Fragility of Asymptotic Agreement under Bayesian Learning[*]

Daron Acemoglu, Victor Chernozhukov, and Muhamet Yildiz[†]

February 2009.

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, vanishingly small individual uncertainty about the signal distributions can lead to substantial (non-vanishing) differences in asymptotic beliefs. Under a uniform convergence assumption, we then characterize the conditions under which a small amount of uncertainty leads to significant asymptotic disagreement.

Keywords: asymptotic disagreement, Bayesian learning, merging of opinions.
JEL Classification: C11, C72, D83.

1 Introduction

1.1 Motivation

The common prior assumption is one of the cornerstones of modern economic analysis. Most models assume that the players in a game have a common prior about the game form and the payoffs. For example, they postulate that some *state* (e.g., a payoff-relevant parameter) θ is drawn from a commonly known distribution G, even though each player may also have additional information about some components of θ. 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 θ and this process will lead to "agreement" among individuals about the distribution of θ. 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 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 there are peer effects is interpreted very differently by two economists starting with different priors. An economist believing that peer effects are small and unimportant 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 at multiple states. Hence, even though the individuals will learn the asymptotic frequency of signals, they may not always be able to infer the state θ 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

1

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 a small amount of uncertainty leads only to a small amount of disagreement asymptotically. Our main result shows that this is never the case: for *every* model, there exists a vanishingly small amount of uncertainty such that under this uncertainty both individuals assign nearly probability 1 that they will asymptotically hold significantly different beliefs about the underlying state. This result implies that learning foundations of common priors are not as strong as generally presumed.

1.2   Formulation

Consider the following example, which illustrates the main ideas presented below. There are two states $\theta \in \{A, B\}$ and binary signals $s_t \in \{a, b\}$. Two individuals with given priors publicly observe a sequence of signals, $\{s_t\}_{t=1}^n$, and form their posteriors about the state θ. Conditional on the state, the signals are independently and identically distributed, but the individuals do not necessarily know this distribution. Letting p_θ be this unknown probability of $s_t = a$ at state θ, we assume that each individual has a possibly non-degenerate belief about p_θ. This belief has a cumulative distribution function F_θ^i; the density, when it exists, is denoted by f_θ^i. The standard model used for analysis of learning and agreement in the literature is the special case of this environment where each F_θ^i puts all of its mass at some p_θ, with $p_A > 1/2 > p_B$. Throughout, we refer to this benchmark as the (or a) "standard model". Thus in contrast to the standard model, where the informativeness of signals is known, the individuals in our environments may face some uncertainty about the informativeness of signals. Consequently, as they observe additional signals, they learn not only about the state θ, but also about the interpretation of the signals. Our general model is introduced in Section 2 and the example here is a special case that is discussed in greater detail in Section 3. In that section, using an asset trading example, we illustrate how the failure of asymptotic agreement may have significant implications for game theoretic and economic analysis.

In Section 4, we first extend existing results on asymptotic learning and agreement. In particular, we show that when the supports of F_A^i and F_B^i are disjoint (which is significantly more general than the standard case where each F_θ^i has point mass), there is "full identification" for both individuals, in the sense that each can identify the state from the long-run frequency

of signals (as $n \to \infty$). In this case, each individual is certain that he will eventually learn the true state (*asymptotic learning*). Moreover, we show that under additional mild conditions, each individual is also certain that they will eventually agree (*asymptotic agreement*).

A key observation of our paper is that asymptotic agreement under full identification rests on how individuals treat zero probability events. In particular, under full identification, individuals do not question their models even along sample paths that are impossible according to their model, because they attribute the frequencies that are outside the support of their model to sampling variation, regardless of how unlikely such sampling variation may be and how large their samples are. An implication is that along such sample paths their beliefs about the future frequencies diverge from the empirical long-run frequency, leading to a form of inconsistency. This inconsistency and the unwillingness of individuals to adjust their models in the face of overwhelming evidence to the contrary are unappealing features of the standard formulation. More importantly, support restrictions are clearly idealizations that equate small probabilities with zero probability; they can be justified only when we know that behavior under small and zero probabilities are similar. Whether or not this is so can only be studied by first considering models without such support restrictions.

Motivated by these observations, we relax all such support restrictions and instead assume that each F_θ^i has full support. Under full support, as each individual observes additional signals, his beliefs about the future frequencies approach the empirical long-run frequency. By Bayes rule, beliefs about the underlying payoff-relevant state are determined by the likelihood ratio f_B^i/f_A^i of the densities at the realized long-run frequencies. Sampling variation (which disappears due to the strong law of large numbers) plays no role in the formation of asymptotic beliefs.

An immediate implication of the full support assumption is that there will be no full identification. Consequently, each individual recognizes that he will never fully learn the true state, as some uncertainty about the informativeness of the signals and the state remains forever. This also implies that, except for the knife-edge cases, each individual is also certain that asymptotic agreement will fail, in the sense that some amount of differences of opinions will remain forever.

The most substantive part of our paper, Section 5, investigates whether the amount of disagreement is small (vanishing) when we are arbitrarily close to a fully-identified model. We consider a standard (fully-identified) model parameterized by a vector $\left(p_A^1, p_B^1, p_A^2, p_B^2\right)$, where each F_θ^i puts probability 1 on a single frequency p_θ^i. We then consider families $\{F_{\theta,m}^i\}$ of dis-

3

tributions with full support that become increasingly concentrated on p_θ^i (as $m \to \infty$). Here, a family $\{F_{\theta,m}^i\}$ is a possible relaxation of the idealized assumptions in the standard model. Our main question is thus whether the amount of asymptotic disagreement vanishes in the limit as $m \to \infty$ (as the model is arbitrarily close to the standard model). Our first main result in this section shows that asymptotic agreement is never robust to all relaxations. More precisely, for every standard model (here parameterized by the vector $\left(p_A^1, p_B^1, p_A^2, p_B^2\right)$), we construct a family $\{F_{\theta,m}^i\}$ of distributions that become more and more concentrated around p_θ^i, such that the amount of asymptotic disagreement eventually exceeds a fixed positive level for almost all sample paths and for all m. This result therefore implies that asymptotic agreement is "fragile"—small perturbations of the standard model lead to significant asymptotic disagreement.

The idea underlying this fragility result is intuitive. As $m \to \infty$ and we approach the standard model, the identification problem vanishes, in the sense that each individual i assigns nearly probability 1 to the event that he will learn the true state. However, even though asymptotic learning applies, asymptotic agreement is considerably more demanding. For asymptotic agreement, each individual must also be certain that the other individual will also eventually learn the true state. While this latter requirement is true in the standard model, it is often not the case *near* the standard model. In particular, near any standard model, each individual assigns a high probability to a small set of long-run frequencies (thus ensuring asymptotic learning). Yet, even if F_θ^1 and F_θ^2 are very close to each other, the likelihood ratios of the densities may remain significantly different on those sets. In particular, j may assign a low probability to the true state at the frequencies i finds likely. In that case, i would be nearly certain that j will fail to learn the true state and the beliefs will be different in the long run. Importantly, this conclusion is true even for (instances of) the standard model when $p_\theta^1 = p_\theta^2$, so that the individuals agree on the likely frequencies.

Our second main result in Section 5 provides a tight characterization of the conditions under which asymptotic agreement is fragile when the families of distributions $\{F_{\theta,m}^i\}$ is such that the resulting likelihood ratios converge *uniformly* to a continuous function. This uniform convergence requirement ensures that asymptotic beliefs are not highly sensitive to the long-run frequency of signals as $m \to \infty$. In the context of a canonical example, our characterization shows that the asymptotic agreement results are fragile when the families of distributions $\{F_{\theta,m}^i\}$ converging to the standard model have regularly-varying (polynomial) tails, such as the Pareto or the log-normal distributions but not when they have rapidly-varying (exponential)

4

tails, such as the normal and the exponential distributions.

Our third main result in Section 5 shows that agreement is "continuous" in the medium run.[1] Consider any standard model with asymptotic agreement and any family $\{F_{\theta,m}^i\}$ of models converging to the standard model. For any model $F_{\theta,m}^i$, individual beliefs may be far apart at the beginning and also asymptotically (as $n \to \infty$). However, we show that in the middle, the beliefs of the two individuals will be arbitrarily close to each other for long periods, provided that m is sufficiently large. The intuition for this result is as follows. The events concerning a few signal realizations correspond to only "coarse" information. This information is similar under $F_{\theta,m}^i$ and the standard model, so that individual beliefs are similar in the two models and the disagreement decreases with more observations for a while. However, eventually, individuals start using "finer" information in updating their beliefs and it is this finer information that is different under the two models. Therefore, eventually beliefs may grow apart under $F_{\theta,m}^i$, while they keep approaching each other under the standard model.

1.3 Interpretation

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.[2] Whether this common prior assumption is warranted therefore depends on the specific setting and what type of information individuals are trying to glean from the data.

The relevance of our results for theoretical modeling depends on whether our full-support assumption is a better approximation to reality and a more useful modeling tool for certain situations than the standard full-identification assumption. The full-support assumption does not rule out that p_B, the unknown probability of $s_t = a$ at state $\theta = B$, is higher than p_A, the unknown probability of $s_t = a$ at state $\theta = A$. That is, the individual finds it possible (though unlikely) that a signal $s_t = a$ can be considered as evidence in favor of state B rather than A. This is because the individual is uncertain not only about the informativeness of the signals

---

[1] We thank a referee for conjecturing such a result and encouraging us to investigate it further.

[2] See Section 3 for an example. 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.

but also about their direction.[3]

It may at first appear that individuals should always know whether a particular signal value is evidence in favor of state A or state B. One might then argue that as an individual observes more and more a signals, he should not decrease his belief that the state is A—that is, beliefs should be *monotone* in frequencies. For example, this reasoning would suggest that if a candidate for an internal promotion has more publications in a particular journal, then his chances for promotion should also be higher. Likewise, if we keep finding radioactive residues in various sites in a country, we should not decrease our belief that the country has a covert nuclear weapons program.[4]

These intuitions are correct in fully-identified models, but not in our more general environment.[5] This is a strength—not a shortcoming—of our model. We now argue that these intuitions are in fact not as compelling as they first appear and rule out a range of relevant empirical and theoretical possibilities.

Consider the internal promotion case first. The department has voted for promotion and the case comes before the president of the university, who is from an unrelated department. The chair of the department tells her that journal A publishes only exceptional contributions to the discipline, while journal B publishes minor contributions to the candidate's field. Suppose that the candidate has 3 publications in journal A and 3 publications in journal B, and the president approves the promotion. Now consider the case where the candidate had 2 more publications in journal A. It is natural to suppose that she would be even more enthusiastic about the case. Would she still be as enthusiastic about the promotion if the candidate had 20 publications in journal A? 200 more publications? 2000 more publications? Clearly, as we increase the number of publications in journal A, the president will eventually start doubting the description that journal A publishes only exceptional contributions and at some point will start putting less weight on publications in journal A. Naturally, before seeing the candidate's publication record, she would have attached a very small probability to seeing 2000 publications in journal A. But faced with such a promotion case, she would start questioning her working hypothesis.

Similarly, finding nuclear residue in a site may be considered a strong evidence for a covert

---

[3]Note, however, that under $\{F_{\theta,m}^i\}$, the probability that p_B is higher than p_A becomes vanishingly small as $m \to \infty$.

[4]We thank an anonymous referee for suggesting these examples.

[5]In fact, the full-identification assumption is considerably stronger than monotonicity and hence full-identification assumption fails whenever asymptotic beliefs are non-monotone on some (possibly unlikely) events. Theorem 5 below shows that our fragility results hold even when asymptotic beliefs are monotone in signal frequencies (though naturally our results are not true under full identification).

6

nuclear weapons program. Finding residues in three different sites would probably be considered a smoking gun. But what if we find nuclear residues everywhere in the country? We would presumably not conclude that there is a nuclear facility in every building in the country, but start entertaining the hypothesis that the measurements are wrong or they are responding to some other compound or to some specific feature of the geography of the country.

1.4   A Brief Literature Review

Blackwell and Dubins (1962) show that when two agents agree on zero probability events (i.e., their beliefs are absolutely continuous with respect to each other), asymptotically, they will make the same predictions about future frequencies of signals. It is well-understood that the absolute continuity assumption is crucial for such a merging of opinions and its relaxation can lead to a failure of merging; see, for example, Freedman (1963, 1965), Diaconis and Freedman, 1986, Miller and Sanchirico (1999), and Stinchcombe (2005).[6] For example, Freedman shows that when there are infinitely many signal values, an individual may put positive probability to the conditional signal distributions that are arbitrarily close to the true signal distribution in the product topology, but his future predictions may diverge dramatically from those of another individual who knows the true signal distribution. This is because posterior beliefs may be quite sensitive to the tail of the conditional signal distribution and the tails are negligible in the product topology. 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, but they may fail to converge to a limiting distribution.

In contrast to the above-mentioned papers, we do not question the absolute continuity assumption, as our full-support assumption implies absolute continuity. In particular, as in Blackwell and Dubins' theorem, the individuals asymptotically agree on the future frequency of signals. Indeed, our results rely on the fact that agreeing about future frequencies is not the same as agreeing about the underlying payoff-relevant state, because of the identification problem that arises in the presence of uncertainty.[7] 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

---

[6] In dynamic games, another source of lack of merging 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, failing to learn their payoffs (see, Fudenberg and Levine, 1993, Fudenberg and Kreps, 1995).

[7] 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.

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, in the asset trading example discussed in Section 3, what is most relevant is not the frequency of information about the dividend process but the actual dividend that the asset will pay. Thus, 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 results show that even when this identification problem is negligible for individual learning, its implications for asymptotic agreement may be significant.

