Fragility of asymptotic agreement under Bayesian learning

Daron Acemoglu
Department of Economics, Massachusetts Institute of Technology

Victor Chernozhukov
Department of Economics, Massachusetts Institute of Technology

Muhamet Yildiz
Department of Economics, Massachusetts Institute of Technology

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 (nonvanishing) 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 *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 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 nondegenerate belief about p_θ. This belief has a cumulative distribution function (c.d.f.) 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 $p_A > \frac{1}{2} > p_B$. 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.)

Asymptotic learning and asymptotic agreement in the standard model rely on the assumption that the supports of F_A^i and F_B^i are disjoint (which we refer to as the *full-identification assumption*). In that case, each individual can identify the state from the long-run frequency of signals (as $n \to \infty$) and eventually learn the true state (*asymptotic learning*). Under additional mild assumptions, they also 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.

In our main results, we investigate 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 $(p_A^1, p_B^1, p_A^2, p_B^2)$, where each F_θ^i puts probability 1 on a single frequency p_θ^i. We then consider families $\{F_{\theta,m}^i\}$ of distributions 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 shows that asymptotic agreement is never robust to all relaxations. More precisely, for every standard model (here parameterized by the vector $(p_A^1, p_B^1, p_A^2, p_B^2)$), 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 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 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) tails, such as the normal and the exponential distributions.

Our third main result shows that agreement is "continuous" in the medium run.[1] Consider any standard model with asymptotic agreement and any family $\{F^i_{\theta,m}\}$ of models converging to the standard model. For any model $F^i_{\theta,m}$, 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 $F^i_{\theta,m}$ 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 $F^i_{\theta,m}$, 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 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

---

[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 (1987, 1998) and Gul (1998). Gul (1998), for instance, questions whether the common-prior assumption makes sense when there is no ex ante stage.

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

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 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). For example, Freedman shows that when there are infinitely

---

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

[5]In fact, the full-identification assumption is considerably stronger than monotonicity and hence the full-identification assumption fails whenever asymptotic beliefs are non-monotone on some (possibly unlikely) events. Theorem 3 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).

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. In Freedman's example, the individual puts zero probability on the true signal distribution. 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.[6] This identification problem leads to different possible interpretations of the same signal sequence by individuals with different priors. In most economic situations, what is important is not future frequencies of signals but some payoff-relevant parameter. For example, what is relevant for economists trying to evaluate a policy is not the frequency of estimates on the effect of similar policies from other researchers, but the impact of this specific policy when (and if) implemented. Similarly, 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 et al. (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 *heterogeneous 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.[7]

In dynamic games, a similar mechanism precludes full learning: some subgames are never visited along the equilibrium path and thus players do not learn the other players'

---

[6]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 nonstationarity of the world.

[7]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.

and nature's moves in those subgames (see for example Fudenberg and Levine 1993 and Fudenberg and Kreps 1995). In that case, players may remain holding differing beliefs about the other players' and nature's moves in those contingencies even after arbitrarily long interactions. In particular, in games against nature, the players may disagree about the consequences in counterfactual situations, and some superstitions may persist (Fudenberg and Levine 2006). Note that, as in our paper, the main reason for lack of learning and asymptotic agreement is the lack of identification: one cannot identify the strategies from the observable moves. We are not aware of any work on nearly identifiable models within that literature.

In our next section, we introduce our general model. In Section 3, we illustrate the asymptotic agreement in the standard model and its fragility using a binary example (as briefly discussed above). Section 4 formally introduces our notion of discontinuity of asymptotic agreement and presents our main results. In Section 5, using an asset trading example, we illustrate how the failure of asymptotic agreement may have significant implications for game-theoretic and economic analysis. Section 6 concludes. In Appendix A, we present a general characterization of asymptotic learning and agreement under full identification. Appendix B contains the proofs of all the results stated in the paper.

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 $\| \cdot \|$. Individual 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.[8] That is, 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 $p_\theta \equiv (p_{\theta,\sigma})_{\sigma \in \Sigma} \in \Delta(\Sigma)$. The key notation introduced in this paragraph and later is summarized in Table 1.

Our main departure from the standard model is that we allow the individuals to be uncertain about the signal frequency 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.

---

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

Variable	Meaning
$\theta \in \Theta$	Unknown fundamental
$s = \{s_t\}$	A sequence of signals observed
$\sigma \in \Sigma$	A possible value for signal s_t
π_θ^i	Prior probability of θ according to i
$\phi_{\theta,n}(s)$	Posterior probability of θ after observing $\{s_1, \ldots, s_n\}$
$\phi_{\theta,\infty}(s)$	Asymptotic probability of θ
$r_{\sigma,n}(s)$	Empirical frequency of σ
$\rho_\sigma(s)$	Empirical long-run frequency of σ
$p_{\theta,\sigma}$	Unknown probability of σ under θ
F_θ^i, f_θ^i	Distribution (c.d.f. and p.d.f.) of $p_{\theta,\sigma}$ according to i
$R_{\theta,\theta'}^i = f_{\theta'}^i / f_\theta^i$	Likelihood ratio

TABLE 1. Key variables.

ASSUMPTION 1 (Full support for signal frequency). *For each i and θ, F_θ^i has a continuous, nonzero, 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.[9]

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

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

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

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

the number of times $s_t = \sigma$ out of first n signals for each signal value $\sigma \in \Sigma$; we write $r_n(s) \equiv (r_{\sigma,n}(s))_{\sigma \in \Sigma}$.[10] Let us write

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

---

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

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

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

at each event $E^{\theta,s,n} = \{(\theta, s') \mid s_t' = s_t \text{ for each } t \le n\}$, 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

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

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

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

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 $s \in \bar{S}$,*

$$\lim_{n \to \infty} \phi^i_{\theta,n}(s) = \frac{1}{1 + \sum_{\theta' \neq \theta} \frac{\pi^i_{\theta'}}{\pi^i_\theta} R^i_{\theta,\theta'}(\rho(s))} \equiv \phi^i_{\theta,\infty}(\rho(s)), \tag{2}$$

where

$$R^i_{\theta,\theta'} \equiv f^i_{\theta'}/f^i_\theta \tag{3}$$

is the asymptotic likelihood ratio for θ and θ'.

All proofs are given in the Appendix.

In (3), $R^i_{\theta,\theta'}(\rho)$ 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 $\phi^i_\infty \equiv (\phi_{\theta,\infty})_{\theta \in \Theta}$.

REMARK 1. In any exchangeable model, the long-run frequency vector $\rho(s)$ is a sufficient statistic for the entire sample path s. Hence, after observing the entire s, a Bayesian individual updates his belief to $\phi^i_{\theta,\infty}(\rho(s))$ by conditioning on $\rho(s)$ whenever $\rho(s)$ has positive ex ante probability (or positive density). In that case, asymptotic beliefs $\phi^i_{\theta,n}(s)$ also converge to $\phi^i_{\theta,\infty}(\rho(s))$, as established by Lemma 1. Hence, under the full-support assumption, asymptotic analysis in our model is equivalent to the analysis of a simpler

model in which the individuals observe a single lower-dimensional signal $\rho(s) \in \Delta(\Sigma)$. This equivalence leads to elementary analyses in most of our proofs.[11]

For us, the following two concepts are of interest.

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

(ii) *Asymptotic agreement*: $\Pr^i(\phi^1_\infty = \phi^2_\infty) = 1$ 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 also discuss the problematic assumptions of the standard models that motivate our alternative full-support assumption. We take $\Theta = \{A, B\}$ and $\Sigma = \{a, b\}$, so that there are two states and the signals are binary.

