), a small amount of uncertainty leads to a substantial amount of asymptotic disagreement.

**Keywords:** asymptotic disagreement, Bayesian learning, merging of opinions.

**JEL Classification:** C11, C72, D83.

*An earlier version of this paper was circulated under the title "Learning and Disagreement in an Uncertain World". We thank Eduardo Faingold, Greg Fisher, Bengt Holmstrom, Matthew Jackson, Drew Fudenberg, Alessandro Lizzeri, Giuseppe Moscarini, Marciano Sinscalchi, Robert Wilson and seminar participants at MIT, the University of British Columbia, University of Illinois at Urbana-Champaign, Stanford and Yale for useful comments and suggestions.

†Department of Economics, Massachusetts Institute of Technology.

# 1 Introduction

The common prior assumption is one of the cornerstones of modern economic analysis. Most models postulate that the players in a game have the "same model of the world," or more precisely, that they have a common prior about the game form and payoff distributions—for example, they all agree that some *state* (payoff-relevant parameter vector) $\theta$ is drawn from a known distribution $G$, even though each may also have additional information about some components of $\theta$. The typical justification for the common prior assumption comes from *learning*; individuals, through their own experiences and the communication of others, will have access to a history of events informative about the state $\theta$, and this process will lead to "agreement" among individuals about the distribution of $\theta$. A strong version of this view is expressed in Savage (1954, p. 48) as the statement that a Bayesian individual, who does not assign zero probability to "the truth," will learn it eventually as long as the signals are informative about the truth. An immediate implication of this result is that two individuals who observe the same sequence of signals will ultimately agree, even if they start with very different priors.

Despite this powerful intuition, disagreement is the rule rather than the exception in practice. For example, there is typically considerable disagreement even among economists working on a certain topic. Similarly, there are deep divides about religious beliefs within populations with shared experiences. In most cases, the source of disagreement does not seem to be differences in observations or experiences. Instead, individuals appear to interpret the available data differently. For example, an estimate showing that subsidies increase investment is interpreted very differently by two economists starting with different priors. An economist believing that subsidies have no effect on investment appears more likely to judge the data or the methods leading to this estimate to be unreliable and thus to attach less importance to this evidence.

In this paper, we investigate the outcome of learning about an underlying state by two Bayesian individuals with different priors when they are possibly *uncertain* about the conditional distributions (or interpretations) of signals. This leads to a potential *identification problem*, as the same long-run frequency of signals may result from different combinations of payoff-relevant variables and different interpretations of the signals. Hence, even though the individuals will learn the asymptotic frequency of signals, they may not always be able to infer the state $\theta$, and initial differences in their beliefs may translate into differences in asymptotic beliefs. When the amount of uncertainty is small, the identification problem is also small in the sense that each individual finds it highly likely that he will eventually assign high probability to the true state. One may then expect that the asymptotic beliefs of the two individuals about the underlying states should be close as well. If so, the common prior assumption would be a good approximation when players have a long common experience and face only a small amount of uncertainty about how the signals are related to the states.

Our focus in this paper is to investigate the validity of this line of argument. In particular, we study whether asymptotic agreement is *continuous at certainty*. Asymptotic agreement is continuous at certainty if a small amount of uncertainty leads only to a small amount of disagreement asymptotically. Our main result shows that asymptotic agreement is discontinuous at certainty for *every* model: for every model there is a vanishingly small amount of uncertainty that is sufficient for each individual to assign nearly probability 1 that they will asymptotically hold significantly different beliefs about the underlying state. This result implies that learning foundations of common prior are not as strong as one might have thought.

Before explaining our main result and its intuition, it is useful to provide some details about the environment we study. Two individuals with given priors observe a sequence of signals, $\{s_t\}_{t=0}^{n}$, and form their posteriors about the state $\theta$. The only non-standard feature of the environment is that these individuals may be uncertain about the distribution of signals conditional on the underlying state. In the simplest case where the state and the signal are binary, e.g., $\theta \in \{A, B\}$, and $s_t \in \{a, b\}$, this implies that $\Pr(s_t = \theta \mid \theta) = p_\theta$ is not a known number, but individuals also have a prior over $p_\theta$, say given by a cumulative distribution function $F_\theta^i$ for each agent $i = 1, 2$. We refer to $F_\theta^i$ as individual's *subjective probability distribution* and to its density $f_\theta^i$ as *subjective (probability) density*. This distribution, which can differ among individuals, is a natural measure of their uncertainty about the informativeness of signals. When subjective probability distributions are non-degenerate, individuals will have some latitude in interpreting the sequence of signals they observe. The presence of subjective probability distributions over the interpretation of the signals introduces an identification problem and implies that, in contrast to the standard learning environments, asymptotic learning and asymptotic agreement are not guaranteed. In particular, when each $F_\theta^i$ has a full support for each $\theta$, there will not be asymptotic learning or asymptotic agreement. Lack of asymptotic agreement implies that two individuals with different priors observing the *same* sequence of signals will reach different posterior beliefs even after observing infinitely many signals. Moreover, individuals attach *ex ante probability 1* that they will disagree after observing the sequence of signals.

Now consider a family of subjective density functions, $\{f^i_{\theta,m}\}$, becoming increasingly concentrated around a single point—thus converging to certainty. When $m$ is large (and uncertainty is small), each individual is almost certain that he will assign nearly probability 1 to the true value of $\theta$. Despite this approximate asymptotic learning, our main result shows that asymptotic agreement may fail. In particular, for any $(p^1_A, p^1_B, p^2_A, p^2_B)$, we can construct sequences of $f^i_{\theta,m}$ that become more and more concentrated around $p^i_\theta$, but with a significant amount of asymptotic disagreement at almost all sample paths for all $m$. This establishes that asymptotic agreement is *discontinuous* at certainty for every model.

Under additional continuity and uniform convergence assumptions on the family $\{f^i_{\theta,m}\}$, we characterize the families of subjective densities under which asymptotic agreement is continuous at certainty. When $f^1_{\theta,m}$ and $f^2_{\theta,m}$ are concentrated around the same $p$, these additional assumptions ensure that asymptotic agreement is continuous at certainty. Otherwise, continuity of asymptotic agreement depends on the tail properties of the family of subjective density functions $\{f^i_{\theta,m}\}$. When this family has *regularly-varying tails* (such as the Pareto or the log-normal distributions), even under the additional regularity conditions that ensure uniform convergence, there will be a substantial amount of asymptotic disagreement. When $\{f^i_{\theta,m}\}$ has rapidly-varying tails (such as the normal distribution), asymptotic agreement will be continuous at certainty.

The intuition for this result is as follows. When the amount of uncertainty is small, each individual believes that he will learn the state $\theta$, but he may still believe that the other individual will fail to learn. Whether or not he believes this depends on how an individual reacts when a frequency of signals different from the one he expects with "almost certainty" occurs. If this "surprise" event ensures that the individual learns $\theta$ (as it does in the case of learning under certainty), then each individual will expect the other to learn when the frequency of signals under their model of the world is realized and thus attaches probability arbitrarily close to 1 that they will asymptotically agree. This is what happens when the family $\{f^i_{\theta,m}\}$ has rapidly-varying tails. However, when the family $\{f^i_{\theta,m}\}$ has regularly-varying (thick) tails, an unexpected frequency of signals will prevent the individual from learning (because he interprets this as a possibility likely even near certainty due to the thick tails). In this case, each individual expects the limiting frequencies to be consistent with his model and the other individual not to learn the true state $\theta$, and concludes that there will be significant asymptotic disagreement.

Lack of asymptotic agreement has important implications for a range of economic situations. We illustrate some of these in a companion paper by studying a number of simple environments where two individuals observe the same sequence of signals before or while playing a game (Acemoglu, Chernozhukov and Yildiz, 2008).

Our results cast doubt on the idea that the common prior assumption may be justified by learning. They imply that in many environments, even when there is little uncertainty so that each individual believes that he will learn the true state, Bayesian learning does not necessarily imply agreement about the relevant parameters. Consequently, the strategic outcomes may be significantly different from those in the common-prior environments.[1] Whether this assumption is warranted therefore depends on the specific setting and what type of information individuals are trying to glean from the data.

Relating our results to the famous Blackwell-Dubins (1962) theorem may help clarify their essence. This theorem shows that when two agents agree on zero-probability events (i.e., their priors are absolutely continuous with respect to each other), asymptotically, they will make the same predictions about future frequencies of signals. Our results do not contradict this theorem, since we impose absolute continuity. Instead, as pointed out above, our results rely on the fact that agreeing about future frequencies is not the same as agreeing about the underlying payoff-relevant variables, because of the identification problem that arises in the presence of uncertainty.[2] This identification problem leads to different possible interpretations of the same signal sequence by individuals with different priors. In most economic situations, what is important is not future frequencies of signals but some payoff-relevant parameter. For example, what is relevant for economists trying to evaluate a policy is not the frequency of estimates on the effect of similar policies from other researchers, but the impact of this specific policy when (and if) implemented. Similarly, what may be relevant in trading assets is not the frequency of information about the dividend process, but the actual dividend that the asset will pay. Thus,

[1] For previous arguments on whether game-theoretic models should be formulated with all individuals having a common prior, see, for example, Aumann (1986, 1998) and Gul (1998). Gul (1998), for instance, questions whether the common prior assumption makes sense when there is no ex ante stage.

[2] In this respect, our paper is also related to Kurz (1994, 1996), who considers a situation in which agents agree about long-run frequencies, but their beliefs fail to merge because of the non-stationarity of the world.

many situations in which individuals need to learn about a parameter or state that will determine their ultimate payoff as a function of their action falls within the realm of the analysis here. Our main result shows that even when this identification problem is negligible for individual learning, its implications to asymptotic agreement may be large.

In this respect, our work differs from papers, such as Freedman (1963, 1965) and Miller and Sanchirico (1999), that question the applicability of the absolute continuity assumption in the Blackwell-Dubins theorem in statistical and economic settings (see also Diaconis and Freedman, 1986, Stinchcombe, 2005). Similarly, a number of important theorems in statistics, for example, Berk (1966), show that when individuals place zero probability on the true data generating process, limiting posteriors will have their support on the set of all identifiable values (though they may fail to converge to a limiting distribution). Our results are different from those of Berk both because in our model individuals always place positive probability on the truth and also because we provide a tight characterization of the conditions for lack of asymptotic learning and agreement.[3] In addition, neither Berk nor any other paper that we are aware of investigates whether asymptotic agreement is continuous at certainty, which is the main focus of our paper.

Our paper is also related to recent independent work by Cripps, Ely, Mailath and Samuelson (2006), who study the conditions under which there will be "common learning" by two agents observing correlated private signals. Cripps, et al. focus on a model in which individuals start with *common priors* and then learn from *private signals* under *certainty* (though they note that their results could be extended to the case of non-common priors). They show that individual learning ensures "approximate common knowledge" when the signal space is finite, but not necessarily when it is infinite. In contrast, we focus on the case in which the agents start with *heterogenous priors* and learn from *public signals* under *uncertainty* or under *approximate certainty.* Since all signals are public in our model, there is no difficulty in achieving approximate common knowledge.[4]

[3] In dynamic games, another source of non-learning (and thus lack of convergence to common prior) is that some subgames are never visited along the equilibrium path and thus players do not observe information that contradict their beliefs about payoffs in these subgames (see, Fudenberg and Levine, 1993, Fudenberg and Kreps, 1995). Our results differ from those in this literature, since individuals fail to learn or fail to reach agreement despite the fact that they receive signals about all payoff-relevant variables.

[4] Put differently, we as whether a player thinks that the other player will learn, whereas Cripps et al. ask whether a player $i$ thinks that the other player $j$ thinks that $i$ thinks that $j$ thinks that ... a player will learn.

The rest of the paper is organized as follows. Section 2 provides a number of preliminary results focusing on the simple case of two states and two signals. Section 3 contain our main results at characterizing the conditions under which agreement is continuous at certainty. Section 4 provides generalizations of these results to an environment with $K$ states and $L \geq K$ signals. Section 5 concludes, while the Appendix contains the proofs omitted from the text.

# 2 The Two-State Model and Preliminary Results

## 2.1 Environment

We start with a two-state model with binary signals. This model is sufficient to establish all our main results in the simplest possible setting. These results are generalized to arbitrary number of states and signal values in Section 4.

There are two individuals, denoted by $i = 1$ and $i = 2$, who observe a sequence of signals $\{s_t\}_{t=0}^n$ where $s_t \in \{a, b\}$. The underlying state is $\theta \in \{A, B\}$, and agent $i$ assigns ex ante probability $\pi^i \in (0, 1)$ to $\theta = A$. The individuals believe that, given $\theta$, the signals are exchangeable, i.e., they are independently and identically distributed with an unknown distribution.[5] That is, the probability of $s_t = a$ given $\theta = A$ is an unknown number $p_A$; likewise, the probability of $s_t = b$ given $\theta = B$ is an unknown number $p_B$—as shown in the following table:

| | $A$ | $B$ |
|---|---|---|
| $a$ | $p_A$ | $1 - p_B$ |
| $b$ | $1 - p_A$ | $p_B$ |

Our main departure from the standard models is that we allow the individuals to be uncertain about $p_A$ and $p_B$. We denote the cumulative distribution function of $p_\theta$ according to individual $i$—namely, his *subjective probability distribution*—by $F_\theta^i$. In the standard models, $F_\theta^i$ is degenerate (Dirac) and puts probability 1 at some $\hat{p}_\theta^i$. In contrast, for most of the analysis, we will impose the following assumption:

[5] See, for example, Billingsley (1995). If there were only one state, then our model would be identical to De Finetti's canonical model (see, for example, Savage, 1954). In the context of this model, De Finetti's theorem provides a Bayesian foundation for classical probability theory by showing that exchangeability (i.e., invariance under permutations of the order of signals) is equivalent to having an independent identical unknown distribution and implies that posteriors converge to long-run frequencies. De Finetti's decomposition of probability distributions is extended by Jackson, Kalai and Smorodinsky (1999) to cover cases without exchangeability.

**Assumption 1** *For each $i$ and $\theta$, $F_\theta^i$ has a continuous, non-zero and finite density $f_\theta^i$ over* $[0,1]$.

The assumption implies that $F_\theta^i$ has *full support* over $[0,1]$. As discussed in Remark 2, Assumption 1 is stronger than necessary for our results, but simplifies the exposition. In addition, throughout we assume that $\pi^1$, $\pi^2$, $F_\theta^1$ and $F_\theta^2$ are known to both individuals.[6]

We consider infinite sequences $s \equiv \{s_t\}_{t=1}^\infty$ of signals and write $S$ for the set of all such sequences. The posterior belief of individual $i$ about $\theta$ after observing the first $n$ signals $\{s_t\}_{t=1}^n$ is

$$\phi_n^i(s) \equiv \Pr^i\left(\theta = A \mid \{s_t\}_{t=1}^n\right),$$

where $\Pr^i\left(\theta = A \mid \{s_t\}_{t=1}^n\right)$ denotes the posterior probability that $\theta = A$ given a sequence of signals $\{s_t\}_{t=1}^n$ under prior $\pi^i$ and subjective probability distribution $F_\theta^i$. Since the sequence of signals, $s$, is generated by an exchangeable process, the order of the signals does not matter for the posterior. It only depends on

$$r_n(s) \equiv \#\{t \le n | s_t = a\},$$

the number of times $s_t = a$ out of first $n$ signals.[7] By the strong law of large numbers, $r_n(s)/n$ converges to some $\rho(s) \in [0,1]$ almost surely according to both individuals. Defining the set

$$\bar{S} \equiv \{s \in S : \lim_{n\to\infty} r_n(s)/n \text{ exists}\}, \tag{1}$$

this observation implies that $\Pr^i\left(s \in \bar{S}\right) = 1$ for $i = 1, 2$. We will often state our results for all sample paths $s$ in $\bar{S}$, which equivalently implies that these statements are true almost surely or with probability 1. Now, a straightforward application of the Bayes rule gives

$$\phi_n^i(s) = \frac{1}{1 + \frac{1-\pi^i}{\pi^i}\frac{\Pr^i(r_n|\theta=B)}{\Pr^i(r_n|\theta=A)}}, \tag{2}$$

[6] Since our purpose is to understand whether learning justifies the common prior assumption, we do not assume a common prior, allowing agents to have differing beliefs even when the beliefs are commonly known.