Our paper is also related to recent independent work by Cripps, Ely, Mailath and Samuelson (2008), 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 (negligible) *uncertainty*. Since all signals are public in our model, there is no difficulty in achieving approximate common knowledge.[8]

2 Model

There are two individuals, denoted by $i \in \{1, 2\}$. The individuals care about a state θ, which comes from a finite set Θ with $K \geq 2$ elements. The individuals cannot observe the state, but they publicly observe a sequence of signals $\{s_t\}_{t=1}^{n}$ where $s_t \in \Sigma$ for some finite set Σ with $L \geq 2$ elements. We designate $\theta \in \Theta$ and $\sigma \in \Sigma$ as a generic state and a generic signal value, respectively. We write $\Delta(\Theta) \subset [0,1]^K$ and $\Delta(\Sigma) \subset [0,1]^L$ for the sets of all probability distributions on Θ and Σ, respectively. We endow $\Delta(\Theta)$ and $\Delta(\Sigma)$ with the supremum norm $\|\cdot\|$. Agent i assigns ex ante probability $\pi_\theta^i > 0$ to θ; we write $\pi^i \equiv (\pi_\theta^i)_{\theta \in \Theta} \in \Delta(\Theta)$ for the vector of prior beliefs. The individuals believe that, given θ, the signals are exchangeable, i.e., they are independently and identically distributed with an unknown distribution.[9] That is,

---

[8] Put differently, we ask 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.

[9] 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

the probability of $s_t = \sigma$ given θ is an unknown number $p_{\theta,\sigma}$. Here, $p_{\theta,\sigma}$ can be considered as the long-run frequency of σ when the true state is θ. We write $p_\theta \equiv (p_{\theta,\sigma})_{\sigma \in \Sigma} \in \Delta(\Sigma)$.

Our main departure from the standard model is that we allow the individuals to be uncertain about p_θ. We denote the cumulative joint distribution function of p_θ according to individual i—namely, his *subjective probability distribution*—by F_θ^i. In the standard model, F_θ^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:

Assumption 1 (Full Support) *For each i and θ, F_θ^i has a continuous, non-zero and finite density f_θ^i over $\Delta(\Sigma)$.*

The assumption implies that F_θ^i has a full support over the simplex $\Delta(\Sigma)$. Assumption 1 is stronger than necessary for our results, but simplifies the exposition. In addition, throughout we assume that π^1, π^2, F_θ^1 and F_θ^2 are known to both individuals.[10]

We consider infinite sequences $s \equiv \{s_t\}_{t=1}^\infty$ of signals and write S for the set of all such sequences. We write

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

for the posterior probability that the true state is θ given a sequence of signals $\{s_t\}_{t=1}^n$ under prior π^i and subjective probability distribution F_θ^i. After observing $\{s_t\}_{t=1}^n$, i assigns probability $\phi_{\theta,n}^i(s)$ to state θ. 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

$$r_{\sigma,n}(s) \equiv \#\{t \leq n \mid s_t = \sigma\},$$

the number of times $s_t = \sigma$ out of first n signals for each signal value $\sigma \in \Sigma$; we write $r_n(s) \equiv (r_{\sigma,n}(s))_{\sigma \in \Sigma}$.[11] Let us write

$$\bar{S} \equiv \left\{s \in S \mid \lim_{n\to\infty} r_{\sigma,n}(s)/n \text{ exists for each } \sigma \in \Sigma\right\} \tag{1}$$

---

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.

[10] 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.

[11] Given the definition of $r_n(s)$, the probability distribution $\Pr^i$ on $\Theta \times S$ is defined by setting

$$\Pr^i\left(E^{\theta,s,n}\right) \equiv \pi_\theta^i \int \prod_{\sigma \in \Sigma} p_{\theta,\sigma}^{r_{\sigma,n}(s)} f_\theta^i(p_\theta)\, dp_\theta$$

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

for the set of sequences under which the empirical frequency $r_{\sigma,n}(s)/n$ converges to some long-run frequency and let us denote this *long-run frequency* of σ under s by

$$\rho_\sigma(s) \equiv \lim_{n\to\infty} r_{\sigma,n}(s)/n. \tag{2}$$

We denote the vector of long-run frequencies by $\rho(s) \equiv (\rho_\sigma(s))_{\sigma\in\Sigma} \in \Delta(\Sigma)$. By the strong law of large numbers, such a limit exists almost surely for both individuals, which implies that $\Pr^i(s \in \bar{S}) = 1$ 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

$$\phi^i_{\theta,n}(s) = \frac{1}{1 + \sum_{\theta'\neq\theta} \frac{\pi^i_{\theta'}}{\pi^i_\theta} \frac{\Pr^i(r_n|\theta')}{\Pr^i(r_n|\theta)}}, \tag{3}$$

where $\Pr^i(r_n|\theta)$ is the probability of observing the signal $s_t = \sigma$ exactly $r_{\sigma,n}$ times out of n signals for each $\sigma \in \Sigma$ with respect to the distribution F^i_θ.

The following lemma provides a useful formula for the *asymptotic belief* of individual i, $\lim_{n\to\infty} \phi^i_{\theta,n}(s)$, and introduces the concept of the asymptotic likelihood ratio. Both the formula and the asymptotic likelihood ratio are used throughout the rest of the paper.

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

$$\lim_{n\to\infty} \phi_{\theta,n}(s) = \frac{1}{1 + \sum_{\theta'\neq\theta} \frac{\pi^i_{\theta'}}{\pi^i_\theta} R^i_{\theta,\theta'}(\rho(s))} \equiv \phi^i_{\theta,\infty}(\rho(s)) \tag{4}$$

where

$$R^i_{\theta,\theta'} \equiv f^i_{\theta'}/f^i_\theta \tag{5}$$

is the asymptotic likelihood ratio for θ and θ'.

    Proof. All proofs are in the Appendix. ∎

    In (5), $R^i_{\theta,\theta'}(\rho)$ is the *asymptotic likelihood ratio* of observing frequency vector ρ when the true state is θ' versus when it is θ. Lemma 1 states that, asymptotically, individual i uses these likelihood ratios and the Bayes rule to compute his asymptotic posterior beliefs about θ. Notice that $\phi^i_{\theta,\infty}$ and $R^i_{\theta,\theta'}$ are defined on the simplex $\Delta(\Sigma)$ of frequency vectors. We write $\phi^i_\infty \equiv (\phi_{\theta,\infty})_{\theta\in\Theta}$.

    The two questions of interest for us are:

1. **Asymptotic learning:** whether $\Pr^i(\phi^i_{\theta,\infty} = 1|\theta) = 1$ for each θ and $i = 1, 2$.

2. **Asymptotic agreement:** whether $\mathrm{Pr}^i\left(\left\|\phi_\infty^1 - \phi_\infty^2\right\| = 0\right) = 1$ 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.[12] 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.[13]

3   A Binary Example

In this section, we illustrate the main idea of the paper using the following simple example. We take $\Theta = \{A, B\}$ and $\Sigma = \{a, b\}$, so that there are two states and the signals are binary.

3.1   A Standard Model

We first consider an instance of the standard model, where it is commonly known that $p_{A,a} = \hat{p}_A$ and $p_{B,b} = \hat{p}_B$ for some $\hat{p}_A, \hat{p}_B \in (1/2, 1]$. Now, suppose that $r_{a,n}$ out of the first n signals are a. The likelihood of this event is $\hat{p}_A^{r_{a,n}}(1 - \hat{p}_A)^{n - r_{a,n}}$ under $\theta = A$ and $\hat{p}_B^{n - r_{a,n}}(1 - \hat{p}_B)^{r_{a,n}}$ under B. The relative likelihood ratio is

$$R\left(r_{a,n}, n\right) = \left(\frac{\hat{p}_A^{r_{a,n}/n}(1 - \hat{p}_A)^{1 - r_{a,n}/n}}{\hat{p}_B^{1 - r_{a,n}/n}(1 - \hat{p}_B)^{r_{a,n}/n}}\right)^n .$$

Suppose that $n \to \infty$ and $r_{a,n}/n \to \rho_a$. If $\hat{p}_A^{\rho_a}(1 - \hat{p}_A)^{1 - \rho_a} > \hat{p}_B^{1 - \rho_a}(1 - \hat{p}_B)^{\rho_a}$ (so that ρ_a is more likely under A than under B), then the expression in the parentheses is greater than 1, and $R\left(r_{a,n}, n\right) \to \infty$. In that case, asymptotically, each individual assigns probability 1 to state A. When the true state is A, $r_{a,n}/n \to \hat{p}_A$. Therefore, as $n \to \infty$, both individuals assign probability 1 to the true state,. Similarly, when the true state is B, $r_{a,n}/n$ converges to $1 - \hat{p}_B$, and since the expression in the parentheses is now less than 1, $R\left(r_{a,n}, n\right) \to 0$. In that case, both individuals assign probability 1 to the true state B, asymptotically. Except for the knife-edge case $\hat{p}_A^{\rho_a}(1 - \hat{p}_A)^{1 - \rho_a} = \hat{p}_B^{1 - \rho_a}(1 - \hat{p}_B)^{\rho_a}$, both individuals asymptotically assign probability 1 to the same state, leading to asymptotic agreement.

---

[12]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 if and only if the corresponding limits hold for almost all long-run frequencies $\rho(s) \in \Delta(\Sigma)$ under the Lebesgue measure, which has also an "objective" meaning.

[13]In a strategic situation, the players may care both about θ and the future signal frequencies. In that case, for asymptotic agreement, one may ask that both posteriors about θ and the future frequencies merge. The two concepts are equivalent under the full support assumption, since, in this case, the beliefs about the future frequencies merge.

3.2 A Model with Small Uncertainty

Now, we consider the case in which each individual faces a small amount of uncertainty about the conditional signal distributions. The individuals' beliefs about the conditional distributions are slightly different. For some small $\epsilon, \lambda \in (0,1)$, each individual i thinks that with probability $1 - \epsilon$, p_θ is in a λ-neighborhood of some $\hat{p}_\theta^i > (1 + \lambda)/2$, but with probability ϵ, 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

$$f_\theta^i (p_\theta) = \begin{cases} \epsilon + (1 - \epsilon)/\lambda & \text{if } p_{\theta,\theta} \in (\hat{p}_\theta^i - \lambda/2, \hat{p}_\theta^i + \lambda/2) \\ \epsilon & \text{otherwise} \end{cases} \tag{6}$$

for each θ and i. Here, $p_{\theta,\theta}$ is $p_{A,a}$ for $\theta = A$ and $p_{B,b}$ for $\theta = B$. Also, one can pick $\hat{p}_\theta^1$, $\hat{p}_\theta^2$, and $\hat{p}_\theta$ close to each other and pick ϵ and λ very small, so that each individual's beliefs are approximately the same as in the standard model discussed in the previous subsection. Now, by (5), the asymptotic likelihood ratio is

$$R_{A,B}^i (\rho) = \begin{cases} \frac{\epsilon\lambda}{1 - \epsilon(1-\lambda)} & if \rho_a \in D_A^i \equiv (\hat{p}_A^i - \lambda/2, \hat{p}_A^i + \lambda/2) \\[2mm] \frac{1 - \epsilon(1-\lambda)}{\epsilon\lambda} & if \rho_a \in D_B^i \equiv (1 - \hat{p}_B^i - \lambda/2, 1 - \hat{p}_B^i + \lambda/2) \\[2mm] 1 & \text{otherwise.} \end{cases} \tag{7}$$

This and other relevant functions are plotted in Figure 1 for $\epsilon \to 0, \lambda \to 0$. The likelihood ratio $R_{A,B}^i (\rho)$ is 1 when ρ_a 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, $\phi_{A,\infty}^i (\rho)$ will also be non-monotone: when ρ_a is small, the signals are not informative, thus $\phi_{A,\infty}^i (\rho)$ is the same as the prior, π_A^i. In contrast, around $\rho_a = 1 - \hat{p}_B^i$, the signals become very informative suggesting that the state is B, thus $\phi_{A,\infty}^i (\rho) \cong 0$. After this point, the signals become uninformative again and $\phi_{A,\infty}^i (\rho)$ goes back to π_A^i. Around $\hat{p}_A^i$, the signals are again informative, but this time favoring state A, so $\phi_{A,\infty}^i (\rho) \cong 1$. Finally, signals again become uninformative and $\phi_{A,\infty}^i (\rho)$ falls back to π_A^i.

Intuitively, when $\rho_a (s)$ is around $1 - \hat{p}_B^i$ or $\hat{p}_A^i$, 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 $\phi_{A,\infty}^i$ 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 θ. This is because, as $\epsilon \to 0$ and $\lambda \to 0$, each individual becomes convinced that the limiting frequencies will be close to either $1 - \hat{p}_B^i$ or $\hat{p}_A^i$. Thus, there is asymptotic learning in this environment.
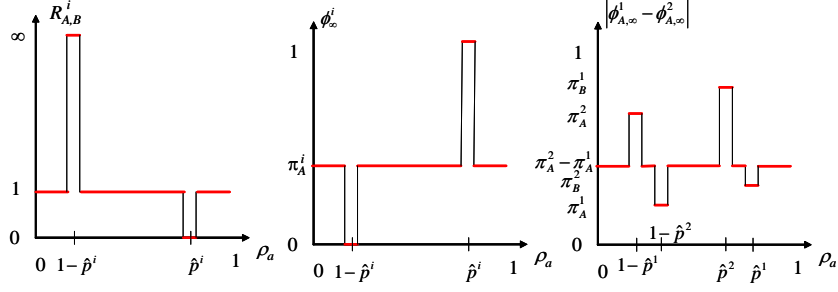
Figure 1: The three panels show, respectively, the approximate values of $R^i_{A,B}$, $\phi^i_{A,\infty}$, and $\left|\phi^1_{A,\infty} - \phi^2_{A,\infty}\right|$ as $\epsilon \to 0$ for $\hat{p}^i_A = \hat{p}^i_B = \hat{p}^i$.

However, asymptotic learning is considerably weaker than asymptotic agreement. Each individual also understands that since $\lambda < \left|\hat{p}^1_\theta - \hat{p}^2_\theta\right|$, 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 θ himself but that the other individual will fail to learn. Therefore, each 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

$$\tilde{z} \equiv \min\left\{\pi^1_A, \pi^2_A, \pi^1_B, \pi^2_B, \left|\pi^1_A - \pi^2_A\right|\right\}. \tag{8}$$

When $\pi^1_A = 1/3$ and $\pi^2_A = 2/3$, 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, $\Pr^i\left(\left|\phi^1_{A,\infty}(\rho) - \phi^2_{A,\infty}(\rho)\right| \geq Z\right) \geq 1 - \epsilon$, where as $\epsilon \to 0$,

$$Z \to z \equiv \min\left\{\pi^1_A, \pi^2_A, \pi^1_B, \pi^2_B\right\}. \tag{9}$$

This "subjective" bound can be as high as $1/2$.