3.1 *Standard model without uncertainty and its problematic assumptions*

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

$$R(r_{a,n}, n) = \left(\frac{\hat{p}^{r_{a,n}/n}(1 - \hat{p})^{1-r_{a,n}/n}}{\hat{p}^{1-r_{a,n}/n}(1 - \hat{p})^{r_{a,n}/n}} \right)^n. \tag{4}$$

---

[11]In contrast, standard models preclude most long-run frequencies as zero-probability events. At those frequencies, the posterior belief $\phi^i_{\theta,\infty}(\rho(s))$ is not well defined, and the Bayes' rule is discontinuous. In that case, the asymptotic beliefs are instead determined by sampling variation, requiring a more involved analysis, which we discuss in the context of our illustrative binary example in Section 3.1 and more generally in Appendix A. Finally, we remark here that the observation that there is an asymptotic sufficient statistics applies to a wide variety of regular statistical models, for example, the models that are LAN (local asymptotically normal). The notion of asymptotic sufficiency as well as of asymptotic optimality based on it has been formalized by Le Cam's theory of limits of statistical experiments; see, e.g., Van der Vaart (1998) for a lucid treatment. Therefore, the insights developed here for the binary and multinomial models extend naturally to such broader contexts.

[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" or "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 about both 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.

Suppose that $n \to \infty$ and $r_{a,n}/n \to \rho_a$ for some $\rho_a > \frac{1}{2}$. Since ρ_a is more likely under A than under B, the expression in the parentheses in (4) is greater than 1, and $R(r_{a,n}, n) \to \infty$. Hence, asymptotically, each individual assigns probability 1 to state A. When the true state is A, $r_{a,n}/n \to \hat{p}$. 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 $1 - \hat{p} < \frac{1}{2}$, and since the expression in the parentheses is now less than 1, $R(r_{a,n}, n) \to 0$. In that case, both individuals assign probability 1 to the true state B, asymptotically. Except for the knife-edge case $\rho_a = \frac{1}{2}$, both individuals asymptotically assign probability 1 to the same state. Therefore, in the standard model, we have both asymptotic learning (i.e., $\Pr^i(\phi^i_{\theta,\infty} = 1|\theta) = 1$ for each θ and $i = 1, 2$) and asymptotic agreement (i.e., $\Pr^i(\phi^1_\infty = \phi^2_\infty) = 1$ for $i = 1, 2$).

A simple intuition for this result is that the underlying state θ is *fully identified* from the limiting frequencies (i.e., $\hat{p} \neq 1 - \hat{p}$), so that both individuals can infer the underlying state from the observation of the limiting frequencies of signals. However, there is more to this result than this simple intuition. Each individual is sure that, at state θ, they will be confronted with a limiting frequency of $\rho_\theta = \hat{p}$, in which case they will conclude that the true state is θ. They are certain that they will not observe an asymptotic frequency of ρ_a other than $\hat{p}$ and $1 - \hat{p}$.

> *What happens if an individual observes a frequency ρ_a of a-signals different from $\hat{p}$ and $1 - \hat{p}$ in a large sample of size n?*

The answer to this question highlights why asymptotic agreement under the standard model rests on problematic assumptions. Although the event in the question has zero probability under the individual's beliefs at the limit $n = \infty$, any frequency has strictly positive probability for $n < \infty$ because of sampling variation. In particular, the individual expects frequency ρ_a to occur with probabilities $(\hat{p}^{\rho_a}(1 - \hat{p})^{1-\rho_a})^n$ and $(\hat{p}^{1-\rho_a}(1 - \hat{p})^{\rho_a})^n$ under states A and B, respectively, when the sample size is n. When $\rho_a > \frac{1}{2}$, this event is infinitely more likely under $\theta = A$ than under $\theta = B$ in the sense that $(\hat{p}^{\rho_a}(1 - \hat{p})^{1-\rho_a})^n/(\hat{p}^{1-\rho_a}(1 - \hat{p})^{\rho_a})^n \to \infty$ as $n \to \infty$. Consequently, he becomes increasingly certain that the long-run frequency of a is $\hat{p}$, which is distinct from ρ_a, and assigns probability 1 to state $\theta = A$. Crucially, the individual reaches this inference despite the fact that as $n \to \infty$, frequency ρ_a has zero probability under $\theta = A$. The fact that identification relies on inference based on such zero-probability events is a problematic aspect of the standard model. What is more, individual i never questions the validity of his model even though the reality is increasingly inconsistent with this model (since $\hat{p} \neq \rho_a \neq 1 - \hat{p}$). 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. Finally, individual i's beliefs (at sample path s) concerning future frequencies diverge from the actual empirical frequency. In particular, we have

$$\lim_{n\to\infty} \Pr^i(s_{n+1} = a|\{s_t\}_{t=1}^{t=n}) \neq \rho_a(s)$$

because $\lim_{n\to\infty} \Pr^i(s_{n+1} = a|\{s_t\}_{t=1}^{t=n}) \in \{\hat{p}, 1 - \hat{p}\}$ while $\rho_a(s) \notin \{\hat{p}, 1 - \hat{p}\}$. In contrast, under the full-support assumption, we always have

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

so that individuals' beliefs about the future frequencies always converge to the empirical frequency.[14]

This discussion suggests that the support restrictions on limiting frequencies in the standard model lead 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 consequences. To investigate these issues systematically, we focus here on the case in which the full-support assumption applies.

Clearly, under the full-support assumption, we cannot have full asymptotic learning:

$$\Pr^i(\phi^i_{\theta,\infty} \neq 1 | \theta) = 1. \tag{5}$$

Moreover, by Lemma 1, except for some knife-edge cases,[15] the initial belief differences are reflected in the asymptotic beliefs, and the individuals fail to reach an asymptotic agreement almost surely:

$$\Pr^i(\phi^1_\infty \neq \phi^2_\infty) = 1. \tag{6}$$

Nevertheless, asymptotic learning and agreement under the standard model may be more robust than (5) and (6) appear to suggest. Under the standard model, individuals expect to assign exactly probability 1 to the true state and the asymptotic disagreement is exactly 0. When the idealized assumptions of the standard model are 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 asymptotic learning and agreement results are robust in the sense that the amount of asymptotic disagreement is small in the neighborhood of the standard models, e.g., when the amount of uncertainty about the conditional signal distributions is small. Our next example shows that this is not necessarily true: in certain cases, there will be substantial asymptotic disagreement even if the amount of uncertainty is negligible and the model looks like the standard one at first glance.

3.2 *Large deviations from asymptotic agreement 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 $1 - \epsilon$, $p_{A,a}$ and $p_{B,b}$ are in the λ neighborhood of some $\hat{p}^i$, but with

---

[14]This claim can be readily proved following the same steps as in the proof of Lemma 3 in Appendix B. Note also that, while the full-support assumption gets rid of the above discrepancy, the beliefs about the future events may still differ from the observed data. In particular, by exchangeability, individuals do not expect the dynamic patterns observed in the past to continue in the future.

[15]The exception is when the equality

$$\sum_{\theta' \neq \theta} \frac{\pi^1_{\theta'}}{\pi^1_\theta} R^1_{\theta,\theta'}(\rho(s)) = \sum_{\theta' \neq \theta} \frac{\pi^2_{\theta'}}{\pi^2_\theta} R^2_{\theta,\theta'}(\rho(s))$$
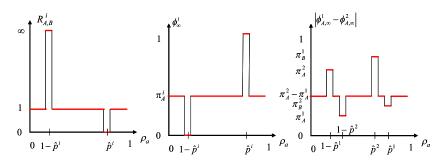
holds for some θ.