[7] Given the definition of $r_n(s)$, the probability distribution $\Pr^i$ on $\{A, B\} \times S$ is

$$\begin{aligned}\Pr^i\left(E^{A,s,n}\right) &\equiv \pi^i \int_0^1 p^{r_n(s)} (1-p)^{n-r_n(s)} f_A^i(p)\,dp, \text{ and} \\ \Pr^i\left(E^{B,s,n}\right) &\equiv \left(1-\pi^i\right) \int_0^1 (1-p)^{r_n(s)} p^{n-r_n(s)} f_B^i(p)\,dp\end{aligned}$$

at each event $E^{\theta,s,n} = \{(\theta, s') | s_t' = s_t \text{ for each } t \le n\}$, where $s \equiv \{s_t\}_{t=1}^\infty$ and $s' \equiv \{s_t'\}_{t=1}^\infty$.

where $\Pr^i(r_n|\theta)$ is the probability of observing the signal $s_t = a$ exactly $r_n$ times out of $n$ signals with respect to the distribution $F_\theta^i$.

The following lemma provides a useful formula for $\phi_\infty^i(s) \equiv \lim_{n\to\infty} \phi_n^i(s)$ for all sample paths $s$ in $\bar{S}$ and also introduces the concept of the asymptotic likelihood ratio. Both the formula and the asymptotic likelihood ratio are crucial for our analyses throughout the paper.

**Lemma 1** *Suppose Assumption 1 holds. Then for all $s \in \bar{S}$,*

$$\phi_\infty^i(\rho(s)) \equiv \lim_{n\to\infty} \phi_n^i(s) = \frac{1}{1 + \frac{1-\pi^i}{\pi^i} R^i(\rho(s))}, \tag{3}$$

*where $\rho(s) = \lim_{n\to\infty} r_n(s)/n$, and $\forall \rho \in [0,1]$,*

$$R^i(\rho) \equiv \frac{f_B^i(1-\rho)}{f_A^i(\rho)} \tag{4}$$

*is the asymptotic likelihood ratio.*

**Proof.** See the Appendix. ■

In equation (4), $R^i(\rho)$ is the *asymptotic likelihood ratio* of observing frequency $\rho$ of $a$ when the true state is $B$ versus when it is $A$. Lemma 1 states that, asymptotically, individual $i$ uses this likelihood ratio and Bayes rule to compute his posterior beliefs about $\theta$.

In the statements about learning, without loss of generality, we suppose that in reality $\theta = A$. The two questions of interest for us are:

1. **Asymptotic learning:** whether $\Pr^i\left(\lim_{n\to\infty} \phi_n^i(s) = 1 | \theta = A\right) = 1$ for $i = 1, 2$.

2. **Asymptotic agreement:** whether $\Pr^i\left(\lim_{n\to\infty} \left|\phi_n^1(s) - \phi_n^2(s)\right| = 0\right) = 1$ for $i = 1, 2$.

Notice that both asymptotic learning and agreement are defined in terms of the ex ante probability assessments of the two individuals. Therefore, asymptotic learning implies that an individual believes that he or she will ultimately learn the truth, while asymptotic agreement implies that both individuals believe that their assessments will eventually converge.[8]

[8] We formulate asymptotic learning and agreement in terms of each individual's initial probability measure so as not to take a position on what the "objective" for "true" probability measure is. Under Assumption 1, asymptotic learning and agreement occur iff the corresponding limits hold for almost all long run frequencies $\rho(s) \in [0,1]$ under Lebesgue measure, which has also an "objective" meaning.

## 2.2 Asymptotic Learning and Agreement with Full Identification

In this subsection, we provide a number of preliminary results on the conditions under which there will be asymptotic learning and agreement. These results will be used as the background for the investigation of the continuity of asymptotic agreement at certainty in the next section. Throughout this subsection we focus on environments where *Assumption 1 does not hold.*

The following result generalizes Savage's (1954) well-known result on asymptotic learning and agreement. Savage's Theorem, which is then stated as Corollary 1 below, is the basis of the argument that Bayesian learning will push individuals towards common beliefs and priors. Let us denote the *support* of a distribution $F$ by $\text{supp}F$ and define $\inf(\text{supp}F)$ to be the infimum of the set $\text{supp}F$ (i.e., the largest $p$ such that $F(p) = 0$). Also let us define the threshold value

$$\hat{\rho}(p_A, p_B) \equiv \frac{\log(p_B/(1-p_A))}{\log(p_B/(1-p_A)) + \log(p_A/(1-p_B))} \in (1-p_B, p_A). \tag{5}$$

(For future reference, this is the unique solution to the equation $p_A^{\rho}(1-p_A)^{1-\rho} = p_B^{1-\rho}(1-p_B)^{\rho}$.)

**Theorem 1** ***(Generalized Asymptotic Learning and Agreement)*** *Define $\hat{\rho}(p_A, p_B)$ as in (5). Assume that for each $\theta$ and $i$, $p_{\theta,i} = \inf(\text{supp}F_{\theta}^{i}) \in (1/2, 1)$ and $1 - p_{B,i} \neq \hat{\rho}(p_{A,j}, p_{B,j}) \neq p_{A,i}$ for all $i \neq j$. Then for all $i \neq j$,*

1. $\Pr^i\left(\lim_{n\to\infty} \phi_n^i(s) = 1 | \theta = A\right) = 1$;
2. $\Pr^i\left(\lim_{n\to\infty} \left|\phi_n^1(s) - \phi_n^2(s)\right| = 0\right) = 1$ *if and only if* $1 - p_{B,i} < \hat{\rho}(p_{A,j}, p_{B,j}) < p_{A,i}$.

**Proof.** Both parts of the theorem are a consequence of the following claim.

**Claim 1** *For any $s \in \bar{S}$,*

$$\lim_{n\to\infty} \phi_n^i(s) = \begin{cases} 1 & \text{if } \rho(s) > \hat{\rho}(p_{A,i}, p_{B,i}) \\ 0 & \text{if } \rho(s) < \hat{\rho}(p_{A,i}, p_{B,i}), \end{cases} \tag{6}$$

*where $\rho(s) = \lim r_n(s)/n$.*

**(Proof of Claim)** Let

$$R_n^i(r_n) \equiv \frac{\Pr^i(r_n|\theta = B)}{\Pr^i(r_n|\theta = A)} = \frac{\int (1-p)^{r_n} p^{n-r_n} dF_B^i}{\int p^{r_n}(1-p)^{n-r_n} dF_A^i}.$$

Take any $\rho > \hat{\rho}(p_{A,i}, p_{B,i})$. Since $1 - p_{B,i} < p_{A,i}$,

$$(1-p_{B,i})^{\rho} p_{B,i}^{1-\rho} < p_{A,i}^{\rho}(1-p_{A,i})^{1-\rho}. \tag{7}$$

The function $p^{\rho}(1-p)^{1-\rho}$ is continuous and concave in $p$, and reaches its maximum at $p = \rho$. Then, (7) implies that there exists $\varepsilon > 0$ and $\hat{p} > p_{A,i}$ such that for all $\tilde{p} \in \operatorname{supp} F_B^i$, $p \in [p_{A,i}, \hat{p}]$,$r_n/n \in (\rho - \varepsilon, \rho + \varepsilon)$,

$$(1-\tilde{p})^{r_n} \tilde{p}^{n-r_n} \leq (1-p_{B,i})^{r_n} p_{B,i}^{n-r_n} < \hat{p}^{r_n}(1-\hat{p})^{n-r_n} \leq p^{r_n}(1-p)^{n-r_n}. \tag{8}$$

The first inequality in (8) implies that

$$\int (1-p)^{r_n} p^{n-r_n} dF_B^i \leq (1-p_{B,i})^{r_n} p_{B,i}^{n-r_n}. \tag{9}$$

On the other hand, the last inequality in (8) implies that

$$\int p^{r_n}(1-p)^{n-r_n} dF_A^i \geq \int_{p \leq \hat{p}} p^{r_n}(1-p)^{n-r_n} dF_A^i \geq F_A^i(\hat{p})\hat{p}^{r_n}(1-\hat{p})^{n-r_n}, \tag{10}$$

where the first inequality follows from non-negativity of $p^{r_n}(1-p)^{n-r_n}$. By dividing the left-hand side [right-hand side] of (9) by the left-hand side [right-hand side] of (10), we therefore obtain

$$0 \leq R_n^i(r_n) \leq \frac{1}{F_A^i(\hat{p})} \left( \frac{(1-p_{B,i})^{r_n/n} p_{B,i}^{1-r_n/n}}{\hat{p}^{r_n/n}(1-\hat{p})^{1-r_n/n}} \right)^n. \tag{11}$$

Equation (6) follows from (11). By (8), when $r_n/n \in [\rho - \varepsilon/2, \rho + \varepsilon/2]$, the expression in parenthesis in (11) is smaller than 1, so that the right-hand side converges to 0 as $n \to \infty$ and $r_n/n \to \rho$. Therefore, $R_n^i(r_n) \to 0$, and thus $\phi_n^i(s) \to 1$. The same argument (switching $A$ and $B$) implies that $\phi_n^i(s) \to 0$ when $\rho < \hat{\rho}(p_{A,i}, p_{B,i})$. □

**(Part 1)** Since $p_{\theta,i} = \inf(\operatorname{supp} F_\theta^i) \in (1/2, 1)$, (6) implies that conditional on $\theta = A$, agent $i$ assigns probability 1 to the event that $s \in \bar{S}$ and $\rho(s) \geq p_{A,i} > \hat{\rho}(p_{A,i}, p_{B,i})$, where the last inequality follows from (5). This implies that $\Pr^i\left(\lim_{n\to\infty} \phi_n^i(s) = 1 | \theta = A\right) = 1$.

**(Part 2: Sufficiency)** We prove that $1 - p_{B,i} < \hat{\rho}(p_{A,j}, p_{B,j}) < p_{A,i}$ implies asymptotic agreement. Suppose $\hat{\rho}(p_{A,j}, p_{B,j}) < p_{A,i}$. Then, conditional on $\theta = A$, (6) implies that $\phi_n^j(s)$ also converges to 1, and therefore $\left|\phi_n^1(s)-\phi_n^2(s)\right| \to 0$. Next, when $\hat{\rho}(p_{A,j},p_{B,j}) > 1-p_{B,i}$, conditional on $\theta = B$, $\phi_n^j(s) \to 0$ and $\phi_n^i(s) \to 0$. This establishes that $\left|\phi_n^1(s)-\phi_n^2(s)\right| \to 0$ and proves sufficiency.

**(Part 2: Necessity)** We prove that asymptotic agreement implies the inequality $1-p_{B,i} < \hat{\rho}(p_{A,j},p_{B,j}) < p_{A,i}$. Suppose the inequality does not hold, and consider the case $p_{A,i} < \hat{\rho}(p_{A,j},p_{B,j})$. Then, $i$ assigns strictly positive probability to the event that $r_n(s)/n \to \rho(s) \in [p_{A,i}, \hat{\rho}(p_{A,j},p_{B,j}))$. But (6) implies $\phi_n^i(s) \to 1$ and $\phi_n^j(s) \to 0$, so that $\left|\phi_n^1(s)-\phi_n^2(s)\right| \to 1$. Therefore, the beliefs diverge almost surely. The argument for the case where $\hat{\rho}(p_{A,j},p_{B,j}) < 1-p_{B,i}$ is analogous and completes the proof of the theorem. ∎

Theorem 1 shows that under the "full identification assumption" that $p_{\theta,i} > 1/2$ for each $\theta$ and $i$, asymptotic learning always obtains. Furthermore, asymptotic agreement depends on the lowest value $p_{\theta,i}$ of $p_\theta$ to which individual $i = 1,2$ assigns positive probability.

An immediate corollary is Savage's theorem.

**Corollary 1** ***(Savage's Theorem)*** *Assume that each $F_\theta^i$ puts probability 1 on $\hat{p}_\theta$ for some $\hat{p}_\theta > 1/2$, i.e., $F_\theta^i(\hat{p}_\theta) = 1$ and $F_\theta^i(p) = 0$ for each $p < \hat{p}_\theta$. Then, for each $i =$ 1,2,*

1. $\Pr^i\left(\lim_{n\to\infty} \phi_n^i(s) = 1 | \theta = A\right) = 1.$
2. $\Pr^i\left(\lim_{n\to\infty} \left|\phi_n^1(s)-\phi_n^2(s)\right| = 0\right) = 1.$

It is useful to spell out the intuition for Theorem 1 and Corollary 1. Let us start with the latter. Corollary 1 states that when the individuals know the conditional distributions of the signals (and hence they agree what those distributions are), they will learn the truth with experience (almost surely as $n \to \infty$) and two individuals observing the same sequence will necessarily come to agree what the underlying state, $\theta$, is. A simple intuition for this result is that the underlying state $\theta$ is *fully identified* from the limiting frequencies, so that both individuals can infer the underlying state from the observation of the limiting frequencies of signals.

However, there is more to this corollary than this simple intuition. Each individual is sure that they will be confronted either with a limiting frequency of $a$ signals equal to $\hat{p}_A$, in which case they will conclude that $\theta = A$, or they will observe a limiting frequency of $1-\hat{p}_B$, and they will conclude that $\theta = B$; and they attach zero probability to the events that they will observe a different asymptotic frequency. What happens if an individual observes a frequency $\rho$ of signals different from $\hat{p}_A$ and $1-\hat{p}_B$ in a large sample of size $n$? The answer to this question will provide the intuition for some of the results that we will present in the next section. Observe that this event has zero probability under the individual's beliefs at the limit $n=\infty$. However, for $n<\infty$ he will assign a strictly positive (but small) probability to such a frequency of signals resulting from *sampling variation*. Moreover, it is straightforward to see that there exists a unique $\hat{\rho}(\hat{p}_A,\hat{p}_B)\in(1-\hat{p}_B,\hat{p}_A)$ given by (5) above such that when $\rho>\hat{\rho}(\hat{p}_A,\hat{p}_B)$, the required sampling variation that leads to $\rho$ under $\theta=B$ is *infinitely greater* (as $n\to\infty$) than the one under $\theta=A$. Consequently, when $\rho>\hat{\rho}(\hat{p}_A,\hat{p}_B)$, the individual will asymptotically assign probability 1 to the event that $\theta=A$. Conversely, when $\rho<\hat{\rho}(\hat{p}_A,\hat{p}_B)$, he will assign probability 1 to $\theta=B$.