Clearly, we can pick $\hat{p}^1_\theta \cong \hat{p}^2_\theta \cong \hat{p}_\theta$ and $\epsilon \cong \lambda \cong 0$, so that each individual's beliefs is approximately the same as in the above instance of the standard model. Yet, our individuals are certain that their beliefs will remain far apart as they observe the public signals, while the individuals in the standard model are certain that their beliefs will merge eventually.[14]

---

[14] Note that both heterogeneous beliefs about the interpretation of signals, i.e., $\hat{p}^1_\theta \neq \hat{p}^2_\theta$, and uncertainty, i.e., $\epsilon \neq 0 \neq \lambda$, are important for this discontinuity.

3.3 A Model of Asset Trading

Failure of asymptotic agreement may have major implications in a game theoretical analysis, as the following simple asset trading example illustrates.[15]

There are three dates, $\tau = 0, 1, 2$, two players and and an asset that yields a dividend only at date 2. The asset pays 1 if the state is A and 0 if the state is B. Player 2 owns the asset, but Player 1 would like to buy it because he is more optimistic: $\pi_A^1 > \pi_A^2$. Between the dates $\tau = 0$ and $\tau = 1$, the players observe a sequence $s \equiv \{s_t\}_{t=1}^{\infty}$ of signals. Player 1 has all the bargaining power: at each date Player 1 makes a take-it-or-leave-it price offer Q_τ, and trade occurs at price Q_τ if Player 2 accepts the offer. We are interested in the subgame-perfect equilibrium of this game.

Let us start with the standard (learning) model. At $\tau = 1$, after observing s, except for the knife-edge case, the individuals have the same belief about θ, and hence they are indifferent toward trading the asset (at price $Q_1 = \phi_{A,\infty}^1(\rho(s)) = \phi_{A,\infty}^2(\rho(s))$). In particular, at $\tau = 0$, both individuals believe that the price at $\tau = 1$ will be 1 if the state is A and 0 if the state is B, leaving both players indifferent. Hence, if trade does not occur at $\tau = 0$, the continuation value of Player 1 is 0, and the continuation value of Player 2 is π_A^2. If they trade at price Q_0, then the continuation value of players 1 and 2 will be $\pi_A^1 - Q_0$ and Q_0, respectively. Thus at date 0, Player 2 will accept an offer if and only if $Q_0 \geq \pi_A^2$. Since $\pi_A^1 > \pi_A^2$, Player 1 offers $Q_0 = \pi_0^2$ at date $\tau = 0$ and trade takes place. Therefore, in any subgame-perfect equilibrium, there is immediate trade at $\tau = 0$.

We next turn to the model with small uncertainty, discussed in the previous subsection. In particular, suppose that in terms of the example there we have $\epsilon \cong \lambda \cong 0$. We will then show that trade is delayed in equilibrium. Now, at date 1, if $\rho_a(s) \in D_A^1 \equiv \left(\hat{p}_A^1 - \lambda/2, \hat{p}_A^1 + \lambda/2\right)$, then the value of the asset for Player 2 is $\phi_{A,\infty}^2(\rho(s)) = \pi_A^2$, and the value of the asset for Player 1 is approximately 1. Hence, at such $\rho(s)$, Player 1 will buy the asset from Player 2 at price $Q_1(\rho(s)) = \pi_A^2$, enjoying gains from trade equal to $1 - \pi_A^2$. On the other hand, if $\rho_a(s) \in D_B^1$ or $\rho_a(s) \in D_A^2$, there will be no trade at date 1. For example, if $\rho_a(s) \in D_B^1$, Player 2 assigns probability π_A^2 to state A, accepting an offer Q_1 only if $Q_1 \geq \pi_A^2$. But since Player 1 assigns nearly probability 0 to state A, he would prefer not to trade at such a price. The continuation value of Player 1 is therefore approximately equal to

$$\pi_A^1\left(1 - \pi_A^2\right)$$

(when $\epsilon \to 0$). The continuation value of Player 2 in this case is π_A^2, since he only trades at his continuation value. Therefore, at date 0, Player 2 would accept a price offer Q_0 only if $Q_0 \geq \pi_A^2$. But such an offer would leave Player 1 at most a surplus of $\pi_A^1 - \pi_A^2$. Since $\pi_A^1 - \pi_A^2$ is strictly less than his continuation value $\pi_A^1 \left(1 - \pi_A^2\right)$, there will be no trade at $\tau = 0$. Instead, in any subgame-perfect equilibrium, Player 1 waits for the information to buy the asset at date 1 (provided that $\rho(s)$ turns out to be in a range where he concludes that the asset pays 1).

This example highlights two important implications of the type of learning analyzed in this paper for game theoretic and economic analysis. First, for players' behavior at $\tau = 1$ after observing s the crucial question is whether their beliefs will be close to each other after observing s at $\tau = 1$. If so, then their behavior will be very similar to that postulated in the standard (learning) model. Second, for players' behavior at $\tau = 0$, the crucial question is whether each player assigns high probability at $\tau = 0$ to the event that their beliefs will be similar at $\tau = 1$ (as in our definition of asymptotic agreement). If so, then the continuation values of each player will be as in the standard model, leading them to behave accordingly. Otherwise, they may behave quite differently, as our example has illustrated.

This example also illustrates that it is not sufficient for each individual to be certain that he will learn the truth for the equilibrium to be similar to that resulting under do standard (learning) model. When individuals are uncertain about p_θ, each is certain that he will learn the true state but is also certain that the other player will fail to do so. This assessment then induces each to wait for the arrival of additional information before trading.[16]

4   Asymptotic Learning and Agreement—Preliminary Results

In this section we return to the general environment introduced in Section 2 and consider two polar benchmarks, one in which the model is fully identified as in the standard model and one under our full-support assumption, Assumption 1. In each case we characterize whether the individuals will eventually learn the truth and whether they will agree asymptotically. Under full identification, asymptotic learning and agreement apply under relatively mild conditions. Under full support, the individuals never learn the truth and their beliefs "generically" diverge. These preliminary results will provide the starting point for our main results on fragility of asymptotic agreement presented in the next section.

---

[16]This contrasts with the intuition that observation of common information should take agents towards common beliefs and make trades less likely. This intuition is correct in models of learning under full identification and is the reason why previous models have generated speculative trade early in the game (e.g., Harrison and Kreps (1978) and Morris (1996)).

4.1 Asymptotic Learning and Agreement with Full Identification

We write $\text{supp}\big(F_\theta^i\big)$ for the smallest closed set to which F_θ^i assigns probability 1. That is, $\int_{\text{supp}(F)} dF_\theta^i = 1$ and for any $x \in \text{supp}\big(F_\theta^i\big)$ and any open neighborhood V of x, $\int_V dF_\theta^i > 0$. In this subsection, we assume that the supports of the frequencies are disjoint under different states, which ensures that the model is fully identified.

Assumption 2 *(**Full Identification**) For all $\theta \neq \theta'$ and i, $\text{supp}\big(F_\theta^i\big) \cap \text{supp}\big(F_{\theta'}^i\big) = \varnothing$.*

We next establish asymptotic learning and agreement under this full-identification assumption and an additional genericity assumption. To state this characterization result, we first define the *cross-entropy distance* between ρ and $p \in \Delta\left(\Sigma\right)$ as:[17]

$$H\left(\rho, p\right) \equiv -\sum_\sigma \rho_\sigma \log\left(p_\sigma\right).$$

We also define

$$P^i\left(\rho\right) \equiv \arg \min_{p \in \cup_\theta \text{supp}\left(F_\theta^i\right)} H\left(\rho, p\right)$$

as the set of frequency vectors p (among those that i initially finds possible) that are closest to ρ according to cross-entropy H. We use $\#P^i\left(\rho\right)$ to denote the number of elements in the set $P^i\left(\rho\right)$. Our genericity assumption is presented next.

Assumption 3 *(**Genericity**) For all $\theta \in \Theta$ and each $i \in \{1, 2\}$, $\text{supp}\big(F_\theta^i\big)$ is in the interior of $\Delta\left(\Sigma\right)$ and for each $\rho \in \text{supp}\big(F_\theta^i\big)$, $\#P^j(\rho) = 1$ (for $j \neq i$).*

This assumption is imposed for expositional simplicity and will be used only in the following result. It holds generically in the sense that any F_θ^i can be approximated by distributions that satisfy it. (Note that for any $\hat{p} \in P^j\left(\rho\right)$ and $\varepsilon \in (0, 1)$, $P^j\left(\varepsilon\rho + (1 - \varepsilon)\hat{p}\right) = \{\hat{p}\}$.)

Theorem 1 *(**Asymptotic Learning and Agreement under Full Identification**) Suppose Assumptions 2 and 3 hold. Then for each $i \in \{1, 2\}$ and $j \neq i$:*

1. $\Pr^i\left(\lim_{n\to\infty} \phi_{\theta,n}^i\left(s\right) = 1 | \theta\right) = 1$.

2. $\Pr^i\left(\lim_{n\to\infty} \left|\phi_{\theta,n}^1\left(s\right) - \phi_{\theta,n}^2\left(s\right)\right| = 0\right) = 1$ if and only if $P^j(\rho) \subseteq \text{supp}(F_\theta^j)$ for all $\rho \in \text{supp}(F_\theta^i)$ and all θ.

---

[17] The cross-entropy function is closely related to Kullback-Leibler divergence in information theory and is often used as a measure of distance between two probability distributions. In particular, the Kullback-Leibler divergence of p from ρ is $H\left(\rho, p\right) - H(p, p)$, where $H\left(p, p\right)$ is also the entropy of p. See, for example, Burnham and Anderson (1989).

The first part of the theorem states that full identification ensures asymptotic learning. The second part provides necessary and sufficient conditions for asymptotic agreement. In particular, under full identification there will be asymptotic agreement if and only if the supports of F_θ^1 and F_θ^2 are close enough according to cross-entropy, in the sense that the closest point among $\cup_{\theta'} \text{supp}(F_{\theta'}^j)$ to a point in $\text{supp}(F_\theta^i)$ lies in $\text{supp}(F_\theta^j)$. It is remarkable that asymptotic agreement is a property of the supports. In particular, the common support assumption implies asymptotic agreement. This is stated in the next corollary.

Corollary 1 (Common Support) *Under Assumptions 2, if* $\text{supp}(F_\theta^1) = \text{supp}(F_\theta^2)$ *for all* θ, *then* $\text{Pr}^i \left(\phi_\infty^1 = \phi_\infty^2 \right) = 1$ *for each* i.

A special case of the corollary is the following well-known result of Savage (1954), which was already discussed in the context of the Binary Example. Savage's result is the basis of the argument that Bayesian learning will push individuals towards common beliefs and priors.

Corollary 2 (Savage's Theorem) *Assume that each* F_θ^i *puts probability 1 on* $\hat{p}_\theta$ *for some* $\hat{p}_\theta$ *such that* $\hat{p}_\theta \neq \hat{p}_{\theta'}$ *for all* $\theta \neq \theta'$. *Then asymptotic learning and agreement always obtain, i.e., for each* $i = 1,2$,

1. $\text{Pr}^i \left(\lim_{n \to \infty} \phi_{\theta,n}^i (s) = 1 | \theta \right) = 1.$

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

It is useful to spell out the intuition for Theorem 1 and Corollary 2. Let us start with the latter. Corollary 2 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 on what the underlying state, θ, is. A simple intuition for this result is that the underlying state θ 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.[18]

However, there is more to this corollary than this simple intuition. Each individual is sure that, at state θ, they will be confronted with a limiting frequency of $\hat{p}_\theta$, in which case they

---

will conclude that the true state is θ; and they attach zero probability to the events that they will observe a different asymptotic frequency. To elaborate this point, consider the following crucial question:

What happens if an individual observes a frequency ρ of signals different from every $\hat{p}_\theta$ in a large sample of size n?

The answer to this question highlights why asymptotic agreement under the standard model rests on problematic assumptions and also provides the intuition for our characterization and some of our main results. First, note that the event in the question has zero probability under the individual's beliefs at the limit $n = \infty$. However, for $n < \infty$ any frequency has strictly positive probability because of *sampling variation.* In particular, the individual expects frequency ρ to occur with probability $e^{-H(\rho,\hat{p}_\theta)n}$ under state θ when the sample size is n. For any θ with $\hat{p}_\theta \in P^i(\rho)$ and θ' with $\hat{p}_{\theta'} \notin P^i(\rho)$, this event is *infinitely more likely* under θ than under θ' in the sense that $e^{-H(\rho,\hat{p}_\theta)n}/e^{-H(\rho,\hat{p}_{\theta'})n} \to \infty$ as $n \to \infty$. Therefore, when $P^i(\rho) = \{\hat{p}_{\theta(\rho)}\}$ for some $\theta(\rho) \in \Theta$, as $n \to \infty$, he becomes increasingly certain that the long-run frequency is $\hat{p}_{\theta(\rho)}$, which is distinct from ρ, and assigns probability 1 to state $\theta(\rho)$. Crucially, the individual reaches this inference despite the fact that as $n \to \infty$, frequency ρ has zero probability under $\theta(\rho)$. The fact that full identification relies on inference based on such zero probability events is the problematic aspect of the standard model (see below for a further discussion).

The intuition for Theorem 1 is very similar to that of Corollary 2. Now, i believes that as $n \to \infty$, under each θ, the empirical frequency will converge to some $\rho \in \mathrm{supp}(F_\theta^i)$, and i will assign probability 1 to state θ. What happens if $\rho \notin \cup_{\theta'}\mathrm{supp}(F_{\theta'}^j)$? Then, j presumes that this has resulted from sampling variation. That is, he thinks that the actual long-run frequency is some $p \in \cup_{\theta'}\mathrm{supp}(F_{\theta'}^j)$, but the sampling variation leads to ρ, which has probability $e^{-H(\rho,p)n}$. Suppose $P^j(\rho)$ is a singleton with its unique element $\hat{p}_{\theta^j(\rho)}$ from $\mathrm{supp}(F_{\theta^j(\rho)}^j)$ for some $\theta^j(\rho) \in \Theta$. Then, once again, as $n \to \infty$, the likelihood of ρ is infinitely greater under $\hat{p}_{\theta^j(\rho)}$ than any other p. Consequently, j assigns probability 1 to the event that the actual long-run frequency is $\hat{p}_{\theta^j(\rho)}$ and the true state is $\theta^j(\rho)$. There will be asymptotic agreement on such a sample path if and only if $\theta^j(\rho) = \theta$. When the supports of F_θ^i and F_θ^j are sufficiently close, this condition is satisfied and there will be asymptotic agreement. In contrast, if the supports were far apart, so that $\theta^j(\rho) \neq \theta$, then i would assign positive probability to the event that their beliefs would diverge to the extremes and we would thus have $\lim_{n\to\infty} \left\| \phi_n^1(s) - \phi_n^2(s) \right\| = 1$.