FIGURE 1. The three panels show, respectively, the approximate values of $R^i_{A,B}$, $\phi^i_{A,\infty}$, and $|\phi^1_{A,\infty} - \phi^2_{A,\infty}|$ as $\epsilon \to 0$.

probability ϵ, the signals are not informative. We will pay special attention to the limit case $\epsilon \to 0$, $\lambda \to 0$, and $\hat{p}^i \to \hat{p}$. In that case, each individual's beliefs are approximately the same as in the standard model discussed in the previous subsection.

Formally, for $\hat{p}^i > \frac{1}{2}(1 + \lambda)$ and $\lambda < |\hat{p}^1 - \hat{p}^2|$, we consider

$$f^i_\theta(p_\theta) = \begin{cases} \epsilon + \frac{1-\epsilon}{\lambda} & \text{if } p_{\theta,\theta} \in (\hat{p}^i - \frac{1}{2}\lambda, \hat{p}^i + \frac{1}{2}\lambda) \\ \epsilon & \text{otherwise} \end{cases}$$

for each θ and i. Here, $p_\theta = (p_{\theta,a}, p_{\theta,b})$, and $p_{\theta,\theta}$ is $p_{A,a}$ for $\theta = A$ and $p_{B,b}$ for $\theta = B$. Now, by (3), the asymptotic likelihood ratio is

$$R^i_{A,B}(\rho) = \begin{cases} \frac{\epsilon\lambda}{1-\epsilon(1-\lambda)} & \text{if } \rho_a \in D^i_A \equiv (\hat{p}^i - \frac{1}{2}\lambda, \hat{p}^i + \frac{1}{2}\lambda) \\ \frac{1-\epsilon(1-\lambda)}{\epsilon\lambda} & \text{if } \rho_a \in D^i_B \equiv (1 - \hat{p}^i - \frac{1}{2}\lambda, 1 - \hat{p}^i + \frac{1}{2}\lambda) \\ 1 & \text{otherwise.} \end{cases} \tag{7}$$

This and other relevant functions are plotted in Figure 1 for the limit case $\epsilon \to 0$, $\lambda \to 0$. The likelihood ratio $R^i_{A,B}(\rho)$ is 1 when ρ_a is small, takes a very high value at $1 - \hat{p}^i$, goes down to 1 afterward, becomes nearly zero around $\hat{p}^i$, and then jumps back to 1. By Lemma 1, $\phi^i_{A,\infty}(\rho)$ will also be non-monotone: when ρ_a is small, the signals are not informative; thus $\phi^i_{A,\infty}(\rho)$ is the same as the prior, π^i_A. In contrast, around $\rho_a = 1 - \hat{p}^i$, the signals become very informative, suggesting that the state is B; thus $\phi^i_{A,\infty}(\rho) \cong 0$. After this point, the signals become uninformative again and $\phi^i_{A,\infty}(\rho)$ goes back to π^i_A. Around $\hat{p}^i$, the signals are again informative, but this time favoring state A, so $\phi^i_{A,\infty}(\rho) \cong 1$. Finally, signals again become uninformative and $\phi^i_{A,\infty}(\rho)$ falls back to π^i_A.

Intuitively, when $\rho_a(s)$ is around $1 - \hat{p}^i$ or $\hat{p}^i$, 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 $\phi^i_{A,\infty}$ is equal to the prior for a large range of limiting frequencies, as $\epsilon \to 0$ and $\lambda \to 0$ each individual attaches probability 1 to the event that he will learn 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 $1 - \hat{p}^i$ or $\hat{p}^i$. Thus, there is approximate asymptotic learning in this environment.

However, asymptotic learning is considerably weaker than asymptotic agreement. Each individual also understands that since $\lambda < |\hat{p}^1 - \hat{p}^2|$, 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 "objective" bound

$$\tilde{z} \equiv \min\{\pi_A^1, \pi_A^2, \pi_B^1, \pi_B^2, |\pi_A^1 - \pi_A^2|\}. \tag{8}$$

When $\pi_A^1 = \frac{1}{3}$ and $\pi_A^2 = \frac{2}{3}$, this bound is equal to $\frac{1}{3}$. 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, $\Pr^i(|\phi_{A,\infty}^1(\rho) - \phi_{A,\infty}^2(\rho)| \geq Z) \geq 1 - \epsilon$, where as $\epsilon \to 0$,

$$Z \to z \equiv \min\{\pi_A^1, \pi_A^2, \pi_B^1, \pi_B^2\}. \tag{9}$$

This "subjective" bound can be as high as $\frac{1}{2}$.

Clearly, we can pick $\hat{p}^1 \cong \hat{p}^2 \cong \hat{p}$ and $\epsilon \cong \lambda \cong 0$, so that each individual's beliefs are 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.[16]

Note that the two models above differ mainly on how they treat the "small probability" events. The standard model in Section 3.1 reduces all such events to zero-probability events and uses the limits of likelihood ratio due to sampling variation under the reduced model to assign degenerate probabilities on states whenever such event occurs. In contrast, the model in Section 3.2 assigns small prior probability on those events and uses the likelihood ratio according to these prior beliefs to assign posterior probabilities on the states when such events occur. (Even though there are clearly more degrees of freedom in choosing small prior probabilities, our qualitative results are independent of these probabilities.) Our main concern in this paper is to investigate when the asymptotic agreement property in the standard model survives under this type of small prior probabilities.

4. Fragility of asymptotic agreement—Main results

In this section, we turn to our general model and 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: as in

---

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

the binary example, one can slightly perturb a standard model of learning under certainty 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.

We consider an arbitrary sequence $\{F^i_{\theta,m}\}_{m\in\mathbb{N}, i\in\{1,2\}, \theta\in\Theta}$ of subjective probability distributions about the signal frequency that converge in weak topology to a Dirac distribution $\delta_{p^i_\theta}$ for each (i, θ) as $m \to \infty$, where $\delta_{p^i_\theta}$ puts probability 1 on $p^i_\theta \in \Delta(\Sigma)$. (We will simply say that $\{F^i_{\theta,m}\}$ converges to $\delta_{p^i_\theta}$.) 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 the 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.

In the sequel, we focus on the case that there is asymptotic learning and asymptotic agreement under certainty for $\delta_{p^i_\theta}$ (as in Section 3.1).[17] As $m \to \infty$, uncertainty about the interpretation of the signals disappears and we converge to a model with asymptotic learning and asymptotic agreement. We check whether these asymptotic learning and asymptotic agreement properties are *continuous at certainty* in the sense that they hold approximately for large m. We first observe that asymptotic learning is indeed continuous at certainty.

THEOREM 1 (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),

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

Equivalently, for any $(\varepsilon, \theta, \theta', i)$ *with* $\varepsilon > 0$ *and* $\theta \neq \theta'$, $\lim_{m\to\infty} \mathrm{Pr}^i_m(R^i_{\theta,\theta'm} < \varepsilon \mid \theta) = 1$.

Here, the condition that $p^i_\theta \neq p^i_{\theta'}$ for all $\theta \neq \theta'$ ensures that there is asymptotic agreement under $\delta_{p^i_\theta}$. Theorem 1 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 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, asymptotic learning in the standard models is robust.

We next define the continuity of asymptotic agreement at certainty and show that, in contrast to asymptotic learning, asymptotic agreement is not robust. That is, there may remain a substantial amount of asymptotic disagreement along the way (for large m).

---

[17]We assume asymptotic learning under certainty only in Theorems 1 and 4, and assume asymptotic agreement under certainty only in Theorem 5, where the assumption is explicitly stated.