The intuition for Theorem 1 is very similar to that of Corollary 1. The assumption that $\inf(\text{supp}F_\theta^i)\in(1/2,1)$ generalizes the assumption that $\hat{p}_\theta>1/2$ in Corollary 1, and is sufficient to ensure that asymptotically each individual will learn the payoff-relevant state $\theta$, and also expects both himself and the other player to do so before observing the sequence of signals. In particular, similar to the intuition for Corollary 1, when individual $i$ observes a frequency $\rho\in(1-p_{B,i},p_{A,i})$, he presumes that this has resulted from sampling variation, and decides whether frequency $\rho$ is more likely under $\theta=A$ or under $\theta=B$. In particular, for each $\theta$, the lowest sampling variation that leads to $\rho$ is attained at $p_{\theta,i}$, and the asymptotic beliefs depend only on how large these variations are. When $\rho>\hat{\rho}(p_{A,i},p_{B,i})$ (and as $n\to\infty$) the necessary sampling variation is infinitely smaller under $\theta=A$ than under $\theta=B$. Consequently, the individual believes with probability 1 that $\theta=A$. Conversely, when $\rho<\hat{\rho}(p_{A,i},p_{B,i})$, he believes with probability 1 that $\theta=B$. Whether there will be asymptotic agreement then purely depends on whether and how different the cutoff values $\hat{\rho}(p_{A,1},p_{B,1})$ and $\hat{\rho}(p_{A,2},p_{B,2})$ are. When they are close, both individuals will interpret the limiting frequency of signals, $\rho$, similarly, even when this is a frequency to which they initially assigned zero probability, and will reach asymptotic agreement.[9]

The next corollary highlights a range of conditions other than those in Corollary 1 that, according to part 2 of Theorem 1, are sufficient for asymptotic agreement.

[9]In contrast, if these cutoff values were far apart, so that $\hat{\rho}(p_{A,j},p_{B,j})\notin(1-p_{B,i},p_{A,i})$, both players would assign positive probability to the event that their beliefs would diverge to the extremes and we would thus have $\lim_{n\to\infty}\left|\phi_n^1(s)-\phi_n^2(s)\right|=1$.

**Corollary 2** ***(Sufficient Conditions for Asymptotic Agreement)*** *Suppose that* $p_{\theta,i} = \inf(\text{supp}F_\theta^i) \in (1/2, 1)$. *Then, there is asymptotic agreement whenever any one of the following conditions hold:*

1. certainty (with symmetry): *each* $F_\theta^i$ *puts probability 1 on some* $\hat{p}^i > 1/2$*;*

2. symmetric support: $\text{supp}F_A^i = \text{supp}F_B^i$ *for each* $i$*;*

3. common support: $\text{supp}F_\theta^1 = \text{supp}F_\theta^2$ *for each* $\theta$*.*

**Proof.** Part 1 of the corollary is a special case of part 2. Under symmetric support assumption, we have $\hat{\rho}(p_{A,i}, p_{B,i}) = 1/2$ for each $i$, so that part 2 of the corollary follows from part 2 of Theorem 1. Finally, part 3 of the corollary follows from the fact that under the common support assumption $\hat{\rho}(p_{A,j}, p_{B,j}) = \hat{\rho}(p_{A,i}, p_{B,i}) \in (1 - p_{B,i}, p_{A,i})$. ■

Corollary 2 shows that various reasonable conditions ensure asymptotic agreement. Asymptotic agreement is implied, for example, by certainty, symmetric support or common support assumptions. In particular, certainty (with symmetry), which corresponds to both individuals believing that limiting frequencies have to be $\hat{p}^i$ or $1 - \hat{p}^i$ (but with $\hat{p}^1 \neq \hat{p}^2$) is sufficient for asymptotic agreement. In this case, each individual is certain about what the limiting frequency will be and therefore believes that the frequency expected by the other individual will not be realized (creating a discrepancy between that individual's initial belief and observation). Nevertheless, with the same reasoning as in the discussion following Corollary 1, each individual also believes that the other individual will ascribe this discrepancy to sampling variation and reach the same conclusion as himself. This is sufficient for asymptotic agreement.

Theorem 1 and Corollary 2 therefore show that results on asymptotic learning and agreement are substantially more general than Savage's original theorem. Nevertheless, these results do rely on the feature that $F_\theta^i(1/2) = 0$ for each $i = 1{,}2$ and each $\theta$ (thus implicitly imposing that Assumption 1 does not hold). This feature implies that both individuals attach zero probability to a range of possible models of the world—i.e., they are certain that $p_\theta$ cannot be less than 1/2. There are two reasons for considering situations in which this is not the case. First, the preceding discussion illustrates why assigning zero probability to certain models of the world is important; it enables individuals to ascribe any frequency of signals that are unlikely under these models to sampling variability. This kind of inference may be viewed as somewhat unreasonable, since individuals are reaching very strong conclusions based on events that have vanishingly small probabilities (since sampling variability vanishes as $n \to \infty$). Second, our motivation of investigating learning under uncertainty suggests that individuals may attach positive (albeit small) probabilities to all possible values of $p_\theta$. This latter feature is the essence of Assumption 1 (the "full support" requirement).

## 2.3 Failure of Asymptotic Learning and Agreement with Full Support

We next impose Assumption 1 and show that under the more general circumstances where $F_\theta^i$ has full support, there will be neither asymptotic learning nor asymptotic agreement.

**Theorem 2** ***(Lack of Asymptotic Learning and Agreement)*** *Under Assumption 1,*

1. $\Pr^i\left(\lim_{n\to\infty} \phi_n^i(s) \neq 1 | \theta = A\right) = 1$ *for* $i = 1{,}2$*;*
2. $\Pr^i\left(\lim_{n\to\infty} \left|\phi_n^1(s) - \phi_n^2(s)\right| \neq 0\right) = 1$ *whenever* $\pi^1 \neq \pi^2$ *and* $F_\theta^1 = F_\theta^2$ *for each* $\theta \in \{A, B\}$*.*

**Proof.** Since $f_B^i(1-\rho(s)) > 0$ and $f_A(\rho(s))$ is finite, $R^i(\rho(s)) > 0$. Hence, by Lemma 1, $\phi_\infty^i(\rho(s)) \neq 1$ for each $s$, establishing the first part. To see the second part, note that, by Lemma 1, for any $s \in \bar{S}$,

$$\phi_\infty^1(\rho(s)) = \phi_\infty^2(\rho(s)) \text{ if and only if } \frac{1-\pi^1}{\pi^1} R^1(\rho(s)) = \frac{1-\pi^2}{\pi^2} R^2(\rho(s)). \tag{12}$$

Since $\pi^1 \neq \pi^2$ and $F_\theta^1 = F_\theta^2$, this implies that for each $s \in \bar{S}$, $\phi_\infty^1(s) \neq \phi_\infty^2(s)$, and thus $\Pr^i\left(\left|\phi_\infty^1(s) - \phi_\infty^2(s)\right| \neq 0\right) = 1$ for $i = 1, 2$. ■

**Remark 1** The assumption that $F_\theta^1 = F_\theta^2$ in this theorem is adopted for simplicity. We can see from (12) that even in the absence of this condition, there will typically be no asymptotic agreement. Theorem 6 in Section 4 states a more general version of this result for the case of multidimensional state and signals, and shows how the assumption that $F_\theta^1 = F_\theta^2$ can be relaxed significantly.

**Remark 2** Assumption 1 is considerably stronger than the necessary conditions for Theorem 2. It is adopted only for simplicity. It can be verified that for lack of asymptotic learning it is sufficient (but not necessary) that the measures generated by the distribution functions $F_A^i(p)$ and $F_B^i(1-p)$ be absolutely continuous with respect to each other. Similarly, for lack of asymptotic agreement, it is sufficient (but not necessary) that the measures generated by $F_A^1(p)$, $F_B^1(1-p)$, $F_A^2(p)$ and $F_B^i(1-p)$ be absolutely continuous with respect each other. For example, if both individuals believe that $p_A$ is either 0.3 or 0.7 (with the latter receiving greater probability) and that $p_B$ is also either 0.3 or 0.7 (with the former receiving greater probability), then there will be neither asymptotic learning nor asymptotic agreement. Throughout we use Assumption 1 both because it simplifies the notation and because it is a natural assumption when we turn to the analysis of asymptotic agreement as the amount of uncertainty vanishes.

Theorem 2 contrasts with Theorem 1 and implies that, with probability 1, each individual will fail to learn the true state. The second part of the theorem states that if the individuals' prior beliefs about the state differ (but they interpret the signals in the same way), then their posteriors will eventually disagree, and moreover, they will both attach probability 1 to the event that their beliefs will eventually diverge. Put differently, this implies that there is "agreement to eventually disagree" between the two individuals, in the sense that they both believe ex ante that after observing the signals they will fail to agree.

Intuitively, when Assumption 1 (in particular, the full support feature) holds, an individual is never sure about the exact interpretation of the sequence of signals he observes and will update his views about $p_\theta$ (the informativeness of the signals) as well as his views about the underlying state. For example, even when signal $a$ is more likely in state $A$ than in state $B$, a very high frequency of $a$ will not necessarily convince him that the true state is $A$, because he may infer that the signals are not as reliable as he initially believed, and they may instead be biased towards $a$. Therefore, the individual never becomes certain about the state, which is captured by the fact that $R^i(\rho)$ defined in (4) never takes the value zero or infinity. Consequently, as shown in (3), his posterior beliefs will be determined by his prior beliefs about the state and also by $R^i$, which tells us how the individual updates his beliefs about the informativeness of the signals as he observes the signals. When two individuals interpret the informativeness of the signals in the same way (i.e., $R^1 = R^2$), the differences in their priors will always be reflected in their posteriors.

In contrast, if an individual were certain about the informativeness of the signals (i.e., if $i$ were sure that $p_\theta = p_\theta^i$ for some $p_\theta^i > 1/2$) as in Theorem 1 and Corollary 2, then he would never question the informativeness of the signals, even when the limiting frequency of $a$ converges to a value different from $p_A^i$ or $1 - p_B^i$, and would interpret such discrepancies as resulting from sampling variation. This would be sufficient for asymptotic agreement when $p_A^i = p_B^i$. The full support assumption in Assumption 1 prevents this type of reasoning and ensures asymptotic disagreement.

# 3 Main Results

In this section, we present our main results concerning the potential discontinuity of asymptotic agreement at certainty. More precisely, we investigate whether as the amount of uncertainty about the interpretation of the signals disappears and we recover the standard model of learning under certainty, the amount of asymptotic disagreement vanishes continuously. We will show that this is not the case, so that one can perturb the standard model of learning under certainty sightly and obtain a model in which there is substantial asymptotic disagreement. We first show that asymptotic agreement is discontinuous at certainty in every model, including the canonical model of learning under certainty, where both individuals share the same beliefs regarding the conditional signal distributions (Theorem 3). We then restrict our perturbations to a class that embodies strong continuity and uniform convergence assumptions. Within this class of perturbations, we characterize the conditions under which asymptotic agreement will be continuous at certainty (Theorem 5).

For any $\hat{p} \in [0,1]$, write $\delta_{\hat{p}}$ for the Dirac distribution that puts probability 1 on $p = \hat{p}$; i.e., $\delta_{\hat{p}}(p) = 1$ if $p \geq \hat{p}$ and $\delta_{\hat{p}}(p) = 0$ otherwise.

Let $\{F_{\theta,m}^i\}_{m\in\mathbb{N},i\in N,\theta\in\Theta}$ ($\{F_{\theta,m}^i\}$ for short) denote an arbitrary sequence of subjective probability distributions converging to a Dirac distribution $\delta_{p_\theta^i}$ for each $(i,\theta)$ as $m \to \infty$:

$$\lim_{m\to\infty} F_{\theta,m}^i(p) = \begin{cases} 1 & \text{if } p > p_\theta^i \\ 0 & \text{if } p < p_\theta^i. \end{cases} \tag{13}$$

(We will simply say that $\{F_{\theta,m}^i\}$ converges to $\delta_{p_\theta^i}$). Throughout it is implicitly assumed that there is asymptotic agreement under $\delta_{p_\theta^i}$ (as in Corollaries 1 and 2). Therefore, as $m \to \infty$, uncertainty about the interpretation of the signals disappears and we converge to a world of asymptotic agreement. We write $\Pr^{i,m}$ for the ex ante probability under $(F_{A,m}^i, F_{B,m}^i)$ and $\phi_{\infty,m}^i$ for the asymptotic posterior belief that $\theta = A$ under $(F_{A,m}^i, F_{B,m}^i)$. Evidently, as $\{F_{\theta,m}^i\}$ converges to $\delta_{p_\theta^i}$, each individual becomes increasingly convinced that he will learn the true state, so that learning is *continuous at certainty.* More formally, for all $\varepsilon > 0$,

$$\lim_{m\to\infty} \Pr^{i,m}\left(\phi^i_{\infty,m} > 1-\varepsilon|\theta = A\right) = 1.$$

This implies that when a model of learning under certainty is perturbed, deviations from full learning will be small and each individual will attach a probability arbitrarily close to 1 that he will eventually learn the payoff-relevant state variable $\theta$. We next define the continuity of asymptotic agreement at certainty.

**Definition 1** *For any given family* $\{F^i_{\theta,m}\}$*, we say that* ***asymptotic agreement is continuous at certainty under*** $\{F^i_{\theta,m}\}$*, if for all* $\varepsilon > 0$ *and for each* $i = 1, 2$*,*

$$\lim_{m\to\infty} \Pr^{i,m}\left(\left|\phi^1_{\infty,m} - \phi^2_{\infty,m}\right| < \varepsilon\right) = 1.$$

*We say that* ***asymptotic agreement is continuous at certainty at*** $(p^1_A, p^1_B, p^2_A, p^2_B)$ *if it is continuous at certainty under every family* $\{F^i_{\theta,m}\}$ *converging to* $\delta_{p^i_\theta}$*.*

Thus, continuity at certainty requires that as the family of subjective probability distributions converge to a Dirac distribution (at which there is asymptotic agreement), the ex ante probability that both individuals assign to the event that they will agree asymptotically becomes arbitrarily close to 1. Hence, asymptotic agreement is *discontinuous at certainty* at $(p^1_A, p^1_B, p^2_A, p^2_B)$ if there exists a family $\{F^i_{\theta,m}\}$ converging to $\delta_{\hat{p}^i_\theta}$ and $\varepsilon > 0$ such that

$$\lim_{m\to\infty} \Pr^{i,m}\left(\left|\phi^1_{\infty,m} - \phi^2_{\infty,m}\right| > \varepsilon\right) > 0$$

for $i = 1, 2$. We will next define a stronger notion of discontinuity.

**Definition 2** *We say that* ***asymptotic agreement is strongly discontinuous at certainty under*** $\{F^i_{\theta,m}\}$ *if there exists* $\varepsilon > 0$ *such that*

$$\lim_{m\to\infty} \Pr^{i,m}\left(\left|\phi^1_{\infty,m} - \phi^2_{\infty,m}\right| > \varepsilon\right) = 1$$

*for* $i = 1, 2$*. We say that* ***asymptotic agreement is strongly discontinuous at certainty at*** $(p^1_A, p^1_B, p^2_A, p^2_B)$ *if it is strongly discontinuous at certainty under some family* $\{F^i_{\theta,m}\}$ *converging to* $\delta_{p^i_\theta}$*.*

Strong discontinuity requires that even as we approach the world of learning under certainty, asymptotic agreement will fail with probability approximately equal to 1 according to both individuals.

## 3.1 Discontinuity of Asymptotic Agreement

The next theorem establishes the strong discontinuity of asymptotic agreement at certainty.

**Theorem 3** ***(Strong Discontinuity of Asymptotic Agreement)*** *Asymptotic agreement is strongly discontinuous at every* $(p_A^1, p_B^1, p_A^2, p_B^2)$ *with* $p_\theta^i \in (1/2, 1)$ *for all* $(\theta, i)$*. Moreover, if* $\pi^1 \neq \pi^2$*, then there exist* $\{F_\theta^i\}$ *converging to* $\delta_{p_\theta^i}$ *and* $\tilde{Z} > 0$ *such that*

$$\left|\phi^1_{\infty,m}(\rho(s)) - \phi^2_{\infty,m}(\rho(s))\right| > \tilde{Z} \text{ for all } m \in \mathbb{N} \text{ and } s \in \bar{S}.$$

The proof of this theorem is provided below. Note that when $p_\theta^1 = p_\theta^2 = \hat{p}_\theta$ for each $\theta$, the limiting world is the canonical learning model (under certainty) described in Savage's Theorem (Corollary 1): both individuals are certain that the probability of observing signal $s = a$ is $\hat{p}_A > 1/2$ if the state is $\theta = A$ and $1 - \hat{p}_B$ if the state is $\theta = B$ (i.e., each $F_\theta^i$ puts probability 1 on $\hat{p}_\theta$). Therefore, this theorem establishes strong discontinuity at certainty for the canonical learning model; even when we are arbitrarily close to this world of certainty, the asymptotic gap in beliefs is bounded away from zero. The condition $p_\theta^i \in (1/2, 1)$ is not needed (see Theorem 7 below). The proof is based on the following example.