18

Now we can see even more explicitly that the behavior of individual i's beliefs when $\rho\left(s\right) \notin \cup_{\theta'}\mathrm{supp}(F_{\theta'}^i)$ is problematic. First, individual i never questions the validity of his model even though the reality is increasingly inconsistent with this model (since $\rho\left(s\right) \notin \cup_{\theta'}\mathrm{supp}(F_{\theta'}^i)$). As we discussed in the Introduction, in practice we expect individuals to revise their models and working hypotheses when faced with overwhelming evidence inconsistent with their assumptions. Second, individual i's beliefs (at sample path s) concerning future frequencies diverge from the actual empirical frequency. In particular, we have

$$\lim_{n\to\infty}(\mathrm{Pr}^i\left(s_{n+1} = \sigma|\left\{s_t\right\}_{t=1}^{t=n}\right))_{\sigma\in\Sigma} \neq \rho\left(s\right)$$

because $\lim_{n\to\infty}(\mathrm{Pr}^i\left(s_{n+1} = \sigma|\left\{s_t\right\}_{t=1}^{t=n}\right))_{\sigma\in\Sigma} \in P^i\left(\rho\left(s\right)\right) \subset \cup_{\theta'}\mathrm{supp}(F_{\theta'}^i)$ while $\rho\left(s\right) \notin \cup_{\theta'}\mathrm{supp}(F_{\theta'}^i)$. This is despite the fact that his model prescribes a symmetric world due to exchangeability. In contrast, under the full-support assumption, we have

$$\lim_{n\to\infty}(\mathrm{Pr}^i\left(s_{n+1} = \sigma|\left\{s_t\right\}_{t=1}^{t=n}\right))_{\sigma\in\Sigma} = \rho\left(s\right) \qquad \forall s \in \bar{S},$$

so that individuals' beliefs about the future frequencies always converge to the empirical frequency.[19]

This discussion suggests that the full-identification assumption leads to a range of problematic conclusions when individuals observe "surprising" frequencies (because individuals always ascribe these frequencies to sampling variability). It also suggests that equating small probability events with zero probability events may have important implications and consequences. To investigate these issues systematically, we next turn to the case in which the full-support assumption applies.

4.2   Failure of Asymptotic Learning and Agreement with Full Support

We next impose Assumption 1 and show that when F_θ^i has full support, there will be neither asymptotic learning nor asymptotic agreement.

Theorem 2 *(Lack of Asymptotic Learning and Agreement Under Full Support)*
Under Assumption 1, for all θ and i,

   1. $\mathrm{Pr}^i\left(\phi_{\theta,\infty}^i\left(\rho\left(s\right)\right) \neq 1|\theta\right) = 1$, *and*

---

[19]This claim can be readily proved following the same steps as in the proof of Lemma 3 in the Appendix since $P^i\left(\rho\right) = \left\{\rho\right\}$ for all ρ and i under the full-support assumption.

19

2. *under the Euclidean metric on real numbers and the $\mathcal{L}^p$ metric on density functions, there exits an open and dense set of models $\left(\pi_\theta^i, f_\theta^i\right)_{i,\theta}$ such that for each $i \in \{1,2\}$, $\Pr^i\left(\phi_{\theta,\infty}^1\left(\rho\left(s\right)\right) \neq \phi_{\theta,\infty}^2\left(\rho\left(s\right)\right)\right) = 1$.*

Under the full support assumption, Theorem 2 leads to the opposite of the conclusions drawn from Theorem 1. The first conclusion of Theorem 2 states that each individual is certain that he will not learn the truth. This is simply because the likelihood ratio $R_{\theta,\theta'}^i$ never takes values of 0 or ∞, and hence, by Lemma 1, i would never put probability 1 on any state. That is, on every sample path $s \in S$, his probability assessment of each state remains bounded away from 0, eventually. The second conclusion states that, generically, he is also certain that their beliefs will never merge, and there will always be some difference of opinions. This is also an immediate implication of Lemma 1. Under the full-support assumption, the asymptotic beliefs vary with the parameters of the individual's subjective model, and asymptotic agreement at any ρ boils down to a family of equality restrictions, indexed by $\theta \in \Theta$:

$$\sum_{\theta' \neq \theta} \frac{\pi_{\theta'}^1}{\pi_\theta^1} R_{\theta,\theta'}^1\left(\rho\left(s\right)\right) = \sum_{\theta' \neq \theta} \frac{\pi_{\theta'}^2}{\pi_\theta^2} R_{\theta,\theta'}^2\left(\rho\left(s\right)\right).$$

Clearly, these equalities are "knife-edge" and a small perturbation always ensures that they are not satisfied almost everywhere (according to the Lebesgue measure on $\Delta\left(\Sigma\right)$). Thus, for almost all learning models $\left(\pi_\theta^i, f_\theta^i\right)_{i,\theta}$, there will be no asymptotic agreement under full support.

There is a simple intuition for the lack of learning. Under full support, an individual is never sure about the exact interpretation of the sequence of signals he observes and will update his views about p_θ (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. This is captured by the fact that $R_{\theta,\theta'}^i\left(\rho\right)$ defined in (5) never takes the value zero or infinity.

Nevertheless, Theorem 1 may be more robust than Theorem 2 appears to suggest. Under the idealized full-identification assumption, Theorem 1 showed that individuals expect to assign exactly probability 1 to the true state and the asymptotic disagreement is exactly 0 (under some additional mile conditions). When this idealized assumption is relaxed, these conclusions may no longer hold. But this is simply because the individuals do not assign exactly probability 1 to any state. This does not, however, rule out the possibility that Theorem 1 is robust in the sense

that the amount of asymptotic disagreement is small in the neighborhood of fully identified models, e.g., when the amount of uncertainty about the conditional signal distributions is small.

5 Fragility of Asymptotic Agreement—Main Results

In this section, we investigate whether as the amount of uncertainty about the interpretation of the signals disappears, the amount of asymptotic disagreement vanishes continuously. We first show that this is not the case: one can perturb a standard model of learning under certainty slightly and obtain a model in which there is a substantial amount of asymptotic disagreement. We establish such a discontinuity *for every model* with certainty, including the canonical model of learning under certainty, where both individuals share the same beliefs regarding the conditional signal distributions. We then restrict our perturbations by imposing a uniform convergence assumption. Within this class, we characterize the perturbations under which the amount of asymptotic disagreement vanishes continuously.

For any $\hat{p} \in \Delta(\Sigma)$, write $\delta_{\hat{p}}$ for the Dirac distribution that puts probability 1 on $p = \hat{p}$. Let $\{F^i_{\theta,m}\}_{m \in \mathbb{N}, i \in \{1,2\}, \theta \in \Theta}$ ($\{F^i_{\theta,m}\}$ for short) denote an arbitrary sequence of subjective probability distributions converging to a Dirac distribution $\delta_{p^i_\theta}$ for each (i, θ) as $m \to \infty$. (We will simply say that $\{F^i_{\theta,m}\}$ converges to $\delta_{p^i_\theta}$). Throughout it is implicitly assumed that there is asymptotic agreement under $\delta_{p^i_\theta}$ (as in Theorem 1). Therefore, as $m \to \infty$, uncertainty about the interpretation of the signals disappears and we converge to a model with asymptotic agreement. We write Pr^i_m for the ex ante probability under $(F^i_{\theta,m})_{\theta \in \Theta}$, $\phi^i_{\theta,\infty,m}$ for the asymptotic posterior belief that true state is θ under $(F^i_{\theta,m})_{\theta \in \Theta}$, and $R^i_{\theta,\theta',m} = f^i_{\theta',m} / f^i_{\theta,m}$ for the long-run likelihood ratio.

We first observe that learning is *continuous at certainty* in the following sense:

Theorem 3 (Continuity of Asymptotic Learning) *Consider any $\{F^i_{\theta,m}\}$ converging to $\delta_{p^i_\theta}$ with $p^i_\theta \neq p^i_{\theta'}$ for all $\theta \neq \theta'$ and i. For any $\varepsilon > 0$ and (θ, i),*

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

Equivalently, for any $\left(\varepsilon, \theta, \theta', i \right)$ with $\varepsilon > 0$ and any $\theta \neq \theta'$,

$$\lim_{m \to \infty} \mathrm{Pr}^i_m \left(R^i_{\theta,\theta'm} < \varepsilon | \theta \right) = 1.$$

This theorem states that as $\{F^i_{\theta,m}\}$ converges to $\delta_{p^i_\theta}$, each individual becomes increasingly convinced that he will learn the true state. Equivalently, when a model of learning under

21

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 θ. Hence, the first part of Theorem 1, which concerns asymptotic learning, is robust.

We next define the continuity of asymptotic agreement at certainty and show that, in contrast to the first part, the second part of Theorem 1 is not robust.

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

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

*We say that asymptotic agreement is **continuous at certainty at** $\left(p^1, p^2 \right) \in \Delta\left(\Sigma \right)^{2K}$ if it is continuous at certainty under all families $\{F_{\theta,m}^i\}$ converging to $\delta_{p_\theta^i}$.*

Continuity at certainty requires that as the family of subjective probability distributions converge to a Dirac distribution (at which there is asymptotic agreement), each individual becomes increasingly certain that asymptotic disagreement will be arbitrarily small. Hence, asymptotic agreement is *discontinuous at certainty* at $\left(p_A^1, p_B^1, p_A^2, p_B^2 \right)$ if there exists a family $\{F_{\theta,m}^i\}$ converging to $\delta_{p_\theta^i}$ and $\varepsilon > 0$ such that

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

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** $\{F_{\theta,m}^i\}$ if there exists $\varepsilon > 0$ such that*

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

*for $i = 1, 2$. We say that **asymptotic agreement is strongly discontinuous at certainty at** $\left(p^1, p^2 \right) \in \Delta\left(\Sigma \right)^{2K}$ if it is strongly discontinuous at certainty under some family $\{F_{\theta,m}^i\}$ converging to $\delta_{p_\theta^i}$.*

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. Finally, we introduce an even a stronger notion of discontinuity.

Definition 3 *We say that* **asymptotic agreement is almost-surely discontinuous at certainty under** $\{F_{\theta,m}^i\}$ *if there exist $\varepsilon > 0$ and $\bar{m} < \infty$ such that for all $m > \bar{m}$ and $\rho \in \Delta(\Sigma)$,*

$$\left\| \phi_{\infty,m}^1(\rho) - \phi_{\infty,m}^2(\rho) \right\| > \varepsilon.$$

We say that **asymptotic agreement is almost-surely discontinuous at certainty at** $\left(p^1, p^2\right) \in \Delta(\Sigma)^{2K}$ *if it is almost-surely discontinuous at certainty under some family $\{F_{\theta,m}^i\}$ (with possibly discontinuous density functions) converging to $\delta_{p_\theta^i}$.*

Almost-sure discontinuity requires that on every sample path $s \in \bar{S}$, with well-defined long-run frequency $\rho(s)$, asymptotic beliefs will be bounded away from each other. Hence, almost surely, there will be a significant asymptotic disagreement. In contrast, strong discontinuity only requires that individuals assign high probabilities to those sample paths. Therefore, almost-sure discontinuity implies strong discontinuity, and strong discontinuity implies discontinuity.

5.1 Discontinuity of Asymptotic Agreement

Theorem 4 *(Strong Discontinuity)* *Asymptotic agreement is strongly discontinuous at certainty at each $\left(p^1, p^2\right) \in \Delta(\Sigma)^{2K}$. Moreover, if $\pi^1 \neq \pi^2$, then asymptotic agreement is almost-surely discontinuous at every $\left(p^1, p^2\right) \in \Delta(\Sigma)^{2K}$.*

That is, for every model with certainty, one can introduce a vanishingly small uncertainty in such a way that the individuals' asymptotic beliefs remain substantially different from each other at almost all sample paths. Therefore, the asymptotic agreement results are always fragile. The proof is based on a perturbation as in the Binary Example introduced in Section 3.

Example 1 *(Discontinuity of Asymptotic Agreement)* In Section 3.2, for each m, take $\epsilon = \lambda = \bar{\epsilon}/m$, $\hat{p}_\theta^1 = \hat{p}_\theta + \lambda$, and $\hat{p}_\theta^2 = \hat{p}_\theta - \lambda$ where $\bar{\epsilon}$ is such that $\phi_{B,\infty}^i(\rho) < \pi_B^j/2$ for $\rho_a \in D_A^i$ and $\phi_{A,\infty}^i(\rho) < \pi_A^j/2$ for $\rho_a \in D_B^i$ whenever $\epsilon = \lambda \leq \bar{\epsilon}$. (Recall from (7) that $D_A^i \equiv \left(\hat{p}_A^i - \lambda/2, \hat{p}_A^i + \lambda/2\right)$ and $D_B^i \equiv \left(1 - \hat{p}_B^i - \lambda/2, 1 - \hat{p}_B^i + \lambda/2\right)$). Such $\bar{\epsilon}$ exists (by asymptotic learning of i). By construction, each $F_{\theta,m}^i$ converges to the beliefs in the standard model, and $\left| \hat{p}_\theta^1 - \hat{p}_\theta^2 \right| > \lambda$ for each θ. To see strong discontinuity, consider the bound $\bar{Z} = z/2 > 0$ with z as in (9). By the choice of $\bar{\epsilon}$, $\left| \phi_{A,\infty,m}^1(\rho) - \phi_{A,\infty,m}^2(\rho) \right| > \bar{Z}$ whenever $\rho_a \in D_A^i \cup D_B^i$. But $\mathrm{Pr}_m^i \left(\rho_a(s) \in D_A^i \cup D_B^i\right) = 1 - \epsilon(1 - \lambda)$, which goes to 1 as $m \to \infty$. Therefore,

$$\lim_{m \to \infty} \mathrm{Pr}_m^i \left(\left| \phi_{A,\infty,m}^1 - \phi_{A,\infty,m}^2 \right| > \bar{Z} \right) = 1. \tag{10}$$

23

This establishes that agreement is strongly discontinuous under $\{F^i_{\theta,m}\}$.