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

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

Continuity at certainty requires that as the family of subjective probability distributions converge to a Dirac distribution, each individual becomes increasingly certain that asymptotic disagreement will be arbitrarily small. Hence, asymptotic agreement is *discontinuous at certainty* under $\{F^i_{\theta,m}\}$ if there exists $\varepsilon > 0$ such that

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

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

DEFINITION 2. We say that *asymptotic agreement is strongly discontinuous at certainty under* $\{F^i_{\theta,m}\}$ if there exists $\varepsilon > 0$ such that for each $i = 1, 2$,

$$\lim_{m\to\infty} \text{Pr}^i_m(\|\phi^1_{\infty,m} - \phi^2_{\infty,m}\| > \varepsilon) = 1.$$

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* $\{F^i_{\theta,m}\}$ if there exist $\varepsilon > 0$ and $\bar{m} < \infty$ such that for all $m > \bar{m}$ and $\rho \in \Delta(\Sigma)$,

$$\|\phi^1_{\infty,m}(\rho) - \phi^2_{\infty,m}(\rho)\| > \varepsilon.$$

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.

REMARK 2. As we explain in Remark 1, $\phi^i_{\theta,\infty}(s)$ is the posterior probability individual i assigns to state θ after observing s. One could then equivalently define the continuity of (asymptotic) learning by the condition that

$$\text{Pr}^i_m\big(\{(\theta, s) : d(\phi^i_m(s), \delta_\theta) > \varepsilon\}\big) \to 0$$

for all $\varepsilon > 0$, where $\phi^i_m(s)$ is the posterior belief about θ after observing s under the probability distribution Pr^i_m about (θ, s), δ_θ is the Dirac distribution, and d is a metric on

probability distributions. The continuity of asymptotic learning according to this defi-
nition also follows readily from Theorem 1. The continuity of (asymptotic) agreement
requires quite a bit more than this, however:

$$\mathrm{Pr}^i_m\big(\{(\theta, s) : d(\phi^i_m(s), \phi^j_m(s)) > \varepsilon\}\big) \to 0$$

for all $\varepsilon > 0$. Under the continuity of (asymptotic) learning, it requires an additional
condition:

$$\mathrm{Pr}^i_m\big(\{(\theta, s) : d(\phi^j_m(s), \delta_\theta) > \varepsilon\}\big) \to 0$$

for all $\varepsilon > 0$. Thus it is also necessary that under nearby models, individuals should
be confident that other individuals will also learn. Such confidence in other individu-
als' learning is highly fragile, as we demonstrate next, underpinning our results on the
discontinuity of (asymptotic) agreement.

4.1 *Discontinuity of asymptotic agreement*

THEOREM 2 (Strong discontinuity). *For each $(p^1, p^2) \in \Delta(\Sigma)^{2K}$, there exists a family
$\{F^i_{\theta,m}\}$ converging to $\delta_{p^i_\theta}$ under which asymptotic agreement is strongly discontinuous at
certainty. If $\pi^1 \neq \pi^2$, then asymptotic agreement is almost-surely discontinuous under
$\{F^i_{\theta,m}\}$.*

That is, for every model with certainty, one can introduce a vanishingly small uncer-
tainty in such a way that the individuals' asymptotic beliefs remain substantially differ-
ent 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
$\epsilon = \lambda = \bar{\epsilon}/m$, $\hat{p}^1 = \hat{p} + \lambda$, and $\hat{p}^2 = \hat{p} - \lambda$, where $\bar{\epsilon}$ is such that $\phi^i_{B,\infty}(\rho) < \frac{1}{2}\pi^j_B$ for
$\rho_a \in D^i_A$ and $\phi^i_{A,\infty}(\rho) < \frac{1}{2}\pi^j_A$ for $\rho_a \in D^i_B$ whenever $\epsilon = \lambda \le \bar{\epsilon}$. (Recall from (7) that $D^i_A \equiv$
$(\hat{p}^i - \frac{1}{2}\lambda, \hat{p}^i + \frac{1}{2}\lambda)$ and $D^i_B \equiv (1 - \hat{p}^i - \frac{1}{2}\lambda, 1 - \hat{p}^i + \frac{1}{2}\lambda)$.) Such $\bar{\epsilon}$ exists (by asymptotic
learning of i). By construction, each $F^i_{\theta,m}$ converges to the beliefs in the standard model,
and $|\hat{p}^1 - \hat{p}^2| > \lambda$. To see strong discontinuity, consider the bound $\bar{Z} = \frac{1}{2}z > 0$ with z as
in (9). By the choice of $\bar{\epsilon}$, $|\phi^1_{A,\infty,m}(\rho) - \phi^2_{A,\infty,m}(\rho)| > \bar{Z}$ whenever $\rho_a \in D^i_A \cup D^i_B$. But
$\mathrm{Pr}^i_m(\rho_a(s) \in D^i_A \cup D^i_B) = 1 - \epsilon(1 - \lambda)$, which goes to 1 as $m \to \infty$. Therefore,

$$\lim_{m \to \infty} \mathrm{Pr}^i_m(|\phi^1_{A,\infty,m} - \phi^2_{A,\infty,m}| > \bar{Z}) = 1.$$

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

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

In the binary example (and in the proof of Theorem 2), 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). 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 3 (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 probability vectors $\hat{p}^i_A = (\hat{p}^i_{A,a}, 1 - \hat{p}^i_{A,a})$ and $\hat{p}^i_B = (1 - \hat{p}^i_{B,b}, \hat{p}^i_{B,b})$ with $\hat{p}^i_{A,a} > \frac{1}{2}$ and $\hat{p}^i_{B,b} > \frac{1}{2}$. There exist a family $\{F^i_{\theta,m}\}$ such that the following statements hold:*

(i) *The variable $F^i_{\theta,m}$ converges to $\delta_{\hat{p}^i_\theta}$ for each θ and i.*

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

(iii) *Agreement is discontinuous at certainty under $\{F^i_{\theta,m}\}$.*

The monotonicity of the likelihood ratio has weakened the conclusion of Theorem 2. Now, asymptotic agreement is discontinuous at certainty, but it is not almost-surely or strongly discontinuous.

In the proof of Theorem 2, 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 pointwise 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.

4.2 *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 we 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. In contrast, the limiting model specifies that each individual i is certain that $p_{\theta,\theta} = \hat{p}$. Consequently, in the limiting model, the asymptotic likelihood ratio is 0 around $\rho_a = \hat{p}$ and is ∞ around $\rho_a = 1 - \hat{p}$. Under uniform convergence, our next theorem establishes that whether asymptotic agreement holds is tied to the value of $R^i_{\theta,\theta',\infty}$ at the limiting frequencies, and it has no connection to the likelihood ratio in the limiting model:

THEOREM 4 (Characterization under uniform convergence). *Under Assumption 1, consider any $\{F_{\theta,m}^i\}$ converging to $\delta_{p_\theta^i}$, where $p_\theta^i \neq p_{\theta'}^i$ for all distinct θ, θ', and (m, i). Assume that for each (θ, θ', i, j), there exists an open neighborhood V_θ^i of p_θ^i on which $R_{\theta,\theta',m}^i$ converges uniformly to $R_{\theta,\theta',\infty}^i$. Then the following statements are true.*

(i) Asymptotic agreement is continuous at certainty under $\{F_{\theta,m}^i\}$ if and only if

$$R_{\theta,\theta',\infty}^j(p_\theta^i) = 0 \quad \forall i \neq j, \theta \neq \theta'.$$

(ii) If $R_{\theta,\theta',\infty}^j(p_\theta^i) \neq 0$ for all $i \neq j$ and $\theta \neq \theta'$, then asymptotic agreement is strongly discontinuous at certainty under $\{F_{\theta,m}^i\}$.