**Example 1** For some small $\epsilon, \lambda \in (0, 1)$, each individual $i$ thinks that with probability $1 - \epsilon$, $p_\theta$ is in a $\lambda$-neighborhood of some $\hat{p}_\theta^i > (1 + \lambda)/2$, but with probability $\epsilon$, the signals are not informative. More precisely, for $\hat{p}_\theta^i > (1 + \lambda)/2$ and $\lambda < |\hat{p}_\theta^1 - \hat{p}_\theta^2|$, we have

$$f_\theta^i(p) = \begin{cases} \epsilon + (1-\epsilon)/\lambda & \text{if } p \in (\hat{p}_\theta^i - \lambda/2, \hat{p}_\theta^i + \lambda/2) \\ \epsilon & \text{otherwise} \end{cases} \tag{14}$$

for each $\theta$ and $i$. Now, by (4), the asymptotic likelihood ratio is

$$R^i(\rho(s)) = \begin{cases} \frac{\epsilon\lambda}{1-\epsilon(1-\lambda)} & \text{if } \rho(s) \in D_A^i \equiv (\hat{p}_A^i - \lambda/2, \hat{p}_A^i + \lambda/2) \\ \\ \frac{1-\epsilon(1-\lambda)}{\epsilon\lambda} & \text{if } \rho(s) \in D_B^i \equiv (1 - \hat{p}_B^i - \lambda/2, 1 - \hat{p}_B^i + \lambda/2) \\ 1 & \text{otherwise.} \end{cases}$$

This and other relevant functions are plotted in Figure 1 for $\epsilon \to 0$, $\lambda \to 0$. The likelihood ratio $R^i(\rho(s))$ is 1 when $\rho(s)$ is small, takes a very high value at $1 - \hat{p}_B^i$, goes down to 1 afterwards, becomes nearly zero around $\hat{p}_A^i$, and then jumps back to 1. By Lemma 1, $\phi_\infty^i(s)$ will also be non-monotone: when $\rho(s)$ is small, the signals are not

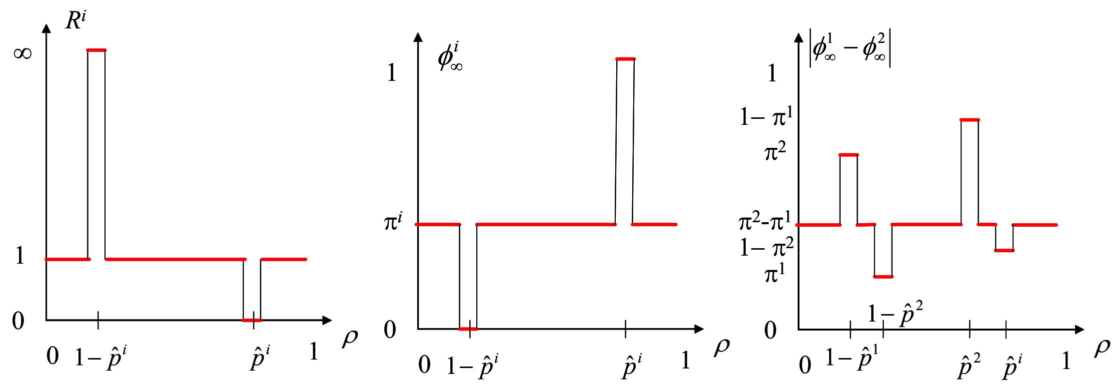


Figure 1: The three panels show, respectively, the approximate values of $R^i(\rho)$, $\phi^i_\infty$, and $\left|\phi^1_\infty - \phi^2_\infty\right|$ as $\epsilon \to 0$, for $\hat{p}^i_A = \hat{p}^i_B = \hat{p}^i$.

informative, thus $\phi^i_\infty(s)$ is the same as the prior, $\pi^i$. In contrast, around $1 - \hat{p}^i_B$, the signals become very informative suggesting that the state is $B$, thus $\phi^i_\infty(s) \cong 0$. After this point, the signals become uninformative again and $\phi^i_\infty(s)$ goes back to $\pi^i$. Around $\hat{p}^i_A$, the signals are again informative, but this time favoring state $A$, so $\phi^i_\infty(s) \cong 1$. Finally, signals again become uninformative and $\phi^i_\infty(s)$ falls back to $\pi^i$. Intuitively, when $\rho(s)$ is around $1 - \hat{p}^i_B$ or $\hat{p}^i_A$, the individual assigns very high probability to the true state, but outside of this region, he sticks to his prior, concluding that the signals are not informative.

The first important observation is that even though $\phi^i_\infty$ is equal to the prior for a large range of limiting frequencies, as $\epsilon \to 0$ and $\lambda \to 0$ each individual attaches probability 1 to the event that he will learn $\theta$. This is because as illustrated by the discussion after Theorem 1, as $\epsilon \to 0$ and $\lambda \to 0$, each individual becomes convinced that the limiting frequencies will be either $1 - \hat{p}^i_B$ or $\hat{p}^i_A$.

However, asymptotic learning is considerably weaker than asymptotic agreement. Each individual also understands that since $\lambda < |\hat{p}^1_\theta - \hat{p}^2_\theta|$, when the long-run frequency is in a region where he learns that $\theta = A$, the other individual will conclude that the signals are uninformative and adhere to his prior belief. Consequently, he expects the posterior beliefs of the other individual to be always far from his. Put differently, as $\epsilon \to 0$ and $\lambda \to 0$, each individual believes that he will learn the value of $\theta$ himself but that the other individual will fail to learn, thus attaches probability 1 to the event that they disagree. This can be seen from the third panel of Figure 1; at each sample path in $\bar{S}$, at least one of the individuals will fail to learn, and the difference between their limiting posteriors will be uniformly higher than the following "objective" bound

$$\tilde{z} = \min\left\{\pi^1, \pi^2, 1-\pi^1, 1-\pi^2, \left|\pi^1 - \pi^2\right|\right\}.$$

When $\pi^1 = 1/3$ and $\pi^2 = 2/3$, this bound is equal to 1/3. In fact, the belief of each individual regarding potential disagreement can be greater than this; each individual believes that he will learn but the other individual will fail to do so. Consequently, for each $i$, $\Pr^i\left(\left|\phi^1_\infty(s) - \phi^2_\infty(s)\right| \geq \bar{Z}\right) \geq 1-\epsilon$, where as $\epsilon \to 0$, $\bar{Z} \to z \equiv \min\{\pi^1, \pi^2, 1-\pi^1, 1-\pi^2\}$. This "subjective" bound can be as high as 1/2.

**Proof of Theorem 3.** We only consider the case $p^1_\theta \geq p^2_\theta$ for $\theta = A, B$; the other cases are identical. In Example 1, for each $m$, take $\epsilon = \lambda = \bar{\epsilon}/m$, $\hat{p}^1_\theta = p^1_\theta + \lambda$, and $\hat{p}^2_\theta = p^2_\theta - \lambda$ where $\bar{\epsilon}$ is such that $1 - \phi^i_\infty(s) < (1-\pi^j)/2$ for $\rho(s) \in D^i_A$ and $\phi^i_\infty(s) < \pi^j/2$ for $\rho(s) \in D^i_B$ whenever $\epsilon = \lambda \leq \bar{\epsilon}$. Such $\bar{\epsilon}$ exists (by asymptotic learning of $i$). By construction, each $F^{i,m}_\theta$ converges to $\delta_{p^i_\theta}$, and $|\hat{p}^1_\theta - \hat{p}^2_\theta| > \lambda$ for each $\theta$. To complete the proof, pick $\bar{Z} = z/2 > 0$. By choice of $\bar{\epsilon}$, $\left|\phi^1_{\infty,m}(s) - \phi^2_{\infty,m}(s)\right| > \bar{Z}$ whenever $\rho(s) \in D^i_A \cup D^i_B$. But $\Pr^{i,m}(\rho(s) \in D^i_A \cup D^i_B) = \epsilon(1-\lambda)$, which goes to 1 as $m \to \infty$. Therefore,

$$\lim_{m\to\infty} \Pr^{i,m}\left(\left|\phi^1_{\infty,m} - \phi^2_{\infty,m}\right| > \bar{Z}\right) = 1. \quad (15)$$

To prove the last statement in the theorem, pick $\tilde{Z} = \tilde{z}/2$, which is positive when $\pi^1 \neq \pi^2$. ■

In the example (and thus in the proof of Theorem 3), the likelihood ratio $R^i(\rho(s))$ and the asymptotic beliefs $\phi^i_\infty(s)$ are *non-monotone* in the frequency $\rho(s)$. This is a natural outcome of uncertainty on conditional signal distributions (see the discussion at the end of Section 2 and Figure 2 below). When $R^i$ is monotone and the amount of uncertainty is small, at each state one of the individuals assigns high probability that both of them will learn the true state and consequently asymptotic disagreement will be small. Nevertheless, asymptotic agreement is still discontinuous at uncertainty when we impose the monotone likelihood ratio property. This is shown in the next theorem.

**Theorem 4** ***(Discontinuity of Asymptotic Agreement under Monotonicity)*** *For any $\hat{p}^i_A, \hat{p}^i_B > 1/2$, $i \in \{1,2\}$, and $\pi^1, \pi^2 \in (0,1)$, there exist a family $\{F^i_{\theta,m}\}$ and $\bar{Z} > 0$ such that:*

1. *for each* $\theta \in \{A, B\}$ *and* $i = 1, 2$, $F^i_{\theta,m}$ *converges to* $\delta_{\hat{p}^i_\theta}$*;*

2. *the likelihood ratio* $R^i_m(\rho)$ *is nonincreasing in* $\rho$ *for each* $i$ *and* $m$*, and*

3. *for each* $i$*,*

$$\lim_{m\to\infty} \Pr^{i,m}\left(\left|\phi^1_{\infty,m} - \phi^2_{\infty,m}\right| > \bar{Z}\right) > 0. \tag{16}$$

**Proof.** See the Appendix. ■

The monotonicity of the likelihood ratio has weakened the conclusion of Theorem 3, so that the limit in (16) is no longer equal to 1, so that asymptotic agreement is discontinuous at certainty, but not strongly so.

Note that in Theorems 3 and 4 the families $\{F^i_{\theta,m}\}$ leading to the discontinuity of asymptotic agreement induce discontinuous likelihood ratios. This is not crucial for the results, however, since smooth approximations to $F^i_{\theta,m}$ would ensure continuity of the likelihood ratios as well. What is important is that the likelihood ratios under families $\{F^i_{\theta,m}\}$ *does not converge uniformly* (instead, convergence is pointwise). We next impose a uniform convergence assumption (as well as additional strong continuity assumptions) and characterize the conditions for discontinuity of asymptotic agreement at certainty.

## 3.2 Agreement and Disagreement with Uniform Convergence

In this subsection, we consider a class of families $\{F^i_{\theta,m}\}$ converging *uniformly* to the Dirac distribution $\delta_{\hat{p}^i}$ for some $\hat{p}^i \in (1/2, 1)$ and show that whether there is discontinuity of asymptotic agreement at certainty depends on the *tail properties* of $\{F^i_{\theta,m}\}$.

We start our analysis by defining the family $\{F^i_{\theta,m}\}$, with a corresponding family of subjective probability density functions $\{f^i_{\theta,m}\}$. The family is parameterized by a *determining* density function $f$. We impose the following conditions on $f$:

(i) $f$ is strictly positive and symmetric around zero;

(ii) there exists $\bar{x} < \infty$ such that $f(x)$ is decreasing for all $x \geq \bar{x}$;

(iii)
$$\tilde{R}(x, y) \equiv \lim_{m\to\infty} \frac{f(mx)}{f(my)} \tag{17}$$

exists in $[0, \infty]$ at all $(x, y) \in \mathbb{R}^2_+$.

Conditions (i) and (ii) are natural and serve to simplify the notation. Condition (iii) introduces the function $\tilde{R}(x, y)$, which will arise naturally in the study of asymptotic agreement and has a natural meaning in asymptotic statistics (see Definitions 1 and 2 below).

In order to vary the amount of uncertainty, we consider mappings of the form $x \mapsto (x - y)/m$, which scale down the real line around $y$ by the factor $1/m$. The family of subjective densities for individuals' beliefs about $p_A$ and $p_B$, $\{f^i_{\theta,m}\}$, will be determined by $f$ and the transformation $x \mapsto (x - \hat{p}^i)/m$.[10] In particular, we consider the following family of densities

$$f^i_{\theta,m}(p) = c^i(m) f\left(m\left(p - \hat{p}^i\right)\right) \tag{18}$$

for each $\theta$ and $i$ where $c^i(m) \equiv 1/\int_0^1 f(m(p - \hat{p}^i))\,dp$ is a correction factor to ensure that $f^i_{\theta,m}$ is a proper probability density function on $[0,1]$ for each $m$. In this family of subjective densities, the uncertainty about $p_A$ is scaled down by $1/m$, and $f^i_{\theta,m}$ converges to the Dirac distribution $\delta_{\hat{p}^i}$ as $m \to \infty$, so that individual $i$ becomes sure about the informativeness of the signals in the limit.

The next theorem characterizes the class of determining functions $f$ for which the resulting family of the subjective densities $\{f^i_{\theta,m}\}$ leads to approximate asymptotic agreement as the amount of uncertainty vanishes.

**Theorem 5** ***(Characterization)*** *Consider the family $\{F^i_{\theta,m}\}$ defined in (18) for some $\hat{p}^i > 1/2$ and $f$, satisfying conditions (i)-(iii) above. Assume that $f(mx)/f(my)$ uniformly converges to $\tilde{R}(x,y)$ over a neighborhood of $(\hat{p}^1 + \hat{p}^2 - 1, |\hat{p}^1 - \hat{p}^2|)$.*

1. *If $\tilde{R}(\hat{p}^1 + \hat{p}^2 - 1, |\hat{p}^1 - \hat{p}^2|) = 0$, then agreement is continuous at certainty under $\{F^i_{\theta,m}\}$.*

2. *If $\tilde{R}(\hat{p}^1 + \hat{p}^2 - 1, |\hat{p}^1 - \hat{p}^2|) \neq 0$, then agreement is strongly discontinuous at certainty under $\{F^i_{\theta,m}\}$.*

**Proof.** Both parts of the theorem are proved using the following claim.

**Claim 2** $\lim_{m\to\infty}\left(\phi^i_{\infty,m}(\hat{p}^i) - \phi^j_{\infty,m}(\hat{p}^i)\right) = 0$ *if and only if* $\tilde{R}(\hat{p}^1 + \hat{p}^2 - 1, |\hat{p}^1 - \hat{p}^2|) = 0$ *(where $\phi^i_{\infty,m}(\hat{p}^i)$ denotes beliefs evaluated under sample paths with $\rho = \hat{p}^i$).*

[10] This formulation assumes that $\hat{p}^i_A$ and $\hat{p}^i_B$ are equal. We can easily assume these to be different, but do not introduce this generality here to simplify the exposition. Theorem 8 allows for such differences in the context of the more general model with multiple states and multiple signals.