To show almost-sure discontinuity, consider the bound $\tilde{Z} = \tilde{z}/2$, which is positive when $\pi^1 \neq \pi^2$, where $\tilde{z}$ is as defined in (8). Clearly, $\left|\phi^1_{A,\infty,m}(\rho) - \phi^2_{A,\infty,m}(\rho)\right| > \bar{Z} \geq \tilde{Z}$ when $\rho_a \in D^i_A \cup D^i_B$, and $\left|\phi^1_{A,\infty,m}(\rho) - \phi^2_{A,\infty,m}(\rho)\right| = \left|\pi^1_A - \pi^2_A\right| > \tilde{Z}$ when $\rho_a \notin D^i_A \cup D^i_B$. That is, at each frequency, the asymptotic disagreement exceeds $\tilde{Z}$, showing that agreement is almost-surely discontinuous under $\{F^i_{\theta,m}\}$.

In the Binary Example (and in the proof of Theorem 4), the likelihood ratio $R^i_{\theta,\theta',m}(\rho(s))$ and the asymptotic beliefs $\phi^i_{\theta,\infty,m}(\rho(s))$ are *non-monotone* in the frequency $\rho(s)$. This is a natural outcome of uncertainty on conditional signal distributions (see the discussion in the Introduction and Figure 2 below). In the Binary Example, when $R^i_{\theta,\theta',m}$ 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 5 (Discontinuity under Monotonicity) *Take* $\Theta = \{A, B\}$ *and* $\Sigma = \{a, b\}$. *Take any* $\pi^1_A, \pi^2_B \in (0,1)$ *and for each* $i \in \{1, 2\}$ *take any* $\hat{p}^i_A$ *and* $\hat{p}^i_B$ *with* $\hat{p}^i_{A,a} > 1/2$ *and* $\hat{p}^i_{B,b} > 1/2$. *There exist a family* $\{F^i_{\theta,m}\}$ *and such that:*

1. *$F^i_{\theta,m}$ converges to $\delta_{\hat{p}^i_\theta}$ for each θ and i;*

2. *the likelihood ratio $R^i_{A,B,m}(\rho)$ is nonincreasing in ρ_a for each i and m, where $\rho = (\rho_a, 1 - \rho_a)$, and*

3. *agreement is discontinuous at certainty under $\{F^i_{\theta,m}\}$.*

The monotonicity of the likelihood ratio has weakened the conclusion of Theorem 4. Now, asymptotic agreement is discontinuous at certainty, but it is not almost-surely or strongly discontinuous.

In the proof of Theorem 4, the family $\{F^i_{\theta,m}\}$ leading to the discontinuity of asymptotic agreement induces discontinuous likelihood ratios. This is not important for the results (except for almost-sure discontinuity) because smooth approximations to $F^i_{\theta,m}$ would ensure continuity of the likelihood ratios. More importantly, the likelihood ratio $R^i_{A,B,m}$ converges point-wise to a continuous function (i.e., $\lim_m R^i_{A,B,m}(\rho) = 1$ for each ρ), but this convergence is not uniform. The failure of uniform convergence is crucial for the stark discontinuity results above, as we establish next.

24

5.2 A Characterization of Continuity with Uniform Convergence

We now assume that the likelihood ratio function $R^i_{\theta,\theta',m}$ *uniformly* converges to a (continuous) function in the relevant regions and characterize the perturbations under which the asymptotic agreement is continuous.

Let $R^i_{\theta,\theta',\infty}(\rho) = \lim_{m\to\infty} R^i_{\theta,\theta',m}(\rho)$ be the pointwise limit of $R^i_{\theta,\theta',m}(\rho)$ in the extended reals. It must be emphasized that the limiting asymptotic likelihood ratio, $R^i_{\theta,\theta',\infty}$, is distinct from the asymptotic likelihood ratio in the limiting model. For example, in Example 1, $R^i_{\theta,\theta',\infty}$ is identically 1. On the other hand, the limiting model specifies that each individual i is certain that $p_{\theta,\theta} = \hat{p}_\theta$. Consequently, in the limiting model, the asymptotic likelihood ratio is 0 around $\rho_a = p_A$ and ∞ around $\rho_a = 1 - \hat{p}_B$. Under uniform convergence, our characterization theorem establishes that whether asymptotic agreement holds is tied to the value of $R^i_{\theta,\theta',\infty}$ at the relevant frequencies, and it has no connection to the likelihood ratio in the limiting model:

Theorem 6 (Characterization under Uniform Convergence) *Under Assumption 1, consider any $\{F^i_{\theta,m}\}$ converging to $\delta_{p^i_\theta}$ where $p^i_\theta \neq p^i_{\theta'}$ for all distinct θ and θ' and (m,i). Assume that for each (θ,θ',i,j), there exists an open neighborhood V^i_θ of p^i_θ on which $R^i_{\theta,\theta',m}$ converges uniformly to $R^i_{\theta,\theta',\infty}$. Then, the following are true.*

1. *Asymptotic agreement is continuous at certainty under $\{F^i_{\theta,m}\}$ if and only if*

$$R^j_{\theta,\theta',\infty}(p^i_\theta) = 0 \qquad \forall i \neq j, \theta \neq \theta'.$$

2. *If $R^j_{\theta,\theta',\infty}(p^i_\theta) \neq 0$ for all $i \neq j$ and $\theta \neq \theta'$, then asymptotic agreement is strongly discontinuous at certainty under $\{F^i_{\theta,m}\}$.*

The characterization establishes that whether asymptotic agreement is continuous at certainty depends on whether the limiting asymptotic likelihood ratio for states θ and θ' according to j, $R^j_{\theta,\theta',\infty}$, is equal to 0 at the frequency p^i_θ that will be realized under state θ according to the limiting model of i. The idea here is intuitive. Individual i is almost certain that, if the state is θ, then the realized frequency will be around p^i_θ and he will assign nearly probability 1 to θ. For the other individual j to agree with him, she must also assign nearly probability 1 to θ at those frequencies, which requires that her likelihood ratio, $R^j_{\theta,\theta',m}$, is nearly 0 at those frequencies. But these ratios are all approximately equal to $R^j_{\theta,\theta',\infty}(p^i_\theta)$ when $R^j_{\theta,\theta',m}$ converges uniformly.

In Theorem 6, the uniform convergence assumption is not superfluous for either direction of the characterization, as the next counterexample shows.

Example 2 *(Counterexample to Necessity)* Without uniform convergence, there may be asymptotic agreement even if $R^j_{\theta,\theta',\infty} \neq 0$. In the Binary Example from Section 3, suppose instead that

$$f^i_{\theta,m}(p) = \begin{cases} 1/m + m(1 - 1/m) & \text{if } p \in \left(\hat{p}_m - \frac{1}{2m}, \hat{p}_m + \frac{1}{2m}\right) \\ 1/m & \text{otherwise} \end{cases}$$

where $\hat{p}_m = 3/4 - 1/m$. The limiting asymptotic likelihood ratio, $R^j_{\theta,\theta',\infty}$, is identically 1. Nevertheless, agreement is continuous under $\{F^i_{\theta,m}\}$ because the individuals have the same model of learning.

(Counterexample to Sufficiency) Without uniform convergence, asymptotic agreement may fail even if $R^j_{\theta,\theta',\infty} = 0$. Consider the following variation of the Binary Example:

$$f^i_{\theta,m}(p_\theta) = \begin{cases} x_m & \text{if } p_{\theta,\theta} \in \left(\hat{p}^i_m - 1/m, \hat{p}_m + 1/m\right) \\ 1/m & \text{if } p_{\theta,\theta} \in [1/2, 1] \setminus \left(\hat{p}^i_m - 7/m, \hat{p}_m + 7/m\right) \\ 1/m^2 & \text{otherwise} \end{cases}$$

for large m where $\hat{p}^1_m = 3/4 + 14/m$, $\hat{p}^2_m = 3/4 + 18/m$, and x_m is such that $f^i_{\theta,m}$ is a density function. One can check that $R^j_{\theta,\theta',\infty}(\rho) = 0$ whenever $\rho_\theta > 1/2$. Nevertheless, agreement is still strongly discontinuous under $\{F^i_{\theta,m}\}$.

We next explore the implications of the characterization results in Theorem 6 in the context of a canonical example.

5.3 A Canonical Example

As in the Binary Example, we take $\Theta = \{A, B\}$ and $\Sigma = \{a, b\}$. We consider a class of "symmetric" families $\{F^i_{\theta,m}\}$ that converge to the Dirac distribution $\delta_{p^i_\theta}$ where $p^i_A = \left(\hat{p}^i, 1 - \hat{p}^i\right)$ and $p^i_B = \left(1 - \hat{p}^i, \hat{p}^i\right)$ for some $\hat{p}^i \in (1/2, 1)$. The family is parameterized by a *determining* density function $f : \mathbb{R} \to \mathbb{R}$ that is strictly positive, symmetric around zero, and monotone in the tails. The subjective density function $f^i_{\theta,m}$ is then induced by f and the transformation $x \mapsto \left(x - \hat{p}^i\right)/m$:

$$f^i_{\theta,m}(\rho) = c^i(m) f\left(m\left(\rho_\theta - \hat{p}^i\right)\right) \tag{11}$$

where $c^i(m) \equiv 1/\int_0^1 f\left(m\left(\rho_\theta - \hat{p}^i\right)\right) d\rho_\theta$ is a correction factor to ensure that $f^i_{\theta,m}$ is a proper probability density function on $\Delta(\Sigma)$. Here, the mapping $x \mapsto \left(x - \hat{p}^i\right)/m$ scales down the real line around $\hat{p}^i$ by the factor $1/m$, scaling down the uncertainty about p_θ by $1/m$. As $m \to \infty$, the uncertainty vanishes and $F^i_{\theta,m}$ converges to $\delta_{p^i_\theta}$. When $\hat{p}^1 = \hat{p}^2$, the individuals have the same subjective densities, leading trivially to asymptotic agreement. To analyze

26

the continuity of asymptotic agreement at certainty, we will assume $\hat{p}^1 \neq \hat{p}^2$. Define $\hat{x} \equiv \hat{p}^1 + \hat{p}^2 - 1 > |\hat{p}^1 - \hat{p}^2| \equiv \hat{y} > 0$.

The relevant asymptotic likelihood ratios are all equal to

$$R_{\theta,\theta',m}^j \left(p_\theta^i\right) = \frac{f\left(m\left(\hat{p}^1 + \hat{p}^2 - 1\right)\right)}{f\left(m\left|\hat{p}^1 - \hat{p}^2\right|\right)} = \frac{f\left(m\hat{x}\right)}{f\left(m\hat{y}\right)}.$$

Hence, under the uniform convergence assumption, whether asymptotic agreement is continuous at certainty under $\{F_{\theta,m}^i\}$ depends on whether

$$\lim_{m\to\infty} \frac{f\left(m\hat{x}\right)}{f\left(m\hat{y}\right)} = 0. \tag{12}$$

This is a familiar condition in statistics. Whether it is satisfied depends on whether f has rapidly-varying (exponential) or regularly-varying (polynomial) tails:

Definition 4 *A density function f has **regularly-varying tails** if*

$$\lim_{m\to\infty} \frac{f(mx)}{f(m)} = H(x) \in \mathbb{R} \qquad \forall x > 0.$$

The condition that $H(x) \in \mathbb{R}$ is relatively weak, but nonetheless has important implications. In particular, it implies that $H(x) \equiv x^{-\alpha}$ for $\alpha \in (0,\infty)$.[20] Moreover, Seneta (1976) shows that the convergence in Definition 4 is uniform on any compact set. Therefore, if f has regularly-varying tails, then $R_{\theta,\theta',m}^j$ converges uniformly to $R_{\theta,\theta',\infty}^j$, where

$$R_{\theta,\theta',\infty}^j \left(p_\theta^i\right) = (\hat{x}/\hat{y})^{-\alpha} > 0$$

for some $\alpha \in (0,\infty)$. Our characterization then shows that asymptotic agreement is strongly discontinuous it certainty under $\{F_{\theta,m}^i\}$. Many common distributions, including the Pareto, log-normal, and t-distributions, have regularly-varying densities.

Definition 5 *A density function f has **rapidly-varying tails** if for every $x > 0$,*

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

Once again, the convergence is uniform on compact intervals that exclude $x = 1$ (as in our case), and hence each $R_{\theta,\theta',m}^j$ converges uniformly to $R_{\theta,\theta',\infty}^j$ where

$$R_{\theta,\theta',\infty}^j \left(p_\theta^i\right) = (\hat{x}/\hat{y})^{-\infty} = 0.$$

---

[20] The proof of this claim is provided in our working paper. See, for example, De Haan (1970).

Our characterization then shows that if f has rapidly varying tails, then agreement is continuous under $\{F_{\theta,m}^i\}$. Examples of densities with rapidly-varying tails include the exponential and the normal densities.

Therefore, whether there is asymptotic agreement depends on the tails:

1. If f has regularly-varying tails, then agreement is continuous at certainty under $\{F_{\theta,m}^i\}$.

2. If f has rapidly-varying tails, then agreement is strongly discontinuous at certainty under $\{F_{\theta,m}^i\}$.

The intuition for these results is as follows. The continuity of agreement is determined by whether $R_{\theta,\theta',m}^j \left(p_\theta^i \right) \cong 0$. That is, whether i thinks that the frequency that will be realized at state θ (namely p_θ^i) will convince the other individual j, too, that the state is θ. Now, since $\hat{y} = \left| \hat{p}^i - \hat{p}^j \right| < \left| \hat{p}^i - \left(1 - \hat{p}^j\right) \right| = \hat{x}$, j also considers frequency p_θ^i as an evidence for state θ. For large m, the strength of this evidence depends on the tail of f. Rapidly-varying tails are increasingly informative. Any difference between $\hat{x}$ and $\hat{y}$ is magnified as we go to the tail frequencies (as m increases). Hence, as $m \to \infty$, j infers from the frequency p_θ^i that the state is θ (i.e., $R_{\theta,\theta',\infty}^j \left(p_\theta^i \right) = 0$). On the other hand, informativeness remains nearly constant on regularly-varying tails. Hence, increasing m does not make the frequency p_θ^i correspond to "stronger evidence"; as a result, $R_{\theta,\theta',\infty}^j \left(p_\theta^i \right)$ remains around $(\hat{x}/\hat{y})^{-\alpha}$, bounded away from 0. Even in the limit $m \to \infty$, j assigns a significant probability to the other state, leading to a significant asymptotic disagreement.