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, $R_{\theta,\theta',\infty}^j$, is equal to 0 at the frequency p_θ^i 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_θ^i 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, $R_{\theta,\theta',m}^j$, is nearly 0 at those frequencies. But these ratios are all approximately equal to $R_{\theta,\theta',\infty}^j(p_\theta^i)$ when $R_{\theta,\theta',m}^j$ converges uniformly.

In Theorem 4, 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 $R_{\theta,\theta',\infty}^j \neq 0$. In the binary example from Section 3, suppose instead that

$$f_{\theta,m}^i(p) = \begin{cases} 1/m + m(1 - 1/m) & \text{if } p \in (\hat{p}_m - 1/(2m), \hat{p}_m + 1/(2m)) \\ 1/m & \text{otherwise,} \end{cases}$$

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

(Counterexample to sufficiency). Without uniform convergence, asymptotic agreement may fail even if $R_{\theta,\theta',\infty}^j = 0$. Consider the variation of the binary example,

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

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

Theorem 4 and the counterexamples above show that uniform convergence of likelihood ratios is what ties continuity of asymptotic agreement to the primitives of the model. Under such a strong continuity concept, we have continuity if and only if the limit of likelihood ratios is zero at the limiting frequencies. It turns out that this condition is closely related to the tail properties of the underlying distributions—as we illustrate next in a canonical example.

4.3 *A canonical example*

As in the binary example, we take $\Theta = \{A, B\}$ and $\Sigma = \{a, b\}$. We consider a class of "symmetric" families $\{F_{\theta,m}^i\}$ that converge to the Dirac distribution $\delta_{p_\theta^i}$, where $p_A^i = (\hat{p}^i, 1 - \hat{p}^i)$ and $p_B^i = (1 - \hat{p}^i, \hat{p}^i)$ for some $\hat{p}^i \in (\frac{1}{2}, 1)$. 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 $f_{\theta,m}^i$ is then induced by f and the transformation $x \mapsto (x - \hat{p}^i)/m$. That is,

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

where $c^i(m) \equiv 1/\int_0^1 f(m(\rho_\theta - \hat{p}^i))\, d\rho_\theta$ is a correction factor to ensure that $f_{\theta,m}^i$ is a proper probability density function (p.d.f.) on $\Delta(\Sigma)$. Here, the mapping $x \mapsto (x - \hat{p}^i)/m$ 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 $F_{\theta,m}^i$ converges to $\delta_{p_\theta^i}$. Note that this model corresponds to the usual structure with small additive noise,

$$p_{\theta,\theta} = \hat{p}_i + \eta/m,$$

where the noise η has density f. We adjust the distribution $f_{\theta,m}^i$ so that $p_{\theta,\theta}$ remains within $[0, 1]$.

When $\hat{p}^1 = \hat{p}^2$, the individuals have the same subjective densities, leading trivially to asymptotic agreement. To analyze 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

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

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

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

This is a familiar condition in mathematical 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

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

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 $\alpha \in (0, \infty)$.[18] Moreover, Seneta (1976) shows that the convergence in Definition 4 is uniform on any compact set. Therefore, if f has regularly varying tails, then $R^j_{\theta, \theta', m}$ converges uniformly to $R^j_{\theta, \theta', \infty}$, where

$$R^j_{\theta, \theta', \infty}(p^i_\theta) = (\hat{x}/\hat{y})^{-\alpha} > 0$$

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

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

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

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

$$R^j_{\theta, \theta', \infty}(p^i_\theta) = (\hat{x}/\hat{y})^{-\infty} = 0.$$

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

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

(i) If f has regularly varying tails, then agreement is continuous at certainty under $\{F^i_{\theta, m}\}$.

(ii) If f has rapidly varying tails, then agreement is strongly discontinuous at certainty under $\{F^i_{\theta, m}\}$.

The intuition for these results is as follows. The continuity of agreement is determined by whether $R^j_{\theta, \theta', m}(p^i_\theta) \cong 0$. That is, whether i thinks that the frequency that will be realized at state θ (namely p^i_θ) will convince the other individual j, too, that the state is θ. Now, since $\hat{y} = |\hat{p}^i - \hat{p}^j| < |\hat{p}^i - (1 - \hat{p}^j)| = \hat{x}$, j also considers frequency p^i_θ 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 $m \to \infty$,

---

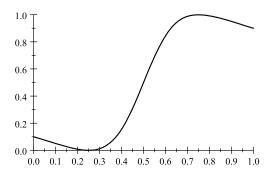[18]See, for example, de Haan (1970).

FIGURE 2. The limit $\lim_{n \to \infty} \phi_n^i(s)$ for Pareto distribution as a function of $\rho(s)$ (for $\alpha = 2$, $\hat{p}^i = \frac{3}{4}$).

j infers from the frequency p_θ^i that the state is θ (i.e., $R_{\theta,\theta',\infty}^j(p_\theta^i) = 0$). In contrast, 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, $R_{\theta,\theta',\infty}^j(p_\theta^i)$ 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 similarly to the Pareto distribution for large m. For simplicity, suppose $\pi_A^1 = \pi_A^2 = \frac{1}{2}$, so that there is no initial disagreement. The likelihood ratio is

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

and the asymptotic probability that the true state is θ is

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

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 $\rho_\theta(s) \cong \hat{p}^i$. In that case, $\phi_{\theta,\infty,m}^i(\rho(s))$ is approximately 1, and $\phi_{\theta,\infty,m}^j(\rho(s))$ is approximately $\hat{y}^{-\alpha}/(\hat{x}^{-\alpha} + \hat{y}^{-\alpha})$. Hence, both individuals believe that the difference between their asymptotic posteriors will be

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

This asymptotic difference is increasing with the difference $\hat{y} \equiv |\hat{p}^1 - \hat{p}^2|$, which corresponds to the difference in the individuals' views on which frequencies of signals are most likely. It is also clear from this expression that this asymptotic difference will converge to zero as $\hat{y} \to 0$ (i.e., as $\hat{p}^1 \to \hat{p}^2$).[19]

---

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

The dichotomy established here is quite intuitive. The fat-tailed regularly varying distributions often naturally arise from thin-tailed rapidly varying distributions when there is uncertainty about the latter distribution. For example, the t-distribution arises from the normal distribution with uncertainty about variance. When individuals are sure about their models (with thin tails), they use the data to resolve the uncertainty about θ. This leads to monotone beliefs and asymptotic agreement. When they face uncertainty about the model as well (with fat tails), they use the data to resolve both modeling uncertainty and the uncertainty about θ. This leads to non-monotone beliefs and asymptotic disagreement.

4.4 *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 a 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, medium-run agreement is continuous at certainty. To state this result, let $\mathbb{E}^i_m$ denote the expectation operator under Pr^i_m, and let $\delta_{p^i_\theta}$ be the Dirac measure that puts probability 1 on p^i_θ.