**(Proof of Claim)** Let $R_m^i(\rho)$ be the asymptotic likelihood ratio as defined in (4) associated with subjective density $f_{\theta,m}^i$. One can easily check that $\lim_{m\to\infty} R_m^i(\hat{p}^i) = 0$. Hence, by (12), $\lim_{m\to\infty}\left(\phi_{\infty,m}^i(\hat{p}^i) - \phi_{\infty,m}^j(\hat{p}^i)\right) = 0$ if and only if $\lim_{m\to\infty} R_m^j(\hat{p}^i) = 0$. By definition,

$$\begin{aligned}\lim_{m\to\infty} R_m^j\left(\hat{p}^i\right) &= \lim_{m\to\infty} \frac{f\left(m\left(1-\hat{p}^1-\hat{p}^2\right)\right)}{f\left(m\left(\hat{p}^1-\hat{p}^2\right)\right)} \\ &= \tilde{R}\left(1-\hat{p}^1-\hat{p}^2, \hat{p}^1-\hat{p}^2\right) \\ &= \tilde{R}\left(\hat{p}^1+\hat{p}^2-1, \left|\hat{p}^1-\hat{p}^2\right|\right),\end{aligned}$$

where the last equality follows by condition (i), the symmetry of the function $f$. This establishes that $\lim_{m\to\infty} R_m^i(\hat{p}^i) = 0$ (and thus $\lim_{m\to\infty}\left(\phi_{\infty,m}^i(\hat{p}^i) - \phi_{\infty,m}^j(\hat{p}^i)\right) = 0$) if and only if $\tilde{R}(\hat{p}^1+\hat{p}^2-1, |\hat{p}^1-\hat{p}^2|) = 0$. $\square$

**(Proof of Part 1)** Take any $\epsilon > 0$ and $\delta > 0$, and assume that $\tilde{R}(\hat{p}^1+\hat{p}^2-1, |\hat{p}^1-\hat{p}^2|) = 0$. We will show that there exists $\bar{m} \in \mathbb{N}$ such that

$$\Pr^i\left(\lim_{n\to\infty}\left|\phi_{n,m}^1(s) - \phi_{n,m}^2(s)\right| > \epsilon\right) < \delta \qquad (\forall m > \bar{m}, i = 1, 2).$$

By Lemma 1, there exists $\epsilon' > 0$ such that $\phi_{\infty,m}^i(\rho(s)) > 1-\epsilon$ whenever $R^i(\rho(s)) < \epsilon'$. There also exists $x_0$ such that

$$\Pr^i\left(\rho(s) \in \left(\hat{p}^i - x_0/m, \hat{p}^i + x_0/m\right)|\theta = A\right) = \int_{-x_0}^{x_0} f(x)\,dx > 1-\delta. \qquad (19)$$

Let $\kappa = \min_{x\in[-x_0,x_0]} f(x) > 0$. Since $f$ monotonically decreases to zero in the tails (see (ii) above), there exists $x_1$ such that $f(x) < \epsilon'\kappa$ whenever $|x| > |x_1|$. Let $m_1 = (x_0+x_1)/(2\hat{p}^i - 1) > 0$. Then, for any $m > m_1$ and $\rho(s) \in (\hat{p}^i - x_0/m, \hat{p}^i + x_0/m)$, we have $|\rho(s) - 1 + \hat{p}^i| > x_1/m$, and hence

$$R_m^i(\rho(s)) = \frac{f\left(m\left(\rho(s)+\hat{p}^i-1\right)\right)}{f\left(m\left(\rho(s)-\hat{p}^i\right)\right)} < \frac{\epsilon'\kappa}{\kappa} = \epsilon'.$$

Therefore, for all $m > m_1$ and $\rho(s) \in (\hat{p}^i - x_0/m, \hat{p}^i + x_0/m)$, we have that

$$\phi_{\infty,m}^i(\rho(s)) > 1-\epsilon. \qquad (20)$$

Again, by Lemma 1, there exists $\epsilon'' > 0$ such that $\phi_{\infty,m}^j(\rho(s)) > 1-\epsilon$ whenever $R_m^j(\rho(s)) < \epsilon''$. Now, for each $\rho(s)$,

$$\lim_{m\to\infty} R_m^j(\rho(s)) = \tilde{R}\left(\rho(s)+\hat{p}^j-1, \left|\rho(s)-\hat{p}^j\right|\right). \qquad (21)$$

Moreover, by the uniform convergence assumption, there exists $\eta > 0$ such that $R_m^j(\rho(s))$ uniformly converges to $\tilde{R}(\rho(s) + \hat{p}^j - 1, |\rho(s) - \hat{p}^j|)$ on $(\hat{p}^i - \eta, \hat{p}^i + \eta)$ and

$$\tilde{R}\left(\rho(s) + \hat{p}^j - 1, \left|\rho(s) - \hat{p}^j\right|\right) < \epsilon''/2$$

for each $\rho(s)$ in $(\hat{p}^i - \eta, \hat{p}^i + \eta)$. Moreover, uniform convergence also implies that $\tilde{R}$ is continuous at $(\hat{p}^1 + \hat{p}^2 - 1, |\hat{p}^1 - \hat{p}^2|)$ (and in this part of the proof, by hypothesis, it takes the value 0). Hence, there exists $m_2 < \infty$ such that for all $m > m_2$ and $\rho(s) \in (\hat{p}^i - \eta, \hat{p}^i + \eta)$,

$$R_m^j(\rho(s)) < \tilde{R}\left(\rho(s) + \hat{p}^j - 1, \left|\rho(s) - \hat{p}^j\right|\right) + \epsilon''/2 < \epsilon''.$$

Therefore, for all $m > m_2$ and $\rho(s) \in (\hat{p}^i - \eta, \hat{p}^i + \eta)$, we have

$$\phi_{\infty,m}^j(\rho(s)) > 1 - \epsilon. \tag{22}$$

Set $\bar{m} \equiv \max\{m_1, m_2, \eta/x_0\}$. Then, by (20) and (22), for any $m > \bar{m}$ and $\rho(s) \in (\hat{p}^i - x_0/m, \hat{p}^i + x_0/m)$, we have $\left|\phi_{\infty,m}^i(\rho(s)) - \phi_{\infty,m}^j(\rho(s))\right| < \epsilon$. Then, (19) implies that $\Pr^i\left(\left|\phi_{\infty,m}^i(\rho(s)) - \phi_{\infty,m}^j(\rho(s))\right| < \epsilon|\theta = A\right) > 1 - \delta$. By the symmetry of $A$ and $B$, this establishes that $\Pr^i\left(|\phi_{\infty,m}^i(\rho(s)) - \phi_{\infty,m}^j(\rho(s))| < \epsilon\right) > 1 - \delta$ for $m > \bar{m}$.

**(Proof of Part 2)** We will find $\epsilon > 0$ such that for each $\delta > 0$, there exists $\bar{m} \in \mathbb{N}$ such that

$$\Pr^i\left(\lim_{n\to\infty}\left|\phi_{n,m}^1(s) - \phi_{n,m}^2(s)\right| > \epsilon\right) > 1 - \delta \qquad (\forall m > \bar{m}, i = 1, 2).$$

Since $\lim_{m\to\infty} R_m^j(\hat{p}^i) = \tilde{R}(\hat{p}^1 + \hat{p}^2 - 1, |\hat{p}^1 - \hat{p}^2|) > 0$, $\lim_{m\to\infty}\phi_{\infty,m}^j(\hat{p}^i) < 1$. We set $\epsilon = \left(1 - \lim_{m\to\infty}\phi_{\infty,m}^j(\hat{p}^i)\right)/2$ and use similar arguments to those in the proof of Part 1 to obtain the desired conclusion. ■

The main assumption in Theorem 5 is that the likelihood ratios $R_m^i(\rho(s))$ converge *uniformly* to a limiting likelihood ratio, given by $\tilde{R}$.[11] In what follows, we say that "noise vanishes uniformly" as a shorthand for the statement that the likelihood ratio $R_m^i(\rho(s))$ converges uniformly to the limiting likelihood ratio. Theorem 5 provides a complete characterization of the conditions for the continuity of asymptotic agreement at certainty under this *uniform convergence assumption*. In particular, this theorem shows that even when the likelihood ratios converge uniformly, asymptotic agreement may fail.

[11] Note that the limiting likelihood ratio $\tilde{R}$ is not related to the likelihood ratio that applies in the ("limiting") model without uncertainty.

In contrast Corollary 2 shows that that there will always be asymptotic agreement in the limit.

The theorem provides a simple condition on the tail of the distribution $f$ that determines whether the asymptotic difference between the posteriors will be small as the amount of uncertainty concerning the conditional distribution of signals vanishes "uniformly". This condition can be expressed as:

$$\tilde{R}(\hat{x}, \hat{y}) \equiv \lim_{m \to \infty} \frac{f(m\hat{x})}{f(m\hat{y})} = 0 \tag{23}$$

where $\hat{x} \equiv \hat{p}^1 + \hat{p}^2 - 1 > |\hat{p}^1 - \hat{p}^2| \equiv \hat{y}$. The theorem shows that if this condition is satisfied, then as uncertainty about the informativeness of the signals disappears the difference between the posteriors of the two individuals will become negligible. Notice that condition (23) is symmetric and does not depend on $i$.

Intuitively, condition (23) is related to the beliefs of one individual on whether the other individual will learn. As the amount of uncertainty concerning the conditional distributions vanishes, we always have that $\lim_{m \to \infty} R^i_m(\hat{p}^i) = 0$, so that each agent believes that he will learn the value of $\theta$ with probability 1. Asymptotic agreement (or lack thereof) depends on whether he also believes the other individual will learn the value of $\theta$. When $\tilde{R}(\hat{x}, \hat{y}) = 0$, an individual who expects a limiting frequency of $\hat{p}^2$ in the asymptotic distribution will still learn the true state when the limiting frequency is $\hat{p}^1$. Therefore, individual 1, who is almost certain that the limiting frequency will be $\hat{p}^1$, still believes that individual 2 will reach the same inference as himself. In contrast, when $\tilde{R}(\hat{x}, \hat{y}) \neq 0$, individual 1 is still certain that limiting frequency of signals will be $\hat{p}^1$ and thus expects to learn himself. However, he understands that, when $\tilde{R}(\hat{x}, \hat{y}) \neq 0$, an individual who expects a limiting frequency of $\hat{p}^2$ will fail to learn the true state when limiting frequency happens to be $\hat{p}^1$. Since he is almost certain that the limiting frequency will be $\hat{p}^1$ (or $1 - \hat{p}^1$), he expects the other agent not to learn the truth and thus he expects the disagreement between them to persist asymptotically.

The theorem exploits this result and the continuity of $\tilde{R}$ to show that the individuals attach probability arbitrarily close to 1 to the event that the asymptotic difference between their beliefs will disappear when (23) holds, and they attach probability 1 to asymptotic disagreement when (23) fails to hold. Thus the behavior of asymptotic beliefs as uncertainty vanishes "uniformly" are completely determined by condition (23), a condition on the tail of $f$.

When $\hat{y} > 0$ (i.e., when $\hat{p}^1 \neq \hat{p}^2$), condition (23) is a familiar condition in statistics. Whether it is satisfied depends on whether $f$ has rapidly-varying (thin) or regularly-varying (thick) tails:

**Definition 3** *A density function $f$ has* regularly-varying tails *if it has unbounded support and satisfies*

$$\lim_{m\to\infty} \frac{f(mx)}{f(m)} = H(x) \in \mathbb{R}$$

*for any $x > 0$.*

The condition in Definition 3 that $H(x) \in \mathbb{R}$ is relatively weak, but nevertheless has important implications. In particular, it implies that $H(x) \equiv x^{-\alpha}$ for $\alpha \in (0, \infty)$. This follows from the fact that in the limit, the function $H(\cdot)$ must be a solution to the functional equation $H(x)H(y) = H(xy)$, which is only possible if $H(x) \equiv x^{-\alpha}$ for $\alpha \in (0, \infty)$.[12] Moreover, Seneta (1976) shows that the convergence in Definition 3 holds locally uniformly, i.e., uniformly for $x$ in any compact subset of $(0, \infty)$. This implies that if a density $f$ has regularly-varying tails, then the assumptions imposed in Theorem 5 (in particular, the uniform convergence assumption) are satisfied. In fact, in this case, $\tilde{R}$ defined in (17) is given by

$$\tilde{R}(x,y) = \left(\frac{x}{y}\right)^{-\alpha},$$

and is everywhere continuous. As this expression suggests, densities with regularly-varying tails behave approximately like power functions in the tails; indeed a density $f(x)$ with regularly-varying tails can be written as $f(x) = \mathcal{L}(x)x^{-\alpha}$ for some *slowly-varying* function $\mathcal{L}$ (with $\lim_{m\to\infty} \mathcal{L}(mx)/\mathcal{L}(m) = 1$). Many common distributions, including the Pareto, log-normal, and t-distributions, have regularly-varying densities. When $f$ has regularly varying tails, $\tilde{R}(\hat{x}, \hat{y}) > 0$, and condition (23) cannot be satisfied. We also define:

**Definition 4** *A density function $f$ has* rapidly-varying tails *if it satisfies*

$$\lim_{m\to\infty} \frac{f(mx)}{f(m)} = x^{-\infty} \equiv \begin{cases} 0 & if \quad x > 1 \\ 1 & if \quad x = 1 \\ \infty & if \quad x < 1 \end{cases}$$

[12] To see this, note that since $\lim_{m\to\infty} (f(mx)/f(m)) = H(x) \in \mathbb{R}$, we have

$$H(xy) = \lim_{m\to\infty} \left(\frac{f(mxy)}{f(m)}\right) = \lim_{m\to\infty} \left(\frac{f(mxy)}{f(my)} \frac{f(my)}{f(m)}\right) = H(x)H(y).$$

See de Haan (1970) or Feller (1971).

*for any* $x > 0$.

As in Definition 3, the above convergence holds locally uniformly (uniformly in $x$ over any compact subset that excludes 1). Examples of densities with rapidly-varying tails include the exponential and the normal densities. When $f$ has rapidly varying tails $\tilde{R}(\hat{x}, \hat{y}) = (\hat{x}/\hat{y})^{-\infty} = 0$, and condition (23) is satisfied.

The next proposition formally states that under the assumptions that noise vanishes uniformly and set $\hat{p}^1 \neq \hat{p}^2$, whether agreement is continuous depends on whether the family of subjective densities converging to "certainty" has regularly or rapidly-varying tails:

**Proposition 1** ***(Tail Properties and Asymptotic Disagreement Under Uniform Convergence)*** *Suppose that the conditions in Theorem 5 are satisfied and that* $\hat{p}^1 \neq \hat{p}^2$. *Then,*

1. *If $f$ has regularly-varying tails, then agreement is continuous at certainty under* $\{F^i_{\theta,m}\}$.
2. *If $f$ has rapidly-varying tails, then agreement is strongly discontinuous at certainty under* $\{F^i_{\theta,m}\}$.

**Proof.** When $f$ has regularly or rapidly varying tails, uniform convergence assumption is satisfied, and the proposition follows from Definitions 3 and 4 and from Theorem 5. ■

Returning to the intuition above, Proposition 1 and the previous definitions make it clear that the failure of asymptotic agreement, under the assumption that $R^i_m(\rho)$ converges to $\tilde{R}$ uniformly, is related to disagreement between the two individuals about limiting frequencies, i.e., $\hat{p}^1 \neq \hat{p}^2$, together with sufficiently thick tails of the subjective probability distribution so that an individual who expects $\hat{p}^2$ should have sufficient uncertainty when confronted with a limiting frequency of $\hat{p}^1$. Along the lines of the intuition given there, this is sufficient for both individuals to believe that they will learn the true value of $\theta$ themselves, but that the other individual will fail to do so. Rapidly-varying tails imply that individuals become relatively certain of their model of the world and thus when individual $i$ observes a limiting frequency $\rho$ close to, but different from $\hat{p}^i$, he will interpret this as being driven by sampling variation and attach a high probability to $\theta = A$. This will guarantee asymptotic agreement between the two individuals. In contrast, with regularly-varying tails, even under the uniform convergence assumptions, limiting frequencies different from $\hat{p}^i$ will be interpreted not as sampling variation, but as potential evidence for $\theta = B$, preventing asymptotic agreement. The following example provides a simple illustration of part 1 of Proposition 1.