To illustrate the nature of asymptotic disagreement under regularly-varying tails further, consider the Pareto distribution with some $\alpha > 0$. This is particularly relevant, since all distributions with regularly-varying tails behave similar to the Pareto-distribution for large m. For simplicity, suppose $\pi_A^1 = \pi_A^2 = 1/2$, so that there is no initial disagreement. The likelihood ratio is

$$R_{\theta,\theta',m}^i \left(\rho\left(s\right) \right) = \left(\frac{\rho_\theta\left(s\right) + \hat{p}^i - 1}{\rho_\theta\left(s\right) - \hat{p}^i} \right)^{-\alpha},$$

and the asymptotic probability that the true state is θ is

$$\phi_{\theta,\infty,m}^i \left(\rho\left(s\right) \right) = \frac{\left(\rho_\theta\left(s\right) - \hat{p}^i \right)^{-\alpha}}{\left(\rho_\theta\left(s\right) - \hat{p}^i \right)^{-\alpha} + \left(\rho_\theta\left(s\right) + \hat{p}^i - 1 \right)^{-\alpha}}$$

for all m. As illustrated in Figure 2, in this case $\phi_{\theta,\infty,m}^i$ is not monotone. To see the magnitude of asymptotic disagreement, consider $\rho_\theta\left(s\right) \cong \hat{p}^i$. In that case, $\phi_{\theta,\infty,m}^i \left(\rho\left(s\right) \right)$ is approximately
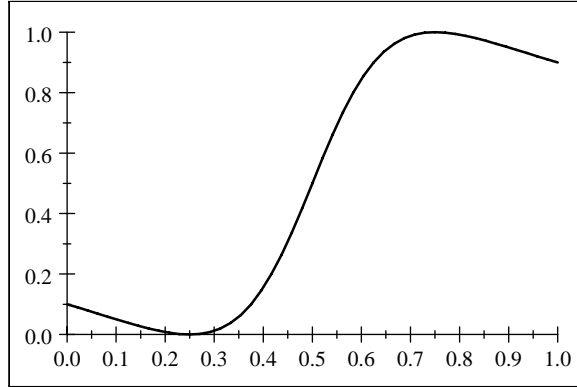
28

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

1, and $\phi^j_{\theta,\infty,m}(\rho(s))$ is approximately $\hat{y}^{-\alpha}/(\hat{x}^{-\alpha} + \hat{y}^{-\alpha})$. Hence, both individuals believe that the difference between their asymptotic posteriors will be

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

This asymptotic difference is increasing with the difference $\hat{y} \equiv \left| \hat{p}^1 - \hat{p}^2 \right|$, 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 $\hat{p}^1 \to \hat{p}^2$).[21]

5.4 Robustness of Agreement in the Medium Run

We have so far established that asymptotic agreement in standard models is fragile in the sense that there may be substantial asymptotic disagreement when one introduces a small amount of uncertainty. In the latter model, the amount of disagreement eventually exceeds a predetermined level, casting doubt on the notion that beliefs of different individuals become eventually similar as they observe the same public information. Nevertheless, it is possible that even in the model with divergent asymptotic beliefs, the individuals' beliefs may be quite similar for a long while before they eventually diverge. Then, one may be able to use the common-prior assumption as an approximation in the medium run, after sufficient amount of learning, but before beliefs eventually diverge. We will now show that this is indeed the case whenever there is asymptotic agreement in the limiting standard model. In other words,

---

[21]The working paper paper version shows that in this canonical example asymptotic disagreement will be small for large m whenever $\hat{y}$ is small.

medium-run agreement is continuous at certainty. (Below $\mathbb{E}_m^i$ is the expectation operator under $\Pr_m^i$, and $\delta_{p_\theta^i}$ is the Dirac measure that puts probability 1 on p_θ^i.)

Theorem 7 *(Continuity of Medium-Run Agreement) Let $\left(p^1, p^2\right) \in \left(\Delta\left(\Sigma\right)^\Theta\right)^2$ be such that there is asymptotic agreement under $(\delta_{p_\theta^i})_{\theta \in \Theta, i \in \{1,2\}}$ (e.g. $p^1 = p^2$ and $p_\theta^i \neq p_{\theta'}^i$ for all distinct θ and θ'). Let $\{F_{\theta,m}^i\}$ be any family converging to $\delta_{p_\theta^i}$. Then, for every $\varepsilon > 0$ and $N < \infty$, there exist $\bar{n} < \infty$ and $\bar{m} < \infty$ such that*

$$\Pr_m^i \left(\left\|\phi_{n,m}^1 - \phi_{n,m}^2\right\| > \varepsilon\right) < \varepsilon \qquad (\forall m > \bar{m}, \forall n \in \{\bar{n}, \bar{n}+1, \ldots, \bar{n}+N\}).$$

Equivalently, for every $\varepsilon > 0$ and $N < \infty$, there exist $\bar{n} < \infty$ and $\bar{m} < \infty$ such that

$$\mathbb{E}_m^i \left[\left\|\phi_{n,m}^1 - \phi_{n,m}^2\right\|\right] < \varepsilon \qquad (\forall m > \bar{m}, \forall n \in \{\bar{n}, \bar{n}+1, \ldots, \bar{n}+N\}).$$

Imagine two individuals who face a small amount of uncertainty about the conditional signal distributions, but their beliefs are similar to those in a standard model. A significant amount of belief differences may remain early in the process because it takes time to reduce the initial belief differences. Their beliefs may also eventually diverge as in the Binary Example. Despite this, Theorem 7 establishes that, in between these two ends, their beliefs will remain arbitrarily close to each other for an arbitrarily long period of time, provided that the amount of uncertainty is sufficiently small.

The idea of the proof is simple. Firstly, in the standard model, the expected disagreement between the individuals' beliefs vanishes as $n \to \infty$, so that it becomes less than $\varepsilon/2$ when n exceeds some $\bar{n}$. Moreover, for a fixed n, since the asymptotic beliefs are continuous and bounded, as $m \to \infty$, the expected disagreement in the model with uncertainty converges to the one under the standard model, so that it becomes within $\varepsilon/2$ neighborhood of the latter difference when m exceeds some $\bar{m}_n$. Hence, the expected disagreement (in both models) will be less than ε whenever $m \geq \bar{m}_n$ and $n \geq \bar{n}$. Therefore, when $m \geq \max_{\bar{n} \leq n \leq \bar{n}+N} \bar{m}_n$, the expected amount of disagreement will be less than ε for each n in between $\bar{n}$ and $\bar{n}+N$.

As n grows, in the standard model, the expected disagreement diminishes and vanishes eventually. In a nearby model with small uncertainty, the expected difference also diminishes and becomes very small for a long while (because it remains close to the one in the standard model), but it may eventually grow larger and become substantial. In a sense, small differences between the two models build up and lead to a substantial difference eventually as the individuals receive more and more information.
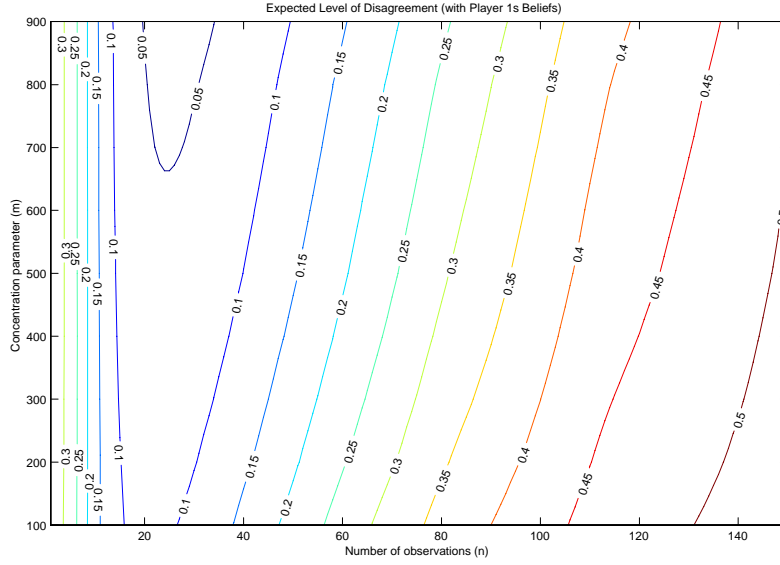
Figure 3: Ex-ante expected disagreement in the Binary Example of Section 3.2 for $\epsilon = \lambda = 1/m$.

For an illustration, consider the Binary Example of Section 3.2 for $\epsilon = \lambda = 1/m$. Take $\pi_A^1 = 2/3$ and $\pi_A^2 = 1/3$ as the prior beliefs. In Figure 3, we plot the ex-ante expected value of the disagreement aftern n observations as a function of m and n. (The value on the contour indicates the expected disagreement along the contour.) For any fixed m, as the individuals observe more signals, the expected disagreement decreases first and starts increasing after a while, eventually approaching $5/9$. As m increases (and uncertainty decreases), it takes longer for the asymptotic effect to take over, allowing the expected disagreement to decrease to lower values and stay low longer before increasing. Theorem 7 shows that this period becomes arbitrarily long as $m \to \infty$. Remarkably, however, even when m is around 1000, so that the individuals know probabilities upto the third digits, the expected disagreement starts growing rapidly only after 30 observations or so.

6 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. It receives support from a number of well-known theorems in statistics, such as Savage (1954) and Blackwell and Dubins (1962).

Nevertheless, existing results assume that conditional distribution (i.e., "interpretations") of signals are known. This is sufficient to ensure that payoff-relevant parameters (states) can be *identified* from limiting frequencies of signals. In many situations, individuals are not only learning about payoff-relevant parameters but also about the interpretation of different signals—i.e., learning would be taking 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. However, learning is under uncertainty in the sense that 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 a small amount of uncertainty may lead to a substantial (non-vanishing) amount of asymptotic disagreement, namely asymptotic agreement is discontinuous at certainty. We first show that asymptotic agreement is discontinuous (and thus fragile) 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 there will be a significant amount of asymptotic disagreement. Under an additional uniform convergence assumption, we also characterize the conditions under which asymptotic agreement is continuous at certainty. According to our characterization, asymptotic disagreement may prevail even as the amount of uncertainty vanishes, depending on the tail properties of the families of subjective probability distributions. These results imply learning foundations of common priors are not as strong as generally presumed.

32

Appendix: Proofs

Proof of Lemma 1. Write

$$\frac{\Pr^i\left(r_n|\theta'\right)}{\Pr^i\left(r_n|\theta\right)} = \frac{\int \prod_\sigma p_{\theta',\sigma}^{r_{\sigma,n}} f_{\theta'}^i(p_{\theta'})dp_{\theta'}}{\int \prod_\sigma p_{\theta,\sigma}^{r_{\sigma,n}} f_{\theta}^i(p_{\theta})dp_{\theta}} = \frac{\dfrac{\int \prod_\sigma p_{\theta',\sigma}^{r_{\sigma,n}} f_{\theta'}^i(p_{\theta'})dp_{\theta'}}{\int \prod_\sigma p_{\theta',\sigma}^{r_{\sigma,n}} dp_{\theta'}}}{\dfrac{\int \prod_\sigma p_{\theta,\sigma}^{r_{\sigma,n}} f_{\theta}^i(p_{\theta})dp_{\theta}}{\int \prod_\sigma p_{\theta,\sigma}^{r_{\sigma,n}} dp_{\theta}}} = \frac{\mathbb{E}^\lambda[f_{\theta'}^i|r_n]}{\mathbb{E}^\lambda[f_{\theta}^i|r_n]}.$$

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 $f_{\theta'}^i(p_{\theta'})$ given r_n under the flat (Lebesgue) prior on $p_{\theta'}$ and the Bernoulli distribution on $\{s_t\}_{t=1}^n$. Denoting this by $\mathbb{E}^\lambda[f_{\theta'}^i|r_n]$, and the denominator, which is similarly defined as the conditional expectation of $f_\theta(p_\theta)$, by $\mathbb{E}^\lambda[f_\theta^i|r_n]$, we obtain the last equality. By Doob's consistency theorem for Bayesian posterior expectation of the parameter, as $r_n \to \rho$ (that is, as $r_{\sigma,n} \to \rho_\sigma$ for each $\sigma \in \Sigma$), we have that $\mathbb{E}^\lambda[f_{\theta'}^i|r_n] \to f_{\theta'}(\rho)$ and $\mathbb{E}^\lambda[f_\theta^i|r_n] \to f_\theta^i(\rho)$ (Doob (1949) shows the convergence for almost all ρ, and Diaconis and Freedman shows the convergence for all ρ in our multinomial model). This establishes

$$\frac{\Pr^i\left(r_n|\theta'\right)}{\Pr^i\left(r_n|\theta\right)} \to R_{\theta,\theta'}^i\left(\rho\right),$$

as defined in (5). Equation (4) then follows from (3). ∎

Proof of Theorem 1. Throughout the proof, Assumptions 2 and 3 are imposed. We first develop the necessary notation.

Notation: For any i, write

$$U^i = \left\{\rho|\#P^i(\rho) = 1\right\}. \tag{13}$$

For every $\rho \in U^i$, write $p^i(\rho)$ for the unique member of $P^i(\rho)$ and $\theta^i(\rho)$ for the unique θ with $p^i(\rho) \in \text{supp}(F_\theta^i)$. Writing $r_n(s) = (r_{\sigma,n}(s))_{\sigma\in\Sigma}$, note that

$$\phi_{\theta,n}^i(s) = \frac{1}{1 + \sum_{\theta'\neq\theta} \frac{\pi_{\theta'}^i}{\pi_\theta^i} \bar{R}_{\theta,\theta',n}^i(r_n(s))} \tag{14}$$

where

$$\bar{R}_{\theta,\theta',n}^i(r_n(s)) = \frac{\int \prod_\sigma p_\sigma^{r_{\sigma,n}(s)} dF_{\theta'}^i(p)}{\int \prod_\sigma p_\sigma^{r_{\sigma,n}(s)} dF_\theta^i(p)} = \frac{\int e^{-H(r_n(s)/n,p)n} dF_{\theta'}^i(p)}{\int e^{-H(r_n(s)/n,p)n} dF_\theta^i(p)}. \tag{15}$$

We will use following lemmas in our proof.