THEOREM 5 (Continuity of medium-run agreement). *Let $(p^1, p^2) \in (\Delta(\Sigma)^\Theta)^2$ be such that there is asymptotic agreement under $(\delta_{p^i_\theta})_{\theta \in \Theta, i \in \{1,2\}}$ (e.g., $p^1 = p^2$ and $p^i_\theta \neq p^i_{\theta'}$ for all distinct θ and θ'). Let $\{F^i_{\theta,m}\}$ be any family converging to $\delta_{p^i_\theta}$. Then, for every $\varepsilon > 0$ and $N < \infty$, there exist $\bar{n} < \infty$ and $\bar{m} < \infty$ such that*

$$\mathrm{Pr}^i_m(\|\phi^1_{n,m} - \phi^2_{n,m}\| > \varepsilon) < \varepsilon \quad (\forall m > \bar{m}, \forall n \in \{\bar{n}, \bar{n}+1, \ldots, \bar{n}+N\}).$$

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

$$\mathbb{E}^i_m[\|\phi^1_{n,m} - \phi^2_{n,m}\|] < \varepsilon \quad (\forall m > \bar{m}, \forall n \in \{\bar{n}, \bar{n}+1, \ldots, \bar{n}+N\}).$$

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 5 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.
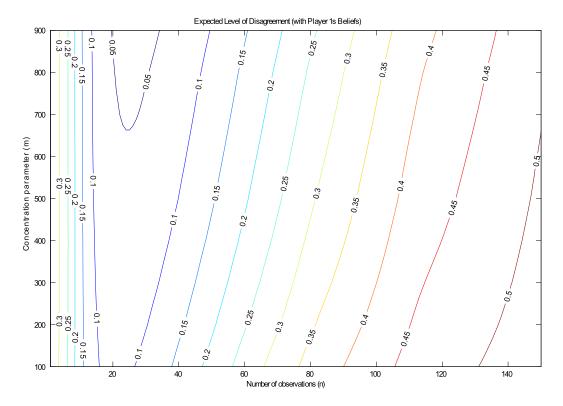
FIGURE 3. Ex ante expected disagreement in the binary example of Section 3.2 for $\epsilon = \lambda = 1/m$.

The idea of the proof is simple. First, in the standard model, the expected disagreement between the individuals' beliefs vanishes as $n \to \infty$, so that it becomes less than $\frac{1}{2}\varepsilon$ 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 an $\frac{1}{2}\varepsilon$ 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.

For an illustration, consider the binary example of Section 3.2 for $\epsilon = \lambda = 1/m$. Take $\pi_A^1 = \frac{2}{3}$ and $\pi_A^2 = \frac{1}{3}$ as the prior beliefs. In Figure 3, we plot the ex ante expected value of the disagreement after 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 $\frac{5}{9}$. 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 for longer before increasing. Theorem 5 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 up to the third digit, the expected disagreement starts growing rapidly only after 30 observations or so.

5. A MODEL OF ASSET TRADING

Failure of asymptotic agreement may have major implications in a game-theoretic analysis. In this section, we illustrate this with a simple asset trading example within the framework of the binary example in Section 3. There are three dates, $\tau = 0, 1, 2$, two players, 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 in Section 3.1. 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 Section 3.2. 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 $\rho_a(s) \in D_A^1 \equiv (\hat{p}^1 - \frac{1}{2}\lambda, \hat{p}^1 + \frac{1}{2}\lambda)$, 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 $1 - \pi_A^2$. Alternatively, 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

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

(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 $\pi_A^1(1 - \pi_A^2)$, 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 the 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.[20]

In our example, the players passively receive information between the two trading periods (and our results do not depend on whether the signals they receive have direct payoff implications, such as information concerning revenue streams or news on initial public offerings (IPOs)). A more major extension of our setting would be to allow them to take actions that impact the information they receive. Though this type of "active learning" may increase their information, unless it somehow restores full identification, it does not change the fundamental confounding problems that are at the root of the lack of asymptotic agreement in our model.

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

---

[20]This contrasts with the intuition that observation of common information should take agents toward 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).

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 toward 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 nondegenerate 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 (nonvanishing) 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 that learning foundations of common priors are not as strong as generally presumed.

Appendix A: Asymptotic agreement under full identification

In this appendix, we present a broad generalization of the asymptotic agreement result in the standard models to our framework. Toward this end, we write $\text{supp}(F_\theta^i)$ 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 $x \in \text{supp}(F_\theta^i)$ and any open neighborhood V of x, $\int_V dF_\theta^i > 0$. 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 from support restrictions). *For all $\theta \neq \theta'$ and i,* $\text{supp}(F_\theta^i) \cap \text{supp}(F_{\theta'}^i) = \varnothing$.

We characterize the cases in which asymptotic agreement obtains under this full-identification assumption and an additional genericity assumption. To state our characterization, we first define the *cross-entropy distance* between ρ and $p \in \Delta(\Sigma)$ as[21]

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

We also define

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

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 $\#P^i(\rho)$ to denote the number of elements in the set $P^i(\rho)$. Our genericity assumption is presented next.

ASSUMPTION 3 (Genericity). *For all $\theta \in \Theta$ and each $i \in \{1, 2\}$, $\text{supp}(F_\theta^i)$ is in the interior of $\Delta(\Sigma)$ and for each $\rho \in \text{supp}(F_\theta^i)$, $\#P^j(\rho) = 1$ (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 $\hat{p} \in P^j(\rho)$ and $\varepsilon \in (0, 1)$, $P^j(\varepsilon\rho + (1-\varepsilon)\hat{p}) = \{\hat{p}\}$.)

THEOREM 6 (Asymptotic learning and agreement under full identification). *Suppose Assumptions 2 and 3 hold. Then for each $i \in \{1, 2\}$ and $j \neq i$,*

(i) $\text{Pr}^i(\phi_{\theta,\infty}^i = 1 | \theta) = 1$

(ii) $\text{Pr}^i(\phi_\infty^1 = \phi_\infty^2) = 1$ *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.*

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

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 $\bigcup_{\theta'} \text{supp}(F_{\theta'}^j)$ 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. As a corollary to this characterization, the third part states that the common support assumption implies asymptotic agreement. The characterization is illustrated in the following example.

---

[21]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 $H(\rho, p) - H(p, p)$, where $H(p, p)$ is also the entropy of p. See, for example, Burnham and Anderson (1989).

EXAMPLE 3. In the binary example, consider the following case. Individual 1 thinks that, for any state θ, $p_{\theta,\theta}$ can be anywhere between $\frac{3}{4}$ and 1. Individual 2 thinks that $p_{A,a}$ can be anywhere between $\frac{3}{4}$ and 1, while $p_{B,b}$ can be anywhere between $\hat{p}$ and 1, where we let $\hat{p} > \frac{1}{2}$ vary. By symmetry, asymptotically, individual 1 assigns probability 1 on state A whenever $\rho_a > \frac{1}{2}$ and assigns probability 1 on state B whenever $\rho_a < \frac{1}{2}$. Individual 2 has a different cutoff,

$$\hat{\rho}_a = \log(4\hat{p})/\log(3\hat{p}/(1-\hat{p})) \in \left(1-\hat{p}, \tfrac{3}{4}\right),$$

which solves the equation $(\frac{3}{4})^\rho(\frac{1}{4})^{1-\rho} = (1-\hat{p})^\rho \hat{p}^{1-\rho}$. He assigns probability 1 on state A if $\rho_a > \hat{\rho}_a$ and on state B if $\rho_a > \hat{\rho}_a$. Whether there is asymptotic agreement depends on whether $\hat{\rho}_a > \frac{1}{4}$. When $\hat{\rho}_a < \frac{1}{4}$, individual 1 assigns positive probability to the event that the state is B and $\hat{\rho}_a < \rho_a < \frac{1}{4}$. On that event, asymptotically, individual 1 assigns probability 1 on state B, while individual 2 assigns probability 1 on state A, leading to extreme asymptotic disagreement. Indeed, the condition in the second part of Theorem 6 fails for that case: for any $\rho \in \mathrm{supp}(F_B^1)$ with $\hat{\rho}_a < \rho_a < \frac{1}{4}$, we have $P^2(\rho) = \{(\frac{3}{4}, \frac{1}{4})\} \not\subseteq \mathrm{supp}(F_B^2)$. In contrast, when $\hat{\rho}_a > \frac{1}{4}$, that condition holds and there is asymptotic agreement. For example, for any $\rho \in \mathrm{supp}(F_B^1)$, we have $\rho_a \leq \frac{1}{4} < \hat{\rho}_a$, and thus $P^2(\rho) = \{\rho\} \subseteq \mathrm{supp}(F_B^1) \subseteq \mathrm{supp}(F_B^2)$. (One can check that in all other cases, the condition holds.) Note that there is asymptotic agreement if and only if $\hat{p}$ is below a threshold, which is approximately 0.9878. $\Diamond$

Theorem 6 is a generalization of 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 toward common beliefs and priors.