**Example 2** Let $f$ be the Pareto distribution and $\pi^1 = \pi^2 = 1/2$. The likelihood ratio is

$$R^i_m(\rho(s)) = \left(\frac{\rho(s) + \hat{p}^i - 1}{\rho(s) - \hat{p}^i}\right)^{-\alpha},$$

and the asymptotic probability of $\theta = A$ is

$$\phi^i_{\infty,m}(\rho(s)) = \frac{(\rho(s) - \hat{p}^i)^{-\alpha}}{(\rho(s) - \hat{p}^i)^{-\alpha} + (\rho(s) + \hat{p}^i - 1)^{-\alpha}}$$

for all $m$. (These expressions hold in the limit $m \to \infty$ under any $f$ with regularly-varying tails.) As illustrated in Figure 2, in this case $\phi^i_{\infty,m}$ is not monotone. To see the magnitude of asymptotic disagreement, consider $\rho(s) \cong \hat{p}^i$. In that case, $\phi^i_{\infty,m}(\rho(s))$ is approximately 1, and $\phi^j_{\infty,m}(\rho(s))$ is approximately $\hat{y}^{-\alpha}/(\hat{x}^{-\alpha} + \hat{y}^{-\alpha})$. Hence, both individuals believe that the difference between their asymptotic posteriors will be

$$\left|\phi^1_{\infty,m} - \phi^2_{\infty,m}\right| \cong \frac{\hat{x}^{-\alpha}}{\hat{x}^{-\alpha} + \hat{y}^{-\alpha}}.$$

This asymptotic difference is increasing with the difference $\hat{y} \equiv |\hat{p}^1 - \hat{p}^2|$, which corresponds to the difference in the individuals' views on which frequencies of signals are most likely. It is also clear from this expression that this asymptotic difference will converge to zero as $\hat{y} \to 0$ (i.e., as $\hat{p}^1 \to \hat{p}^2$).

The last statement in the example is in fact generally true when noise vanishes uniformly and $\tilde{R}$ is continuous. This is explored in the next proposition.

**Proposition 2** ***(Limits to Asymptotic Disagreement)*** *In Theorem 5, in addition, assume that $\tilde{R}$ is continuous on the set $D = \{(x,y) \mid -1 \leq x \leq 1, |y| \leq \bar{y}\}$ for some $\bar{y} > 0$. Then for every $\epsilon > 0$ and $\delta > 0$, there exist $\lambda > 0$ and $\bar{m} \in (0,\infty)$ such that whenever $|\hat{p}^1 - \hat{p}^2| < \lambda$,*

$$\Pr^i\left(\lim_{n\to\infty}\left|\phi^1_{n,m} - \phi^2_{n,m}\right| > \epsilon\right) < \delta \qquad (\forall m > \bar{m}, i = 1,2).$$

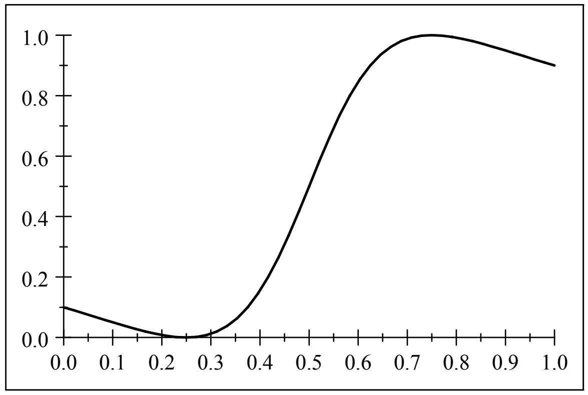


Figure 2: $\lim_{n\to\infty}\phi_n^i(s)$ for Pareto distribution as a function of $\rho(s)$ [$\alpha = 2$, $\hat{p}^i = 3/4$.]

**Proof.** To prove this proposition, we modify the proof of Part 1 of Theorem 5 and use the notation in that proof. Since $\tilde{R}$ is continuous on the compact set $D$ and $\tilde{R}(x, 0) = 0$ for each $x$, there exists $\lambda > 0$ such that $\tilde{R}(\hat{p}^1 + \hat{p}^2 - 1, |\hat{p}^1 - \hat{p}^2|) < \epsilon''/4$ whenever $|\hat{p}^1 - \hat{p}^2| < \lambda$. Fix any such $\hat{p}^1$ and $\hat{p}^2$. Then, by the uniform convergence assumption, there exists $\eta > 0$ such that $R_m^j(\rho(s))$ uniformly converges to $\tilde{R}(\rho(s) + \hat{p}^j - 1, |\rho(s) - \hat{p}^j|)$ on $(\hat{p}^i - \eta, \hat{p}^i + \eta)$ and

$$\tilde{R}\left(\rho(s) + \hat{p}^j - 1, \left|\rho(s) - \hat{p}^j\right|\right) < \epsilon''/2$$

for each $\rho(s)$ in $(\hat{p}^i - \eta, \hat{p}^i + \eta)$. The rest of the proof is identical to the proof of Part 1 in Theorem 5. ■

This proposition implies that in the case where noise vanishes uniformly and the individuals are almost certain about the informativeness of signals, any significant difference in their asymptotic beliefs must be due to differences in their subjective densities regarding the signal distribution—that is, $|\hat{p}^1 - \hat{p}^2|$ cannot be too small. In particular, when $\hat{p}^1 = \hat{p}^2$, we must have $\tilde{R}(\hat{x}, \hat{y}) = 0$, and thus, from Theorem 5, there will be convergence to asymptotic agreement. Notably, however, the requirement that $\hat{p}^1 = \hat{p}^2$ is rather strong. For example, Corollary 2 established that under certainty there is asymptotic agreement for all $\hat{p}^1, \hat{p}^2 > 1/2$.

In closing this section, let us reiterate that the key assumption in Proposition 2 is that $R_m^i(\rho)$ *uniformly* converges to a continuous limiting likelihood ratio $\tilde{R}$. In contrast, recall that Theorem 3 establishes that a slight uncertainty may lead to substantial asymptotic disagreement with nearly probability 1 even when $\hat{p}^1 = \hat{p}^2$. The crucial difference is that in Theorem 3 the likelihood ratios converge to a continuous limiting likelihood function pointwise, but *not uniformly.*

# 4 Generalizations

The previous section provided our main results in an environment with two states and two signals. In this section, we show that the results from the previous two sections generalize to an environment with $K \geq 2$ states and $L \geq K$ signals. All the proofs for this section are contained in the Appendix and to economize on space, we do not provide the analog of Theorem 1.

Suppose $\theta \in \Theta$, where $\Theta \equiv \{A_1, ..., A_K\}$ is a set containing $K \geq 2$ distinct elements. We refer to a generic element of the set by $A_k$. Similarly, let $s_t \in \{a_1, ..., a_L\}$, with $L \geq K$ signal values. As before, define $s \equiv \{s_t\}_{t=1}^{\infty}$, and for each $l = 1, ..., L$, let

$$r_{n,l}(s) \equiv \#\{t \leq n | s_t = a_l\}$$

be the number of times the signal $s_t = a_l$ out of first $n$ signals. Once again, the strong law of large numbers implies that, according to both individuals, for each $l = 1, ..., L$, $r_{n,l}(s)/n$ almost surely converges to some $\rho_l(s) \in [0,1]$ with $\sum_{l=1}^{L} \rho_l(s) = 1$. Define $\rho(s) \in \Delta(L)$ as the vector $\rho(s) \equiv (\rho_1(s), ..., \rho_L(s))$, where $\Delta(L) \equiv \left\{p = (p_1, \ldots, p_L) \in [0,1]^L : \sum_{l=1}^{L} p_l = 1\right\}$, and let the set $\bar{S}$ be

$$\bar{S} \equiv \{s \in S : \lim_{n \to \infty} r_{n,l}(s)/n \text{ exists for each } l = 1, ..., L\}. \tag{24}$$

With analogy to the two-state-two-signal model in Section 2, let $\pi_k^i > 0$ be the prior probability individual $i$ assigns to $\theta = A_K$, $\pi^i \equiv (\pi_1^i, ..., \pi_K^i)$, and $p_{\theta,l}$ be the frequency of observing signal $s = a_l$ when the true state is $\theta$. When players are certain about $p_{\theta,l}$'s as in usual models, immediate generalizations of Theorems 1 and 1 apply. With analogy to before, we define $F_\theta^i$ as the *joint subjective probability distribution* of conditional frequencies $p_\theta \equiv (p_{\theta,1}, ..., p_{\theta,L})$ according to individual $i$. Since our focus is learning under uncertainty, we impose an assumption similar to Assumption 1.

**Assumption 2** *For each $i$ and $\theta$, the distribution $F_\theta^i$ over $\Delta(L)$ has a continuous, non-zero and finite density $f_\theta^i$ over $\Delta(L)$.*

This assumption can be weakened along the lines discussed in Remark 2 above.

We also define $\phi^i_{k,n}(s) \equiv \Pr^i(\theta = A_k \mid \{s_t\}^n_{t=0})$ for each $k = 1, ..., K$ as the posterior probability that $\theta = A_k$ after observing the sequence of signals $\{s_t\}^n_{t=0}$, and

$$\phi^i_{k,\infty}(\rho(s)) \equiv \lim_{n\to\infty} \phi^i_{k,n}(s).$$

Given this structure, it is straightforward to generalize the results in Section 2. Let us now define the transformation $T_k : \mathbb{R}^K_+ \to \mathbb{R}^{K-1}_+$, such that

$$T_k(x) = \left(\frac{x_{k'}}{x_k}; k' \in \{1, ..., K\} \setminus k\right).$$

Here $T_k(x)$ is taken as a column vector. This transformation will play a useful role in the theorems and the proofs. In particular, this transformation will be applied to the vector $\pi^i$ of priors to determine the ratio of priors assigned the different states by individual $i$. Let us also define the norm $\|x\| = \max_l |x_l|$ for $x = (x_1, \ldots, x_L) \in \mathbb{R}^L$.

The next lemma generalizes Lemma 1 (proof omitted).

**Lemma 2** *Suppose Assumption 2 holds. Then for all $s \in \bar{S}$,*

$$\phi^i_{k,\infty}(\rho(s)) = \frac{1}{1 + \frac{\sum_{k' \neq k} \pi^i_{k'} f^i_{A_{k'}}(\rho(s))}{\pi^i_k f^i_{A_k}(\rho(s))}}.$$

Our first theorem in this section parallels Theorem 2 and shows that under Assumption 2 there will be lack of asymptotic learning, and under a relatively weak additional condition, there will also asymptotic disagreement.

**Theorem 6** ***(Generalized Lack of Asymptotic Learning and Agreement)*** *Suppose Assumption 2 holds for $i = 1,2$, then for each $k = 1, ..., K$, and for each $i = 1,2$,*

1. $\Pr^i\left(\phi^i_{k,\infty}(\rho(s)) \neq 1 | \theta = A_k\right) = 1$, *and*
2. $\Pr^i\left(\left|\phi^1_{k,\infty}(\rho(s)) - \phi^2_{k,\infty}(\rho(s))\right| \neq 0\right) = 1$ *whenever* $\Pr^i((T_k(\pi^1) - T_k(\pi^2))' T_k(f^i(\rho(s)) = 0) = 0$ *and* $F^1_\theta = F^2_\theta$ *for each* $\theta \in \Theta$.

The additional condition in part 2 of Theorem 6, that $\Pr^i((T_k(\pi^1) - T_k(\pi^2))' T_k(f^i(\rho(s)) = 0) = 0$, plays the role of differences in priors in Theorem 2 (here "$'$" denotes the transpose of the vector in question). In particular, if this condition did not hold, then at some $\rho(s)$, the relative asymptotic likelihood of some states could be the same according to two individuals with different priors and they would interpret at least some sequences of signals in a similar manner and achieve asymptotic agreement. It is important to note that the condition that $\Pr^i((T_k(\pi^1) - T_k(\pi^2))' T_k(f^i(\rho(s)) = 0) = 0$ is relatively weak and holds generically—i.e., if it did not hold, a small perturbation of $\pi^1$ or $\pi^2$ would restore it.[13] The Part 2 of Theorem 6 therefore implies that asymptotic disagreement occurs *generically.*

We next define continuity and discontinuity of asymptotic agreement at certainty in this more general case. A family of subjective probability distributions is again denoted by $\{F^i_{\theta,m}\}$. Throughout $\{F^i_{\theta,m}\}$ converge to a Dirac distribution $\delta_{p^i_\theta}$, where $p^i_\theta \in \Delta(L)$, and $\delta_{p^i_\theta}$ is such that there is asymptotic agreement (that is, there is asymptotic agreement when learning is under uncertainty). The corresponding *asymptotic* beliefs are denoted by $\phi^1_{k,\infty,m}$ and $\phi^2_{k,\infty,m}$ for $k = 1, ..., K$ and $m \in \mathbb{N}$.

**Definition 5** *Asymptotic agreement is* ***continuous at certainty under family*** $\{F^i_{\theta,m}\}$ *if for all* $\varepsilon > 0$*, for each* $k = 1, ..., K$ *and for each* $i = 1, 2$*,*

$$\lim_{m\to\infty} \Pr^{i,m}\left(\left|\phi^1_{k,\infty,m} - \phi^2_{k,\infty,m}\right| < \varepsilon\right) = 1.$$

*Asymptotic agreement is* ***continuous at certainty at*** $(p^1, p^2) \in \Delta(L)^{2K}$ *if it is continuous at certainty under all families* $\{F^i_{\theta,m}\}$ *converging to* $\delta_{p^i_\theta}$*.*

**Definition 6** *Asymptotic agreement is* ***strongly discontinuous at certainty under family*** $\{F^i_{\theta,m}\}$ *if there exists* $\varepsilon > 0$ *such that*

$$\lim_{m\to\infty} \Pr^{i,m}\left(\left|\phi^1_{k,\infty,m} - \phi^2_{k,\infty,m}\right| > \varepsilon\right) = 1$$

*for each* $k = 1, ..., K$ *and each* $i = 1, 2$*. Asymptotic agreement is* ***strongly discontinuous at certainty at*** $(p^1, p^2) \in \Delta(L)^{2K}$ *if asymptotic agreement is strongly discontinuous at certainty under some family* $\{F^i_{\theta,m}\}$ *converging to* $\delta_{p^i_\theta}$*.*

The next result generalizes Theorem 3:

[13] More formally, the set of solutions $\mathcal{S} \equiv \{(\pi^1, \pi^2, \rho) \in \Delta(L)^2 : (T_k(\pi^1) - T_k(\pi^2))' T_k(f^i(\rho)) = 0\}$ has Lebesgue measure 0. This is a consequence of the Preimage Theorem and Sard's Theorem in differential topology (see, for example, Guillemin and Pollack, 1974, pp. 21 and 39). The Preimage Theorem implies that if $y$ is a regular value of a map $f : X \to Y$, then $f^{-1}(y)$ is a submanifold of $X$ with dimension equal to $\dim X - \dim Y$. In our context, this implies that if 0 is a regular value of the map $(T_k(\pi^1) - T_k(\pi^2))' T_k(f^i(\rho))$, then the set $\mathcal{S}$ is a two dimensional submanifold of $\Delta(L)^3$ and thus has Lebesgue measure 0. Sard's theorem implies that 0 is generically a regular value.