Lemma 2 *Under Assumptions 2 and 3, for any i, as $n \to \infty$, if $r_n(s)/n \to \rho \in U^i$, then for all $\theta' \neq \theta^i(\rho)$,*

$$\bar{R}_{\theta^i(\rho),\theta',n}^i(r_n(s)) \to 0.$$

Proof of Lemma: By definition, $H\left(\rho, p^i(\rho)\right) < H(\rho, p)$ for all $p \in \text{supp}(F_{\theta'}^i)$. Since H is continuous and $\text{supp}(F_{\theta'}^i)$ is closed, this implies that there exist open neighborhoods V_p and V_ρ of $p^i(\rho)$ and ρ, respectively, such that

$$\hat{H} \equiv \sup_{\rho'\in V_\rho, p\in V_p} H(\rho', p) < \inf_{\rho'\in V_\rho, p\in supp(F_{\theta'}^i)} H(\rho', p) \equiv \hat{H}'.$$

33

Since $r_n(s)/n \to \rho$, there exists $\bar{n} < \infty$ such that $r_n(s)/n \in V_\rho$ for all $n > \bar{n}$. Take any $n > \bar{n}$. Then, by definition of $\hat{H}'$,

$$\int e^{-H(r_n(s)/n,p)n} dF^i_{\theta'}(p) \leq e^{-\hat{H}'n}. \tag{16}$$

Moreover,

$$\int e^{-H(r_n(s)/n,p)n} dF^i_{\theta^i(\rho)}(p) \geq \int_{p \in V_p} e^{-H(r_n(s)/n,p)n} dF^i_{\theta^i(\rho)}(p) \geq e^{-\hat{H}n} \Pr^i\left(V_p|\theta^i(\rho)\right), \tag{17}$$

where the first equality is by non-negativity of exponential function and the second inequality is by definition of $\hat{H}$. Note that since $p^i(\rho) \in \text{supp}(F^i_{\theta^i(\rho)})$ and V_p is an open neighborhood of $p^i(\rho)$, by definition of supp, $\Pr^i\left(V_p|\theta^i(\rho)\right) > 0$. Substituting (16) and (17) in definition (15), we then obtain

$$\bar{R}^i_{\theta^i(\rho),\theta',n}(r_n(s)) \leq \frac{e^{-\hat{H}'n}}{e^{-\hat{H}n}\Pr^i\left(V_p|\theta^i(\rho)\right)} = \frac{e^{-(\hat{H}'-\hat{H})n}}{\Pr^i\left(V_p|\theta^i(\rho)\right)}.$$

Since $\hat{H}' > \hat{H}$, the right-hand side goes to zero, showing that $R^i_{\theta^i(\rho),\theta',n}(r_n(s)) \to 0$. $\qquad$ *QED*

Lemma 3 *For any i and $s \in \bar{S}$ with $\rho(s) \in U^i$, as $n \to \infty$, $\phi^i_{\theta^i(\rho(s)),n}(s) \to 1$.*

Proof of Lemma: This lemma follows from Lemma 2 and (14). $\qquad$ *QED*

We are now ready to prove the theorem.

(Proof of Part 1) Take any θ. Firstly, by Doob's consistency theorem,

$$\Pr^i\left(r_n(s)/n \to \rho(s) \in \text{supp}\left(F^i_\theta\right)|\theta\right) = 1. \tag{18}$$

Moreover, for any $\rho \in \text{supp}(F^i_\theta)$, $P^i(\rho) = \{\rho\}$, yielding $\rho \in U^i$ with $p^i(\rho) = \rho$ and $\theta^i(\rho) = \theta$. Then, Lemma 3 establishes that
$$\Pr^i\left(\phi^i_{\theta,n}(s) \to 1|\theta\right) = 1.$$

(Proof of Part 2: Sufficiency) Take any $\hat{\theta}$. By Assumption 3, $\text{supp}(F^i_{\hat{\theta}}) \subseteq U^j$. Assume that for every $\rho \in \text{supp}(F^i_{\hat{\theta}}) \subseteq U^j$, $P^j(\rho) \subseteq \text{supp}(F^j_{\hat{\theta}})$, so that $\theta^j(\rho) = \hat{\theta}$. Since $\rho \in \text{supp}(F^i_{\hat{\theta}})$, $\theta^i(\rho) = \hat{\theta}$ as in part 1. Then, whenever $r_n(s)/n \to \rho(s) \in \text{supp}(F^i_{\hat{\theta}})$, by Lemma 3, $\phi^i_{\hat{\theta},n}(s) \to 1$ and $\phi^j_{\hat{\theta},n}(s) \to 1$. Consequently, $\left|\phi^1_{\theta,n}(s) - \phi^2_{\theta,n}(s)\right| \to 0$ for each $\theta \in \Theta$. Therefore, by (18), for each $\theta \in \Theta$,

$$\Pr^i\left(\left|\phi^1_{\theta,n}(s) - \phi^2_{\theta,n}(s)\right| \to 0|\hat{\theta}\right) = 1.$$

Since $\hat{\theta}$ is arbitrary, this shows that, for all θ,

$$\Pr^i\left(\left|\phi^1_{\theta,n}(s) - \phi^2_{\theta,n}(s)\right| \to 0\right) = 1.$$

(Proof of Part 2: Necessity) Suppose that for some $\hat{\theta}$ and $\hat{\rho} \in \text{supp}(F^i_{\hat{\theta}})$, $P^j(\hat{\rho}) \not\subseteq \text{supp}(F^j_{\hat{\theta}})$, so that $\theta^j(\hat{\rho}) \neq \hat{\theta} = \theta^i(\hat{\rho})$. Recall that $\hat{\rho} \in U^i$ and $\hat{\rho} \in U^j$ (by Assumption 3), with well-defined $\theta^i(\hat{\rho})$ and $\theta^j(\hat{\rho})$. Now, since H is continuous, P^j is upper-semicontinuous. Hence, there exists an open neighborhood $\hat{V} \subset U^j$ of $\hat{\rho}$ such that $\theta^j(\rho) = \theta^j(\hat{\rho}) \neq \hat{\theta}$ for each $\rho \in \hat{V}$. Thus, for any $s \in \bar{S}$ with $\rho(s) \in \hat{V} \cap \text{supp}(F^i_{\hat{\theta}})$, Lemma 3 implies that $\phi^i_{\hat{\theta},n}(s) \to 1$ and $\phi^j_{\theta^j(\hat{\rho}),n}(s) \to 1$, so that $\left|\phi^1_{\hat{\theta},n}(s) - \phi^2_{\hat{\theta},n}(s)\right| \to 1$. Since $\hat{V}$ is an open neighborhood of $\hat{\rho} \in \text{supp}(F^i_{\hat{\theta}})$, $\Pr^i\left(\rho(s) \in \hat{V} \cap \text{supp}(F^i_{\hat{\theta}})|\hat{\theta}\right) > 0$, showing that

$$\Pr^i\left(\left|\phi^1_{\hat{\theta},n}(s) - \phi^2_{\hat{\theta},n}(s)\right| \to 1\right) > 0,$$

and thus contradicting asymptotic agreement. ∎

Proof of Theorem 2. **(Proof of Part 1)** This part immediately follows from Lemma 1, as each $\pi^i_{\theta'} f_{\theta'}(\rho(s))$ is positive, and $\pi^i_\theta f_\theta(\rho(s))$ is finite.

(Proof of Part 2) By Lemma 1, $\phi^1_{\theta,\infty}(\rho) - \phi^2_{\theta,\infty}(\rho) = 0$ if and only if $\sum_{\theta' \neq \theta} R^1_{\theta,\theta'}(\rho(s)) \pi^1_{\theta'}/\pi^1_\theta = \sum_{\theta' \neq \theta} R^2_{\theta,\theta'}(\rho(s)) \pi^2_{\theta'}/\pi^2_\theta$. There is an open and dense set of parameters for which the equality on the right-hand side is satisfied only at a set of frequencies $\rho(s)$ with Lebesgue measure zero. Since we have densities, in any such case, $\Pr^i$ also puts probability zero on those frequencies, yielding $\Pr^i\left(\left|\phi^1_\infty(s) - \phi^2_\infty(s)\right| \neq 0\right) = 1$ for $i = 1, 2$. ∎

Proof of Theorem 3. Fix any $(\varepsilon, \theta, \theta', i)$ with $\varepsilon > 0$ and $\theta \neq \theta'$. We will show that $\lim_{m \to \infty} \Pr^i_m\left(R^i_{\theta,\theta',m} \geq \varepsilon|\theta\right) = 0$. Let V be a neighborhood of p^i_θ such that $p^i_{\theta'} \notin \bar{V}$, where $\bar{V}$ is the closure of V. Define

$$D_m = \left\{\rho \in V \mid R^i_{\theta,\theta',m}(\rho) \geq \varepsilon\right\}.$$

By definition,

$$\Pr^i_m\left(R^i_{\theta,\theta',m} \geq \varepsilon|\theta\right) \leq \Pr^i_m(D_m|\theta) + 1 - \Pr^i_m(V|\theta). \tag{19}$$

But

$$\begin{aligned} \Pr^i_m(D_m|\theta) &= \int_{\rho \in D_m} f^i_{\theta,m}(\rho) \, d\rho \\ &\leq \frac{1}{\varepsilon} \int_{\rho \in D_m} f^i_{\theta',m}(\rho) \, d\rho = \frac{1}{\varepsilon} \Pr^i_m(D_m|\theta') \\ &\leq \frac{1}{\varepsilon} \Pr^i_m(V|\theta') \end{aligned} \tag{20}$$

where the first inequality follows from the fact that $R^i_{\theta,\theta',m} = f^i_{\theta'}/f^i_\theta \geq \varepsilon$ on D_m and the second inequality holds since $D_m \subseteq V$. Combining (19) and (20), we obtain

$$0 \leq \Pr^i_m\left(R^i_{\theta,\theta',m} \geq \varepsilon|\theta\right) \leq \frac{1}{\varepsilon} \Pr^i_m(V|\theta') + 1 - \Pr^i_m(V|\theta).$$

Now, since $F^i_{\theta'} \to \delta_{p^i_{\theta'}}$ and $p^i_{\theta'} \notin \bar{V}$, $\Pr^i_m(V|\theta') \to 0$. Likewise, since $F^i_\theta \to \delta_{p^i_\theta}$ and $p^i_\theta \in V$, $\Pr^i_m(V|\theta) \to 1$. Therefore, the upper bound goes to 0, completing the proof. ∎

Proof of Theorem 4. Pick sequences $p^i_{\theta,m}$ and $\bar{\epsilon} > 0$ such that $p^i_{\theta,m} \to p^i_\theta$ and $\left\|p^i_{\theta,m} - p^j_{\theta',m}\right\| > \bar{\epsilon}/m$ for all $(\theta, i) \neq (\theta', j)$. For each (θ, i), define

$$D^i_{\theta,m} \equiv \left\{p \in \Delta(\Sigma) : 3\left\|p - p^i_{\theta,m}\right\| \leq \bar{\epsilon}/m\right\},$$

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

$$f^i_{\theta,m}(\rho) \equiv \begin{cases} x^i_{\theta,m} & \text{if } \rho \in D^i_{\theta,m} \\ 1/m & \text{otherwise,} \end{cases}$$

where $x^i_{\theta,m}$ is normalized so that $f^i_{\theta,m}$ is a probability density function. By construction of sequences $f^i_{\theta,m}$ and $p^i_{\theta,m}$, $F^i_{\theta,m} \to \delta_{p^i_\theta}$ for each (θ, i). We will show that agreement is strongly discontinuous under $\{F^i_{\theta,m}\}$. Now

$$\phi^i_{\theta,\infty,m}(\rho) = \frac{1}{1 + \frac{1-\pi^i_\theta}{\pi^i_\theta m x^i_{\theta,m}}}$$

35

if $\rho \in D^i_{\theta,m}$ for some θ and $\phi^i_{\infty,m}(\rho) = \pi^i$ otherwise. Note that $\phi^i_{\theta,\infty,m}(\rho) \to 1$ if $\rho \in D^i_{\theta,m}$. Moreover, since the sets $D^i_{\theta,m}$ and $D^j_{\theta',m}$ are disjoint for each θ', $\phi^j_{\infty,m}(\rho) = \pi^j$ when $\rho \in D^i_{\theta,m}$. Hence, there exist $\bar{m}$ such that for any $m \geq \bar{m}$ and any $\rho \in D^i_m \equiv \cup_\theta D^{i,m}_\theta$,

$$\left\| \phi^i_{\infty,m}(\rho) - \phi^j_{\infty,m}(\rho) \right\| > \varepsilon$$

where $\varepsilon \equiv \min_{j,\theta} \left(1 - \pi^j_\theta \right)/2$. But for each θ, $\Pr^i_m \left(D^i_{\theta,m} | \theta \right) \geq 1 - 1/m$, showing that $\Pr^i_m \left(D^i_m \right) \geq 1 - 1/m$. Therefore,

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

For the second part of the theorem, take $\pi^1 \neq \pi^2$. Then, by construction, for each ρ, $\left\| \phi^i_{\infty,m}(\rho) - \phi^j_{\infty,m}(\rho) \right\| > \min \left\{ \varepsilon, \left\| \pi^1 - \pi^2 \right\| \right\} > 0$, showing that agreement almost-surely discontinuous under $\{F^i_{\theta,m}\}$. ∎

Proof of Theorem 5. For each $m \gg 1$, let

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

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_θ is close to 1 for large m. Then,

$$R^i_{A,B,m}(\rho) = \begin{cases} 1/\varepsilon^2 & \text{if } \rho_\alpha < 1 - \hat{p}^i_B - \lambda/2, \\ x_B/\varepsilon^2 & \text{if } 1 - \hat{p}^i_B - \lambda/2 \leq \rho_a \leq 1 - \hat{p}^i_B + \lambda/2, \\ 1 & \text{if } 1 - \hat{p}^i_B + \lambda/2 < \rho_a < \hat{p}^i_A - \lambda/2, \\ \varepsilon^2/x_A & \text{if } \hat{p}^i_A - \lambda/2 \leq \rho_a \leq \hat{p}^i_A + \lambda/2, \\ \varepsilon^2 & \text{if } \rho_a > \hat{p}^i_A + \lambda/2, \end{cases}$$