COROLLARY 1 (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 $i = 1, 2$,*

(i) $\mathrm{Pr}^i(\phi_{\theta,\infty}^i = 1|\theta) = 1$

(ii) $\mathrm{Pr}^i(\phi_\infty^1 = \phi_\infty^2) = 1$.

APPENDIX B: PROOFS

PROOF OF LEMMA 1. Write

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

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 $f^i_{\theta'}(p_{\theta'})$ given r_n under the flat (Lebesgue) prior on $p_{\theta'}$ and the Bernoulli distribution on $\{s_t\}^n_{t=1}$. Denoting this by $\mathbb{E}^\lambda[f^i_{\theta'}|r_n]$ and denoting the denominator, which is similarly defined as the conditional expectation of $f_\theta(p_\theta)$, by $\mathbb{E}^\lambda[f^i_\theta|r_n]$, 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 $\mathbb{E}^\lambda[f^i_{\theta'}|r_n] \to f_{\theta'}(\rho)$ and $\mathbb{E}^\lambda[f^i_\theta|r_n] \to f^i_\theta(\rho)$ (Doob 1949 shows the convergence for almost all ρ, and Diaconis and Freedman 1990 show the convergence for all ρ in our multinomial model). This establishes

$$\frac{\text{Pr}^i(r_n|\theta')}{\text{Pr}^i(r_n|\theta)} \to R^i_{\theta,\theta'}(\rho),$$

as defined in (3). Equation (2) then follows from (1).   □

PROOF OF THEOREM 1. Fix any $(\varepsilon, \theta, \theta', i)$ with $\varepsilon > 0$ and $\theta \neq \theta'$. We will show that $\lim_{m\to\infty} \text{Pr}^i_m(R^i_{\theta,\theta',m} \geq \varepsilon|\theta) = 0$. 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

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

By definition,

$$\text{Pr}^i_m(R^i_{\theta,\theta',m} \geq \varepsilon|\theta) \leq \text{Pr}^i_m(D_m|\theta) + 1 - \text{Pr}^i_m(V|\theta). \tag{10}$$

But

$$\begin{aligned}
\text{Pr}^i_m(D_m|\theta) &= \int_{\rho \in D_m} f^i_{\theta,m}(\rho)\, d\rho \\
&\leq \frac{1}{\varepsilon} \int_{\rho \in D_m} f^i_{\theta',m}(\rho)\, d\rho = \frac{1}{\varepsilon} \text{Pr}^i_m(D_m|\theta') \\
&\leq \frac{1}{\varepsilon} \text{Pr}^i_m(V|\theta'),
\end{aligned} \tag{11}$$

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 (10) and (11), we obtain

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

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

PROOF OF THEOREM 2. Pick sequences $p^i_{\theta,m}$ and $\bar{\epsilon} > 0$ such that $p^i_{\theta,m} \to p^i_\theta$ and $\|p^i_{\theta,m} - p^j_{\theta',m}\| > \bar{\epsilon}/m$ for all $(\theta, i) \neq (\theta', j)$. For each (θ, i), define

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

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

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

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', $\phi^i_{\infty,m}(\rho) = \pi^j$ when $\rho \in D^i_{\theta,m}$. Hence, there exist $\bar{m}$ such that for any $m \geq \bar{m}$ and any $\rho \in D^i_m \equiv \bigcup_\theta D^{i,m}_\theta$,

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

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

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

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

PROOF OF THEOREM 3. For each $m \gg 1$, let

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

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 $x_\theta = 1 - \varepsilon(\hat{p}^i_{\theta'} - \frac{1}{2}\lambda) - \varepsilon^3(1 - \hat{p}^i_{\theta'} - \frac{1}{2}\lambda) \in (0, 1)$. Here, x_θ is close to 1 for large m. Then

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

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

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

and hence

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

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

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

$$
= \pi^1_B > 0
$$

and

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

$$
= \pi^2_A > 0,
$$

completing the proof.   □

PROOF OF THEOREM 4. *Part (i): Sufficiency.* Fix any $\hat{\theta}$, and assume that $R^j_{\hat{\theta},\theta,\infty}(p^i_{\hat{\theta}}) = 0$ for each $\theta \neq \hat{\theta}$. We will show that for every $\varepsilon > 0$,

$$
\lim_{m\to\infty} \Pr^i(\phi^1_{\hat{\theta},\infty,m} > 1 - \varepsilon, \phi^2_{\hat{\theta},\infty,m} > 1 - \varepsilon | \hat{\theta}) = 1, \tag{12}
$$

which implies that for each θ,

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

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

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

To prove (12), it suffices to prove that for every $\varepsilon > 0$, and every i, j and every θ,

$$
\lim_{m\to\infty} \Pr^i_m(R^j_{\hat{\theta},\theta,m} < \varepsilon | \hat{\theta}) = 1. \tag{13}
$$

Since Proposition 1 has established (13) for $j = i$ already, we only need to prove (13) 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 $R^j_{\hat\theta,\theta,\infty}(\rho) < \frac{1}{2}\varepsilon$ 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 $p^i_{\hat\theta}$, $\mathrm{Pr}^i_m(\hat V | \hat\theta) \to 1$, proving (13).

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

$$\lim_{m\to\infty} \mathrm{Pr}^i_m(|\phi^1_{\theta,\infty,m} - \phi^2_{\theta,\infty,m}| > \varepsilon_0|\theta) = 1. \tag{14}$$

This implies that

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

showing that agreement is discontinuous at certainty under $\{F^i_{\theta,m}\}$. To prove (14), we set

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

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 $R^j_{\theta,\theta',m}(\rho) > \frac{1}{2}R^j_{\theta,\theta',\infty}(p^i_\theta)$ 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, $\mathrm{Pr}^i_m(\hat V | \theta) \to 1$, showing that

$$\lim_{m\to\infty} \mathrm{Pr}^i_m(\phi^j_{\theta,\infty,m}(\rho) < 1 - 2\varepsilon_0|\theta) = 1. \tag{15}$$

Alternatively, by Proposition 1,

$$\lim_{m\to\infty} \mathrm{Pr}^i_m(\phi^i_{\theta,\infty,m}(\rho) > 1 - \varepsilon_0|\theta) = 1. \tag{16}$$

When $\phi^j_{\theta,\infty,m} < 1 - 2\varepsilon_0$ and $\phi^i_{\theta,\infty,m} > 1 - \varepsilon_0$, we have $|\phi^1_{\theta,\infty,m} - \phi^2_{\theta,\infty,m}| > \varepsilon_0$. Therefore, (15) and (16) imply (14), completing the proof of part (i).