**Theorem 7** ***(Generalized Strong Discontinuity of Asymptotic Agreement)*** *Asymptotic agreement is strongly discontinuous at each* $(p^1, p^2) \in \Delta(L)^{2K}$.

Towards generalizing Theorem 5, we now formally present the appropriate families of probability densities and introduce the necessary notation:

**Assumption 3** *For each* $\theta \in \Theta$ *and* $m \in \mathbb{N}$, *let the subjective density* $f^i_{\theta,m}$ *be defined by*

$$f^i_{\theta,m}(p) = c(i,\theta,m)\, f(m(p - \hat{p}(i,\theta)))$$

*where* $c(i,\theta,m) \equiv 1/\int_{p\in\Delta(L)} f(m(p-\hat{p}(i,\theta)))\,dp$, $\hat{p}(i,\theta) \in \Delta(L)$ *with* $\hat{p}(i,\theta) \neq \hat{p}(i,\theta')$ *whenever* $\theta \neq \theta'$, *and* $f : \mathbb{R}^L \to \mathbb{R}$ *is a positive, continuous probability density function that satisfies the following conditions:*

*(i)* $\lim_{h\to\infty} \max_{\{x:\|x\|\geq h\}} f(x) = 0$,

*(ii)*

$$\tilde{R}(x,y) \equiv \lim_{m\to\infty} \frac{f(mx)}{f(my)} \tag{25}$$

*exists at all* $x, y$, *and*

*(iii) convergence in (25) holds uniformly over a neighborhood of each* $(\hat{p}(i,\theta) - \hat{p}(j,\theta'), \hat{p}(i,\theta) - \hat{p}(j,\theta))$

Writing $\phi^i_{k,\infty,m}(\rho(s)) \equiv \lim_{n\to\infty} \phi^i_{k,n,m}(s)$ for the asymptotic posterior of individual $i$ with subjective density $f^i_{\theta,m}$, we are now ready to state the generalization of Theorem 5.

**Theorem 8** ***(Generalized Asymptotic Agreement and Disagreement Under Uniform Convergence)*** *Under Assumption 3, the following are true:*

1. *Suppose that* $\tilde{R}(\hat{p}(i,\theta) - \hat{p}(j,\theta'), \hat{p}(i,\theta) - \hat{p}(j,\theta)) = 0$ *for each distinct* $\theta$ *and* $\theta'$. *Then, asymptotic agreement is continuous under* $\{F^i_{\theta,m}\}$.

2. *Suppose that* $\tilde{R}(\hat{p}(i,\theta) - \hat{p}(j,\theta'), \hat{p}(i,\theta) - \hat{p}(j,\theta)) \neq 0$ *for each distinct* $\theta$ *and* $\theta'$. *Then, asymptotic agreement is strongly discontinuous under* $\{F^i_{\theta,m}\}$.

These theorems therefore show that the results about lack of asymptotic learning and asymptotic agreement derived in the previous section do not depend on the assumption that there are only two states and binary signals. It is also straightforward to generalize Propositions 2 and 1 to the case with multiple states and signals; we omit this to avoid repetition.

We assumed both the number of signal values and states are finite. This assumption can be dropped in the expense of introducing technical issues that are not central to our focus here.

# 5 Concluding Remarks

The standard approach in game theory and economic modeling assumes that individuals have a "common prior," meaning that they have beliefs consistent with each other regarding the game forms, institutions, and possible distributions of payoff-relevant parameters. This presumption is often justified by the argument that sufficient common experiences and observations, either through individual observations or transmission of information from others, will eliminate disagreements, taking agents towards common priors. This presumption receives support from a number of well-known theorems in statistics, such as Savage (1954) and Blackwell and Dubins (1962).

Nevertheless, existing theorems apply to environments in which learning occurs under *certainty*, that is, individuals are certain about the meaning of different signals. Certainty is sufficient to ensure that payoff-relevant variables can be *identified* from limiting frequencies of signals. In many situations, individuals are not only learning about a payoff-relevant parameter but also about the interpretation of different signals, i.e., learning takes place under *uncertainty*. For example, many signals favoring a particular interpretation might make individuals suspicious that the signals come from a biased source. This may prevent *full identification* (in the standard sense of the term in econometrics and statistics). In such situations, information will be useful to individuals but may not lead to full learning.

This paper investigates the conditions under which learning under uncertainty will take individuals towards common priors and asymptotic agreement. We consider an environment in which two individuals with different priors observe the same infinite sequence of signals informative about some underlying parameter. Learning is under *uncertainty*, however, because each individual has a non-degenerate subjective probability distribution over the likelihood of different signals given the values of the parameter. When subjective probability distributions of both individuals have full support, they will never agree, even after observing the same infinite sequence of signals.

Our main results provide conditions under which asymptotic agreement is fragile or discontinuous at certainty (meaning that as the amount of uncertainty in the environment diminishes, we remain away from asymptotic agreement). We first show that asymptotic agreement is discontinuous at certainty for every model. In particular, a vanishingly small amount of uncertainty about the signal distribution can guarantee that both individuals attach probability arbitrarily close to 1 that they will asymptotically disagree. Under additional strong continuity and uniform convergence assumptions, we also characterize the conditions under which asymptotic agreement is continuous at certainty. Even under these assumptions, asymptotic disagreement may prevail as the amount of uncertainty vanishes, provided that the family of subjective distributions has regularly-varying tails (such as for the Pareto, the log-normal or the t-distributions). In contrast, with rapidly-varying tails (such as the normal and the exponential distributions), convergence to certainty leads to asymptotic agreement.

Lack of common beliefs and common priors has important implications for economic behavior in a range of circumstances. The type of learning outlined in this paper interacts with economic behavior in various different situations. The companion paper, Acemoglu, Chernozhukov and Yildiz (2008), illustrates the influence of learning under uncertainty and lack of asymptotic agreement on games of coordination, games of common interest, bargaining, asset trading and games of communication.

# 6 Appendix: Omitted Proofs

**Proof of Lemma 1.** Write

$$\begin{aligned}
\frac{\Pr^i(r_n|\theta=B)}{\Pr^i(r_n|\theta=A)} &= \frac{\int_0^1 p^{r_n}(1-p)^{n-r_n} f_B(1-p)dp}{\int_0^1 p^{r_n}(1-p)^{n-r_n} f_A(p)dp} \\
&= \frac{\frac{\int_0^1 p^{r_n}(1-p)^{n-r_n} f_B(1-p)dp}{\int_0^1 p^{r_n}(1-p)^{n-r_n} dp}}{\frac{\int_0^1 p^{r_n}(1-p)^{n-r_n} f_A(p)dp}{\int_0^1 p^{r_n}(1-p)^{n-r_n} dp}} \\
&= \frac{\mathbb{E}^\lambda[f_B(1-p)|r_n]}{\mathbb{E}^\lambda[f_A(p)|r_n]}.
\end{aligned}$$

Here, the first equality is obtained by dividing the numerator and the denominator by the same term. The resulting expression on the numerator is the conditional expectation of $f_B(1-p)$ given $r_n$ under the flat (Lebesgue) prior on $p$ and the Bernoulli distribution on $\{s_t\}_{t=0}^n$. Denoting this by $\mathbb{E}^\lambda[f_B(1-p)|r_n]$, and the denominator, which is similarly defined as the conditional expectation of $f_A(p)$, by $\mathbb{E}^\lambda[f_A(p)|r_n]$, we obtain the last equality. By Doob's consistency theorem for Bayesian posterior expectation of the parameter, as $r_n \to \rho$, we have that $\mathbb{E}^\lambda[f_B(1-p)|r_n] \to f_B(1-\rho)$ and $\mathbb{E}^\lambda[f_A(p)|r_n] \to f_A(\rho)$ (see, e.g., Doob, 1949, Ghosh and Ramamoorthi, 2003, Theorem 1.3.2). This establishes

$$\frac{\Pr^i(r_n|\theta=B)}{\Pr^i(r_n|\theta=A)} \to R^i(\rho),$$

as defined in (4). Equation (3) then follows from (2). ■

**Proof of Theorem 4.** For each $m \gg 1$, let

$$f^i_{\theta,m}(\rho) = \begin{cases} x_\theta/\lambda & \text{if } \rho \in \left[\hat{p}^i_\theta - \lambda/2, \hat{p}^i_\theta + \lambda/2\right], \\ \varepsilon^3 & \text{if } \rho < 1 - \hat{p}^i_{\theta'} - \lambda/2, \\ \varepsilon & \text{otherwise,} \end{cases}$$

where $\theta' \neq \theta$, $\varepsilon = \lambda = 1/m$, $\hat{p}^1_A = \hat{p}_A + \lambda$, $\hat{p}^1_B = \hat{p}_B - \lambda$, $\hat{p}^2_A = \hat{p}_A - \lambda$, $\hat{p}^2_B = \hat{p}_B + \lambda$, and $x_\theta = 1 - \varepsilon\left(\hat{p}^i_{\theta'} - \lambda/2\right) - \varepsilon^3\left(1 - \hat{p}^i_{\theta'} - \lambda/2\right) \in (0,1)$. Here, $x_\theta$ is close to 1 for large $m$. Then,

$$R^{i,m}(\rho) = \begin{cases} 1/\varepsilon^2 & \text{if } \rho < 1 - \hat{p}^i_B - \lambda/2, \\ x_B/\varepsilon^2 & \text{if } 1 - \hat{p}^i_B - \lambda/2 \le \rho \le 1 - \hat{p}^i_B + \lambda/2, \\ 1 & \text{if } 1 - \hat{p}^i_B + \lambda/2 < \rho < \hat{p}^i_A - \lambda/2, \\ \varepsilon^2/x_A & \text{if } \hat{p}^i_A - \lambda/2 \le \rho \le \hat{p}^i_A + \lambda/2, \\ \varepsilon^2 & \text{if } \rho > \hat{p}^i_A + \lambda/2, \end{cases}$$

which is clearly decreasing when $m$ is large. For $\varepsilon \cong 0$, we have

$$R^{i,m}(\rho) \cong \begin{cases} \infty & \text{if } \rho \le 1 - \hat{p}^i_B + \lambda/2, \\ 1 & \text{if } 1 - \hat{p}^i_B + \lambda/2 < \rho < \hat{p}^i_A - \lambda/2, \\ 0 & \text{if } \rho \ge \hat{p}^i_A - \lambda/2, \end{cases}$$

and hence

$$\phi_{\infty}^{i,m}(\rho) \cong \begin{cases} 0 & \text{if } \rho \leq 1 - \hat{p}_B^i + \lambda/2, \\ \pi^i & \text{if } 1 - \hat{p}_B^i + \lambda/2 < \rho < \hat{p}_A^i - \lambda/2, \\ 1 & \text{if } \rho \geq \hat{p}_A^i - \lambda/2. \end{cases}$$

Notice that when $\rho \in \left[\hat{p}_A^2 - \lambda/2, \hat{p}_A^2 + \lambda/2\right]$, we have $\rho < \hat{p}_A^1 - \lambda/2$, so that $\phi_{\infty}^{2,m}(\rho) \cong 1$ and $\phi_{\infty}^{1,m}(\rho) \cong \pi^1$, yielding $\left|\phi_{\infty}^{1,m}(\rho) - \phi_{\infty}^{2,m}(\rho)\right| \cong 1 - \pi^1$. Similarly, when $\rho \in \left[1 - \hat{p}_B^1 - \lambda/2, \hat{p}_B^1 + \lambda/2\right]$, we have $\phi_{\infty}^{1,m}(\rho) \cong 0$ and $\phi_{\infty}^{2,m}(\rho) \cong \pi^2$, so that $\left|\phi_{\infty}^{1,m}(\rho) - \phi_{\infty}^{2,m}(\rho)\right| \cong \pi^2$. In order to complete the proof of theorem, we then pick $\bar{Z} = \min\left\{\pi^2, 1 - \pi^1\right\}/2$. In that case,

$$\lim_{m\to\infty} \Pr^{1,m}\left(\left|\phi_{\infty}^{1,m}(\rho) - \phi_{\infty}^{2,m}(\rho)\right| > \bar{Z}\right) = \lim_{m\to\infty} \Pr^{1,m}\left(\rho \in \left[1 - \hat{p}_B^1 - \lambda/2, \hat{p}_B^1 + \lambda/2\right]\right) = 1 - \pi^1 > 0,$$

and

$$\lim_{m\to\infty} \Pr^{2,m}\left(\left|\phi_{\infty}^{1,m}(\rho) - \phi_{\infty}^{2,m}(\rho)\right| > \bar{Z}\right) = \lim_{m\to\infty} \Pr^{2,m}\left(\rho \in \left[\hat{p}_A^2 - \lambda/2, \hat{p}_A^2 + \lambda/2\right]\right) = \pi^2 > 0,$$

completing the proof. ■

**Proof of Theorem 6.**

**(Proof of Part1)** This part immediately follows from Lemma 2, as each $\pi_{k'}^i f_{A_{k'}}(\rho(s))$ is positive, and $\pi_k^i f_{A_k}(\rho(s))$ is finite.

**(Proof of Part 2)** Assume $F_\theta^1 = F_\theta^2$ for each $\theta \in \Theta$. Then, by Lemma 2, $\phi_{k,\infty}^1(\rho) - \phi_{k,\infty}^2(\rho) = 0$ if and only if $\left(T_k\left(\pi^1\right) - T_k\left(\pi^2\right)\right)' T_k\left(\left(f_\theta^1(\rho)\right)_{\theta\in\Theta}\right) = 0$. The latter inequality has probability 0 under both probability measures $\Pr^1$ and $\Pr^2$ by hypothesis. ■

**Proof of Theorem 7.** Pick sequences $p_\theta^{i,m} \to p_\theta^i$ and $\bar{\epsilon} > 0$ such that $\left\|p_\theta^{1,m} - p_{\theta'}^{2,m}\right\| > \bar{\epsilon}/m$ for all $\theta, \theta'$ (including $\theta = \theta'$). For each $(\theta, i)$, define

$$D_\theta^{i,m} = \left\{p \in \Delta(L) : 3\left\|p - p_\theta^{i,m}\right\| \leq \bar{\epsilon}/m\right\},$$

which will be the set of likely frequencies at state $\theta$ according to $i$. Notice that $D_\theta^{i,m} \cap D_{\theta'}^{i',m} \neq \varnothing$ iff $\theta = \theta'$ and $i = i'$. Define

$$f_{\theta,m}^i(\rho) = \begin{cases} x_{\theta,m}^i & \text{if } \rho \in D_\theta^{i,m} \\ 1/m & \text{otherwise,} \end{cases}$$

where $x_{\theta,m}^i$ is normalized so that $f_{\theta,m}^i$ is a probability density function. By construction of sequences $f_{\theta,m}^i$ and $p_\theta^{i,m}$, $F_{\theta,m}^i \to \delta_{p_\theta^i}$ for each $(\theta, i)$. We will show that agreement is discontinuous under $\{F_{\theta,m}^i\}$. Now

$$\phi_{\infty,\theta,m}^i(\rho) = \frac{1}{1 + \frac{1-\pi_\theta^i}{\pi_\theta^i m x_{\theta,m}^i}}$$

if $\rho \in D_\theta^{i,m}$ for some $\theta$ and $\phi_{\infty,m}^i(\rho) = \pi^i$ otherwise. Note that $\phi_{\infty,\theta,m}^i(\rho) \to 1$ if $\rho \in D_\theta^{i,m}$. Moreover, since the sets $D_\theta^{i,m}$ are disjoint (as we have seen above), $\phi_{\infty,m}^j(\rho) = \pi^j$ when $\rho \in D_\theta^{i,m}$. Hence, there exist $\bar{m}$ such that for any $m \geq \bar{m}$ and any $\rho \in D^{i,m} \equiv \cup_\theta D_\theta^{i,m}$,

$$\left\|\phi_{\infty,m}^i(\rho) - \phi_{\infty,m}^j(\rho)\right\| > \varepsilon$$

where $\varepsilon \equiv \min_{j,\theta}\left(1-\pi_\theta^j\right)/2$. But for each $\theta$, $\Pr^{i,m}\left(\rho \in D_\theta^{i,m}|\theta\right) \geq 1-1/m$, showing that $\Pr^{i,m}\left(\rho \in D^{i,m}\right) \geq 1-1/m$. Therefore,

$$\lim_{m\to\infty} \Pr^{i,m}\left(\left\|\phi^i_{\infty,m}-\phi^j_{\infty,m}\right\| > \varepsilon\right) = 1.$$