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

$$R^i_{A,B,m}(\rho) \cong \begin{cases} \infty & \text{if } \rho_a \leq 1 - \hat{p}^i_B + \lambda/2, \\ 1 & \text{if } 1 - \hat{p}^i_B + \lambda/2 < \rho_a < \hat{p}^i_A - \lambda/2, \\ 0 & \text{if } \rho_a \geq \hat{p}^i_A - \lambda/2, \end{cases}$$

and hence

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

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

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

and

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

completing the proof. ∎

Proof of Theorem 6. **(Part 1: Sufficiency)** Fix any $\hat{\theta}$, and assume that $R^j_{\hat{\theta},\theta,\infty}\left(p^i_{\hat{\theta}}\right) = 0$ for each $\theta \neq \hat{\theta}$. We will show that for every $\varepsilon > 0$,

$$\lim_{m \to \infty} \mathrm{Pr}^i \left(\phi^1_{\hat{\theta},\infty,m} > 1 - \varepsilon, \phi^2_{\hat{\theta},\infty,m} > 1 - \varepsilon | \hat{\theta} \right) = 1, \tag{21}$$

which implies that, for each θ,

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

Since $\hat{\theta}$ is arbitrary, this yields the desired inequality for each θ:

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

To prove (21), it suffices to prove that for every $\varepsilon > 0$, and every i, j and every θ,

$$\lim_{m \to \infty} \mathrm{Pr}^i_m \left(R^j_{\hat{\theta},\theta,m} < \varepsilon | \hat{\theta} \right) = 1. \tag{22}$$

Since Proposition 3 has established (22) for $j = i$ already, we only need to prove (22) for $j \neq i$. Since $R^j_{\hat{\theta},\theta,m}$ converges uniformly to $R^j_{\hat{\theta},\theta,\infty}$ and each $R^j_{\hat{\theta},\theta,m}$ is continuous, $R^j_{\hat{\theta},\theta,\infty}$ is continuous at $p^i_{\hat{\theta}}$. Hence, there exists an open neighborhood $\hat{V} \subset V^i_{\hat{\theta}}$ of $p^i_{\hat{\theta}}$ such that $R^j_{\hat{\theta},\theta,\infty}(\rho) < \varepsilon/2$ for each $\rho \in \hat{V}$. Since $R^j_{\hat{\theta},\theta,m}$ converges uniformly to $R^j_{\hat{\theta},\theta,\infty}$ over $\hat{V}$, this implies that there exists $\bar{m} < \infty$ such that $R^j_{\hat{\theta},\theta,m}(\rho) < \varepsilon$ for each $\rho \in \hat{V}$ and $m > \bar{m}$. But since $F^i_{\hat{\theta},m}$ converges to $\delta_{p^i_{\hat{\theta}}}$ and $\hat{V}$ is an open neighborhood of $p^i_{\hat{\theta}}$, $\mathrm{Pr}^i_m \left(\hat{V} | \hat{\theta} \right) \to 1$, proving (22).

(Part 1: Necessity) Suppose that $R^j_{\theta,\theta',\infty}\left(p^i_\theta\right) \neq 0$ for some $i \neq j$ and $\theta \neq \theta'$. We will show that there exists $\varepsilon_0 > 0$ such that

$$\lim_{m \to \infty} \mathrm{Pr}^i_m \left(\left| \phi^1_{\theta,\infty,m} - \phi^2_{\theta,\infty,m} \right| > \varepsilon_0 | \theta \right) = 1. \tag{23}$$

This implies that

$$\lim_{m \to \infty} \mathrm{Pr}^i_m \left(\left| \phi^1_{\theta,\infty,m} - \phi^2_{\theta,\infty,m} \right| > \varepsilon_0 \right) \geq \pi^i_\theta > 0,$$

showing that agreement is discontinuous at certainty under $\{F^i_{\theta,m}\}$. To prove (23), we set

$$\varepsilon_0 \equiv \frac{1}{2} \left(1 - \frac{1}{1 + \frac{\pi^j_{\theta'}}{\pi^j_\theta} R^j_{\theta,\theta',\infty}\left(p^i_\theta\right)/2} \right) > 0.$$

Now, as in the sufficiency part, we use the uniform convergence of $R^j_{\theta,\theta',m}$ to $R^j_{\theta,\theta',\infty}$ and continuity of $R^j_{\theta,\theta',\infty}$ to conclude that there exist an open neighborhood $\hat{V} \subset V^i_\theta$ of p^i_θ and $\bar{m} < \infty$ such that $R^j_{\theta,\theta',m}(\rho) > R^j_{\theta,\theta',\infty}\left(p^i_\theta\right)/2$ for all $\rho \in \hat{V}$ and for all $m > \bar{m}$. But for any such m and ρ, $\phi^j_{\theta,\infty,m}(\rho) < 1 - 2\varepsilon_0$. Once again, $\mathrm{Pr}^i_m \left(\hat{V} | \theta \right) \to 1$, showing that

$$\lim_{m \to \infty} \mathrm{Pr}^i_m \left(\phi^j_{\theta,\infty,m}(\rho) < 1 - 2\varepsilon_0 | \theta \right) = 1. \tag{24}$$

On the other hand, by Proposition 3,

$$\lim_{m \to \infty} \mathrm{Pr}^i_m \left(\phi^i_{\theta,\infty,m}(\rho) > 1 - \varepsilon_0 | \theta \right) = 1. \tag{25}$$

37

When $\phi_{\theta,\infty,m}^j < 1 - 2\varepsilon_0$ and $\phi_{\theta,\infty,m}^i > 1 - \varepsilon_0$, we have $\left|\phi_{\theta,\infty,m}^1 - \phi_{\theta,\infty,m}^2\right| > \varepsilon_0$. Therefore, (24) and (25) imply (23), completing the proof of Part 1.

(Part 2) Assume that $R_{\theta,\theta',\infty}^j\left(p_\theta^i\right) \neq 0$ for all $i \neq j$ and $\theta \neq \theta'$. Then, by (23), there exists $\varepsilon_1 > 0$ such that $\mathrm{Pr}_m^i\left(\left|\phi_{\theta,\infty,m}^1 - \phi_{\theta,\infty,m}^2\right| > \varepsilon_1|\theta\right) \to 1$ for every θ. (To compute ε_1, replace $R_{\theta,\theta',\infty}^j\left(p_\theta^i\right)$ with $\min_{\theta,\theta'}\{R_{\theta,\theta',\infty}^j\left(p_\theta^i\right)\}$ in the definition of ε_0). Therefore,

$$\lim_{m\to\infty} \mathrm{Pr}_m^i\left(\left\|\phi_{\infty,m}^1 - \phi_{\infty,m}^2\right\| > \varepsilon_1\right) = 1,$$

showing that agreement is strongly discontinuous at certainty under $\left\{F_{\theta,m}^i\right\}$. ∎

Proof of Theoerm 7. It suffices to show that $\mathbb{E}_m^i\left[\left\|\phi_{n,m}^1 - \phi_{n,m}^2\right\|\right] < \varepsilon$ because $\mathrm{Pr}_m^i\left(\left\|\phi_{n,m}^1 - \phi_{n,m}^2\right\| > \varepsilon\right) < \varepsilon$ whenever $\mathbb{E}_m^i\left[\left\|\phi_{n,m}^1 - \phi_{n,m}^2\right\|\right] < \varepsilon^2$. In our proof we will use $\mathbb{E}_\infty^i$ and $\phi_{n,\infty}^i$ for the expectation operator and the posterior belief at n, respectively, under the standard model $(\delta_{p_\theta^i})_{\theta\in\Theta,i\in\{1,2\}}$. First, we observe that

$$\lim_{n\to\infty} \mathbb{E}_\infty^i\left[\left\|\phi_{n,\infty}^1 - \phi_{n,\infty}^2\right\|\right] = \mathbb{E}_\infty^i\left[\lim_{n\to\infty}\left\|\phi_{n,\infty}^1 - \phi_{n,\infty}^2\right\|\right] = 0, \tag{26}$$

where the first equality is by the Bounded Convergence Theorem, and the second equality is by the hypothesis of the theorem that there is asymptotic agreement under the standard model, i.e., $\lim_{n\to\infty}\left\|\phi_{n,\infty}^1 - \phi_{n,\infty}^2\right\| = 0$ almost surely.

Next, we introduce:

Claim 1 *For any fixed n, we have*

$$\lim_{m\to\infty} \mathbb{E}_m^i\left[\left\|\phi_{n,m}^1 - \phi_{n,m}^2\right\|\right] = \mathbb{E}_\infty^i\left[\left\|\phi_{n,\infty}^1 - \phi_{n,\infty}^2\right\|\right]. \tag{27}$$

Given this result, fix any $\varepsilon > 0$ and $N < \infty$. By (26), there exist $\bar{n} < \infty$ such that for all $n \geq \bar{n}$,

$$\mathbb{E}_\infty^i\left[\left\|\phi_{n,\infty}^1 - \phi_{n,\infty}^2\right\|\right] < \varepsilon/2.$$

On the other hand, for each $n \geq \bar{n}$, by (27), there exists $\bar{m}_n < \infty$ such that for all $m > \bar{m}_n$,

$$\mathbb{E}_m^i\left[\left\|\phi_{n,m}^1 - \phi_{n,m}^2\right\|\right] < \mathbb{E}_\infty^i\left[\left\|\phi_{n,\infty}^1 - \phi_{n,\infty}^2\right\|\right] + \varepsilon/2.$$

By picking $\bar{m} = \max_{\bar{n}\leq n\leq\bar{n}+N} \bar{m}_n$, we conclude that for each $m > \bar{m}$ and $n \in \{\bar{n}, \bar{n}+1, \ldots, \bar{n}+N\}$,

$$\mathbb{E}_m^i\left[\left\|\phi_{n,m}^1 - \phi_{n,m}^2\right\|\right] < \mathbb{E}_\infty^i\left[\left\|\phi_{n,\infty}^1 - \phi_{n,\infty}^2\right\|\right] + \varepsilon/2 < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

This establishes the desired result. The proof is completed by providing a proof for the Claim.

Proof of Claim: For any r_n, we write

$$\mathrm{Pr}_m^i\left(r_n\right) = \sum_\theta \pi_\theta^i \int \prod_\sigma p_{\theta,\sigma}^{r_{n,\sigma}} dF_{\theta,m}^i\left(p_\theta\right) > 0$$

for the probability of observing r_n. Since $\prod_\sigma p_{\theta,\sigma}^{r_{n,\sigma}}$ is a continuous function of p_θ and $F_{\theta,m}^i \to \delta_{p_\theta^i}$,

$$\lim_{m\to\infty} \mathrm{Pr}_m^i\left(r_n\right) = \sum_\theta \pi_\theta^i \prod_\sigma \left(p_{\theta,\sigma}^i\right)^{r_{n,\sigma}} \equiv \mathrm{Pr}_\infty^i\left(r_n\right) > 0, \tag{28}$$

where $\mathrm{Pr}_\infty^i\left(r_n\right)$ is the probability of r_n under the standard model. We also compute that

$$\begin{aligned}
\mathbb{E}_m^i\left[\left|\phi_{\theta,n,m}^1 - \phi_{\theta,n,m}^2\right|\right] &= \sum_{r_n} \mathrm{Pr}_m^i\left(r_n\right)\left|\frac{\pi_\theta^i \int \prod_\sigma p_{\theta,\sigma}^{r_{n,\sigma}} dF_{\theta,m}^i\left(p_\theta\right)}{\mathrm{Pr}_m^i\left(r_n\right)} - \frac{\pi_\theta^j \int \prod_\sigma p_{\theta,\sigma}^{r_{n,\sigma}} dF_{\theta,m}^j\left(p_\theta\right)}{\mathrm{Pr}_m^j\left(r_n\right)}\right| \\
&= \sum_{r_n} \left|\pi_\theta^i \int \prod_\sigma p_{\theta,\sigma}^{r_{n,\sigma}} dF_{\theta,m}^i\left(p_\theta\right) - \frac{\mathrm{Pr}_m^i\left(r_n\right)}{\mathrm{Pr}_m^j\left(r_n\right)}\pi_\theta^j \int \prod_\sigma p_{\theta,\sigma}^{r_{n,\sigma}} dF_{\theta,m}^j\left(p_\theta\right)\right|.
\end{aligned}$$

38

Now, as $m \to \infty$, $\int \prod_\sigma p_{\theta,\sigma}^{r_{n,\sigma}} dF_{\theta,m}^i (p_\theta) \to \prod_\sigma \left(p_{\theta,\sigma}^i \right)^{r_{n,\sigma}}$ (as above) and $\Pr_m^i (r_n) / \Pr_m^j (r_n) \to \Pr_\infty^i (r_n) / \Pr_\infty^j (r_n)$ by (28) and the Continuous Mapping Theorem. Therefore,

$$
\begin{aligned}
\lim_{m \to \infty} \mathbb{E}_m^i \left[\left| \phi_{\theta,n,m}^1 - \phi_{\theta,n,m}^2 \right| \right] &= \sum_{r_n} \left| \pi_\theta^i \prod_\sigma \left(p_{\theta,\sigma}^i \right)^{r_{n,\sigma}} - \frac{\Pr_\infty^i (r_n)}{\Pr_\infty^j (r_n)} \pi_\theta^j \prod_\sigma \left(p_{\theta,\sigma}^j \right)^{r_{n,\sigma}} \right| \\
&= \mathbb{E}_m^i \left[\left| \phi_{\theta,n,\infty}^1 - \phi_{\theta,n,\infty}^2 \right| \right] .
\end{aligned}
$$

Since $\max_\theta$ is continuous, one more application of the Continuous Mapping Theorem yields (27), proving the Claim. This completes the proof of the theorem. ∎

References

Acemoglu, Daron, Victor Chernozhukov, and Muhamet Yildiz (2006): "Learning and Disagreement in an Uncertain World", 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.

Burnham, Kenneth and David Anderson (1989): *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, Springer, New York.

Cripps, Martin, Jeffrey Ely, George Mailath and Larry Samuelson (2008): "Common Learning" *Econometrica*, 76, 909-933.

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.

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.

Harrison, Michael and David Kreps (1978): "Speculative Investor Behavior in a Stock Market with Heterogenous Expectations," *Quarterly Journal of Economics*, 92, 323-36.

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.

Morris, Stephen (1996): "Speculative Investor Behavior and Learning," *Quarterly Journal of Economics*, Vol. 111, No. 4, 1111-33.

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.