■

**Proof of Theorem 8.** Our proof utilizes the following two lemmas.

**Lemma A.**

$$\lim_{m\to\infty} \phi^i_{k,\infty,m}(p) = \frac{1}{1+\sum_{k'\neq k}\frac{\pi^i_{k'}}{\pi^i_k}\tilde{R}\left(p-\hat{p}\left(i,A_{k'}\right),p-\hat{p}\left(i,A_k\right)\right)}.$$

**Proof.** By condition (i), $\lim_{m\to\infty} c\left(i,A_k,m\right)=1$ for each $i$ and $k$. Hence, for every distinct $k$ and $k'$,

$$\lim_{m\to\infty}\frac{f^i_{A_{k'}}(p)}{f^i_{A_k}(p)} = \lim_{m\to\infty}\frac{c\left(i,A_{k'},m\right)}{c\left(i,A_k,m\right)}\lim_{m\to\infty}\frac{f\left(m\left(p-\hat{p}\left(i,A_{k'}\right)\right)\right)}{f\left(m\left(p-\hat{p}\left(i,A_k\right)\right)\right)} = \tilde{R}\left(p-\hat{p}\left(i,A_{k'}\right),p-\hat{p}\left(i,A_k\right)\right).$$

Then, Lemma A follows from Lemma 2. ■

**Lemma B.** *For any $\tilde{\varepsilon}>0$ and $h>0$, there exists $\tilde{m}$ such that for each $m>\tilde{m}$, $k\leq K$, and each $\rho(s)$ with $\left\|\rho(s)-\hat{p}\left(i,A_k\right)\right\|<h/m$,*

$$\left|\phi^i_{k,\infty,m}(\rho(s))-\lim_{m\to\infty}\phi^i_{k,\infty,m}\left(\hat{p}\left(i,A_k\right)\right)\right|<\tilde{\varepsilon}. \tag{26}$$

**Proof.** Since, by hypothesis, $\tilde{R}$ is continuous at each $\left(\hat{p}(i,\theta)-\hat{p}\left(j,\theta'\right),\hat{p}(i,\theta)-\hat{p}(j,\theta)\right)$, by Lemma A, there exists $h'>0$, such that

$$\left|\lim_{m\to\infty}\phi^i_{k,\infty,m}(\rho(s))-\lim_{m\to\infty}\phi^i_{k,\infty,m}\left(\hat{p}\left(i,A_k\right)\right)\right|<\tilde{\varepsilon}/2 \tag{27}$$

and by condition (iii), there exists $\tilde{m}>h/h'$ such that

$$\left|\phi^i_{k,\infty,m}(\rho(s))-\lim_{m\to\infty}\phi^i_{k,\infty,m}(\rho(s))\right|<\tilde{\varepsilon}/2. \tag{28}$$

holds uniformly in $\left\|\rho(s)-\hat{p}\left(i,A_k\right)\right\|<h'$. The inequalities in (27) and (28) then imply (26). ■

■

**Lemma C.** $\lim_{m\to\infty}\left(\phi^i_{k,\infty,m}\left(\hat{p}\left(i,A_k\right)\right)-\phi^j_{k,\infty,m}\left(\hat{p}\left(i,A_k\right)\right)\right)=0$ iff $\tilde{R}\left(\hat{p}\left(i,A_k\right)-\hat{p}\left(j,A_{k'}\right),\hat{p}\left(i,A_k\right)-\hat{p}\left(j,A\right.\right.$ 0 for each $k'\neq k$.

**Proof. Proof.** Since $\tilde{R}\left(\hat{p}\left(i,A_k\right)-\hat{p}\left(i,A_{k'}\right),0\right)=0$ for each $k'\neq k$ (by condition (i)), Lemma A implies that $\lim_{m\to\infty}\phi^i_{k,\infty,m}\left(\hat{p}\left(i,A_k\right)\right)=1$. Hence, $\lim_{m\to\infty}\left(\phi^i_{k,\infty,m}\left(\hat{p}\left(i,A_k\right)\right)-\phi^j_{k,\infty,m}\left(\hat{p}\left(i,A_k\right)\right)\right)=$ 0 if and only if $\lim_{m\to\infty}\phi^j_{k,\infty,m}\left(\hat{p}\left(i,A_k\right)\right)=1$. Since each ratio $\pi^j_{k'}/\pi^j_k$ is positive, by Lemma

A, the latter holds if only if $\tilde{R}(\hat{p}(i, A_k) - \hat{p}(j, A_{k'}), \hat{p}(i, A_k) - \hat{p}(j, A_k)) = 0$ for each $k' \neq k$. ■

**(Proof of Part 1)** Fix $\epsilon > 0$ and $\delta > 0$. We will find $\bar{m} \in \mathbb{N}$ such that

$$\Pr^i\left(\left\|\phi^1_{\infty,m}(s) - \phi^2_{\infty,m}(s)\right\| > \epsilon\right) < \delta \qquad (\forall m > \bar{m}, i = 1, 2).$$

Fix any $i$ and $k$. Since each $\pi^j_{k'}/\pi^j_k$ is finite, by Lemma 2, there exists $\epsilon' > 0$, such that $\phi^i_{k,\infty,m}(\rho(s)) > 1 - \epsilon$ whenever $f^i_{A_{k'}}(\rho(s))/f^i_{A_k}(\rho(s)) < \epsilon'$ holds for every $k' \neq k$. Now, by (i), there exists $h_{0,k} > 0$, such that

$$\Pr^i(\|\rho(s) - \hat{p}(i, A_k)\| \leq h_{0,k}/m | \theta = A_k) = \int_{\|x\| \leq h_{0,k}} f(x)\, dx > (1 - \delta).$$

Let

$$Q_{k,m} = \{p \in \Delta(L) : \|p - \hat{p}(i, A_k)\| \leq h_{0,k}/m\}$$

and $\kappa \equiv \min_{\|x\| \leq h_{0,k}} f(x) > 0$. By (i), there exists $h_{1,k} > 0$ such that, whenever $\|x\| > h_{1,k}$, $f(x) < \epsilon'\kappa/2$. There exists a sufficiently large constant $m_{1,k}$ such that for any $m > m_{1,k}$, $\rho(s) \in Q_{k,m}$, and any $k' \neq k$, we have $\|\rho(s) - \hat{p}(i, A_{k'})\| > h_{1,k}/m$, and

$$\frac{f(m(\rho(s) - \hat{p}(i, A_{k'})))}{f(m(\rho(s) - \hat{p}(i, A_k)))} < \frac{\epsilon'\kappa}{2}\frac{1}{\kappa} = \frac{\epsilon'}{2}.$$

Moreover, since $\lim_{m\to\infty} c(i, \theta, m) = 1$ for each $i$ and $\theta$, there exists $m_{2,k} > m_{1,k}$ such that $c(i, A_{k'}, m)/c(i, A_k, m) < 2$ for every $k' \neq k$ and $m > m_{2,k}$. This implies

$$f^i_{A_{k'}}(\rho(s))/f^i_{A_k}(\rho(s)) < \epsilon',$$

establishing that

$$\phi^i_{k,\infty,m}(\rho(s)) > 1 - \epsilon. \tag{29}$$

Now, for $j \neq i$, assume that $\tilde{R}\left(\hat{p}(i, \theta) - \hat{p}\left(j, \theta'\right), \hat{p}(i, \theta) - \hat{p}(j, \theta)\right) = 0$ for each distinct $\theta$ and $\theta'$. Then, by Lemma A, $\lim_{m\to\infty} \phi^j_{k,\infty,m}(\hat{p}(i, A_k)) = 1$, and hence by Lemma B, there exists $m_{3,k} > m_{2,k}$ such that for each $m > m_{3,k}$, $\rho(s) \in Q_{k,m}$,

$$\phi^j_{k,\infty,m}(\rho(s)) > 1 - \epsilon. \tag{30}$$

Notice that when (29) and (30) hold, we have $\left\|\phi^1_{\infty,m}(s) - \phi^2_{\infty,m}(s)\right\| < \epsilon$. Then, setting $\bar{m} = \max_k m_{4,k}$, we obtain the desired inequality for each $m > \bar{m}$:

$$\begin{aligned}
\Pr^i\left(\left\|\phi^1_{\infty,m}(s) - \phi^2_{\infty,m}(s)\right\| < \epsilon\right) &= \sum_{k \leq K} \Pr^i\left(\left\|\phi^1_{\infty,m}(s) - \phi^2_{\infty,m}(s)\right\| < \epsilon | \theta = A_k\right) \Pr^i(\theta = A_k) \\
&\geq \sum_{k \leq K} \Pr^i(\rho(s) \in Q_{k,m} | \theta = A_k) \Pr^i(\theta = A_k) \\
&\geq \sum_{k \leq K} (1 - \delta)\pi^i_k \\
&= 1 - \delta.
\end{aligned}$$

**(Proof of Part 2)** Assume that $\tilde{R}\left(\hat{p}(i,\theta)-\hat{p}\left(j,\theta'\right),\hat{p}(i,\theta)-\hat{p}(j,\theta)\right)\neq 0$ for each distinct $\theta$ and $\theta'$. We will find $\epsilon>0$ such that for each $\delta>0$, there exists $\bar{m}\in\mathbb{N}$ such that

$$\Pr^i\left(\left\|\phi^1_{\infty,m}(s)-\phi^2_{\infty,m}(s)\right\|>\epsilon\right)>1-\delta \qquad (\forall m>\bar{m}, i=1,2).$$

Now, since each $\pi^j_{k'}/\pi^j_k$ is positive, Lemma A implies that $\lim_{m\to\infty}\phi^j_{k,\infty,m}(\hat{p}(i,A_k))<1$ for each $k$. Let

$$\epsilon=\min_k\left\{1-\lim_{m\to\infty}\phi^j_{k,\infty,m}(\hat{p}(i,A_k))\right\}/3>0.$$

Then, by Part 1, for each $k$, there exists $m_{2,k}$ such that for every $m>m_{2,k}$ and $\rho(s)\in Q_{k,m}$, we have $\phi^i_{k,\infty}(\rho(s))>1-\epsilon$. By Lemma B, there also exists $m_{5,k}>m_{2,k}$ such that for every $m>m_{5,k}$ and $\rho(s)\in Q_{k,m}$,

$$\phi^j_{k,\infty,m}(\rho(s))<\lim_{m\to\infty}\phi^j_{k,\infty,m}(\hat{p}(i,A_k))+\epsilon\leq 1-2\epsilon<\phi^i_{k,\infty}(\rho(s))-\epsilon.$$

This implies that $\left\|\phi^1_{\infty,m}(\rho(s))-\phi^2_{\infty,m}(\rho(s))\right\|>\epsilon$. Setting $\bar{m}=\max_k m_{5,k}$ and changing $\left\|\phi^1_{\infty,m}(s)-\phi^2_{\infty,m}(s)\right\|<\epsilon$ at the end of the proof of Part 1 to $\left\|\phi^1_{\infty,m}(s)-\phi^2_{\infty,m}(s)\right\|>\epsilon$, we obtain the desired inequality. ■

# 7 References

Acemoglu, Daron, Victor Chernozhukov, and Muhamet Yildiz (2008): "Economic Applications of Learning under Uncertainty", mimeo.

Aumann, Robert (1986): "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica*, 55, 1-18.

Aumann, Robert (1998): "Common priors: A Reply to Gul," *Econometrica*, 66-4, 929-938.

Berk, Robert (1966): "Limiting Behavior of Posterior Distributions When the Model Is Incorrect," *Annals of Mathematical Statistics*, 37, 51-58.

Billingsley, Patrick (1995): *Probability and Measure*, Third Edition, Wiley and Sons, New York.

Blackwell, David and Lester Dubins (1962): "Merging of Opinions with Increasing Information," *Annals of Mathematical Statistics*, 33, 882-887.

Cripps, Martin, Jeffrey Ely, George Mailath and Larry Samuelson (2006): "Common Learning" Cowles Foundation Discussion Paper No. 1575, August 2006.

Doob, J. L. (1949): "Application of the Theory of Martingales," in *Le Calcul des Probabilities et ses Applications*, 23-27, Paris.

Diaconis Persi and David Freedman (1986): "On Inconsistent Bayesian Estimates of Location" *Annals of Statistics*, 14, 68-87.

Dixit, Avinash and Jorgen Weibull (2007): "Political Polarization," Mimeo, Princeton University and Stockholm School of Economics.

Feller, William (1971): *An Introduction to Probability Theory and Its Applications.* Volume II, 2nd ed. Wiley, New York.

Freedman, David (1963): "On the Asymptotic Behavior of Bayes Estimates in the Discrete Case I" *Annals of Mathematical Statistics*, 34, 1386-1403.

Freedman, David (1965): "On the Asymptotic Behavior of Bayes Estimates in the Discrete Case II" *Annals of Mathematical Statistics*, 36, 454-456.

Fudenberg Drew and David Kreps (1995): "Learning in Extensive Games, I: Self-Confirming Equilibria," *Games and Economic Behavior*, 8, 20-55.

Fudenberg, Drew and David Levine (1993): "Self-Confirming Equilibria" *Econometrica*, 61, 523- 546.

Fudenberg, Drew and David Levine (2005): "Learning and Belief Based Trade," *Latin American Journal of Economics*, 42, 199-27.

Geanakoplos, John and Heraklis Polemarchakis (1982): "We Can't Disagree Forever," *Journal of Economic Theory*, 28, 192-200.

Ghosh J. K. and R. V. Ramamoorthi (2003): *Bayesian Nonparametrics*, Springer-Verlang, New York.

Gul, Faruk (1998): "A Comment on Aumann's Bayesian view", *Econometrica*, 66-4, 923-927.

Guillemin Victor and Alan Pollack (1974): *Differential Topology*, Prentice Hall, New Jersey.

Haan, L. de (1970): *On Regular Variation and Its Application to the Weak Convergence of Sample Extremes,* Mathematical Centre Tract 32, Mathematics Centre, Amsterdam, Holland.

Jackson, Matthew O., Ehud Kalai and Rann Smorodinsky (1999): "Bayesian Representation of Stochastic Processes under Learning: De Finetti Revisited," *Econometrica*, 67, 875-893.

Kurz, Mordecai (1994): "On the Structure and Diversity of Rational Beliefs," *Economic Theory*, 4, 877-900.

Kurz, Mordecai (1996): "Rational Beliefs and Endogenous Uncertainty: An Introduction," *Economic Theory*, 8, 383-397.

Miller, Ronald I. and Chris W. Sanchirico (1999): "The Role of Option with Continuity in Merging of Opinions and Rational Learning," *Games and Economic Behavior*, 29, 170-190.

Savage, Leonard J. (1954): *The Foundations of Statistics*, Dover reprint, New York, 1972.

Seneta, Eugène (1976): *Regularly Varying Functions,* Lecture Notes in Mathematics, 508, Springer, Berlin.

Stinchcombe, Maxwell (2005): "The Unbearable Flightness of Bayesians: Generically Erratic Updating," University of Texas, Austin, mimeo.