Part (ii). Assume that $R^j_{\theta,\theta',\infty}(p^i_\theta) \neq 0$ for all $i \neq j$ and $\theta \neq \theta'$. Then, by (14), there exists $\varepsilon_1 > 0$ such that $\mathrm{Pr}^i_m(|\phi^1_{\theta,\infty,m} - \phi^2_{\theta,\infty,m}| > \varepsilon_1|\theta) \to 1$ for every θ. (To compute ε_1, replace $R^j_{\theta,\theta',\infty}(p^i_\theta)$ with $\min_{\theta,\theta'}\{R^j_{\theta,\theta',\infty}(p^i_\theta)\}$ in the definition of ε_0.) Therefore,

$$\lim_{m\to\infty} \mathrm{Pr}^i_m(\|\phi^1_{\infty,m} - \phi^2_{\infty,m}\| > \varepsilon_1) = 1,$$

showing that agreement is strongly discontinuous at certainty under $\{F^i_{\theta,m}\}$.   □

PROOF OF THEOREM 5. It suffices to show that $\mathbb{E}^i_m[\|\phi^1_{n,m} - \phi^2_{n,m}\|] < \varepsilon$ because $\mathrm{Pr}^i_m(\|\phi^1_{n,m} - \phi^2_{n,m}\| > \varepsilon) < \varepsilon$ whenever $\mathbb{E}^i_m[\|\phi^1_{n,m} - \phi^2_{n,m}\|] < \varepsilon^2$. 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

$$\lim_{n \to \infty} \mathbb{E}_\infty^i[\|\phi_{n,\infty}^1 - \phi_{n,\infty}^2\|] = \mathbb{E}_\infty^i\left[\lim_{n \to \infty} \|\phi_{n,\infty}^1 - \phi_{n,\infty}^2\|\right] = 0, \tag{17}$$

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., $\lim_{n \to \infty} \|\phi_{n,\infty}^1 - \phi_{n,\infty}^2\| = 0$ almost surely.

Next, we introduce a claim.

CLAIM 1. *For any fixed n, we have*

$$\lim_{m \to \infty} \mathbb{E}_m^i[\|\phi_{n,m}^1 - \phi_{n,m}^2\|] = \mathbb{E}_\infty^i[\|\phi_{n,\infty}^1 - \phi_{n,\infty}^2\|]. \tag{18}$$

Given this result, fix any $\varepsilon > 0$ and $N < \infty$. By (17), there exist $\bar{n} < \infty$ such that for all $n \geq \bar{n}$,

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

Alternatively, for each $n \geq \bar{n}$, by (18), there exists $\bar{m}_n < \infty$ such that for all $m > \bar{m}_n$,

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

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 $n \in \{\bar{n}, \bar{n}+1, \ldots, \bar{n}+N\}$,

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

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

PROOF OF CLAIM 1. For any r_n, we write

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

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

$$\lim_{m \to \infty} \Pr_m^i(r_n) = \sum_\theta \pi_\theta^i \prod_\sigma (p_{\theta,\sigma}^i)^{r_{n,\sigma}} \equiv \Pr_\infty^i(r_n) > 0, \tag{19}$$

where $\Pr_\infty^i(r_n)$ is the probability of r_n under the standard model. We also compute that

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

$$= \sum_{r_n} \left| \pi_\theta^i \int \prod_\sigma p_{\theta,\sigma}^{r_{n,\sigma}} \, dF_{\theta,m}^i(p_\theta) - \frac{\Pr_m^i(r_n)}{\Pr_m^j(r_n)} \pi_\theta^j \int \prod_\sigma p_{\theta,\sigma}^{r_{n,\sigma}} \, dF_{\theta,m}^j(p_\theta) \right|.$$

Now, as $m \to \infty$, $\int \prod_\sigma p_{\theta,\sigma}^{r_{n,\sigma}} \, dF_{\theta,m}^i(p_\theta) \to \prod_\sigma (p_{\theta,\sigma}^i)^{r_{n,\sigma}}$ (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 (19) and the continuous mapping theorem. Therefore,

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

$$= \mathbb{E}_m^i[|\phi_{\theta,n,\infty}^1 - \phi_{\theta,n,\infty}^2|].$$

Since $\max_\theta$ is continuous, one more application of the continuous mapping theorem yields (18), proving the claim. This completes the proof of the theorem. $\qquad\square$

PROOF OF THEOREM 6. Throughout the proof, Assumptions 2 and 3 are imposed. We first develop the necessary notation.

NOTATION. For any i, write

$$U^i = \{\rho \mid \#P^i(\rho) = 1\}.$$

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 $p^i(\rho) \in \mathrm{supp}(F_\theta^i)$. Writing $r_n(s) = (r_{\sigma,n}(s))_{\sigma \in \Sigma}$, note that

$$\phi_{\theta,n}^i(s) = \frac{1}{1 + \sum_{\theta' \neq \theta} \frac{\pi_{\theta'}^i}{\pi_\theta^i} \bar{R}_{\theta,\theta',n}^i(r_n(s))}, \tag{20}$$

where

$$\bar{R}_{\theta,\theta',n}^i(r_n(s)) = \frac{\int \prod_\sigma p_\sigma^{r_{\sigma,n}(s)} \, dF_{\theta'}^i(p)}{\int \prod_\sigma p_\sigma^{r_{\sigma,n}(s)} \, dF_\theta^i(p)} = \frac{\int e^{-H(r_n(s)/n,p)n} \, dF_{\theta'}^i(p)}{\int e^{-H(r_n(s)/n,p)n} \, dF_\theta^i(p)}. \tag{21}$$

We will use the 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. By definition, $H(\rho, p^i(\rho)) < H(\rho, p)$ for all $p \in \mathrm{supp}(F_{\theta'}^i)$. Since H is continuous and $\mathrm{supp}(F_{\theta'}^i)$ is closed, this implies that there exist open neighborhoods V_p and V_ρ of $p^i(\rho)$ and ρ, respectively, such that

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

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 $\hat{H}'$,

$$\int e^{-H(r_n(s)/n,p)n} \, dF_{\theta'}^i(p) \le e^{-\hat{H}'n}. \tag{22}$$

Moreover,

$$\int e^{-H(r_n(s)/n,p)n}\,dF^i_{\theta^i(\rho)}(p) \geq \int_{p\in V_p} e^{-H(r_n(s)/n,p)n}\,dF^i_{\theta^i(\rho)}(p) \geq e^{-\hat{H}n}\Pr^i(V_p|\theta^i(\rho)), \quad (23)$$

where the first equality is by nonnegativity of the exponential function and the second inequality is by definition of $\hat{H}$. Note that since $p^i(\rho) \in \operatorname{supp}(F^i_{\theta^i(\rho)})$ and V_p is an open neighborhood of $p^i(\rho)$, by definition of supp, $\Pr^i(V_p|\theta^i(\rho)) > 0$. Substituting (22) and (23) in definition (21), we then obtain

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

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

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

　　This lemma follows from Lemma 2 and (20).
　　We are now ready to prove Theorem 6.

PROOF OF PART (i). Take any θ. First, by Doob's consistency theorem,

$$\Pr^i(r_n(s)/n \to \rho(s) \in \operatorname{supp}(F^i_\theta)|\theta) = 1. \quad (24)$$

Moreover, for any $\rho \in \operatorname{supp}(F^i_\theta)$, $P^i(\rho) = \{\rho\}$, yielding $\rho \in U^i$ with $p^i(\rho) = \rho$ and $\theta^i(\rho) = \theta$. Then Lemma 3 establishes that

$$\Pr^i(\phi^i_{\theta,n}(s) \to 1|\theta) = 1.$$

PROOF OF PART (ii): Sufficiency. Take any $\hat{\theta}$. By Assumption 3, $\operatorname{supp}(F^i_{\hat{\theta}}) \subseteq U^j$. Assume that for every $\rho \in \operatorname{supp}(F^i_{\hat{\theta}}) \subseteq U^j$, $P^j(\rho) \subseteq \operatorname{supp}(F^j_{\hat{\theta}})$, so that $\theta^j(\rho) = \hat{\theta}$. Since $\rho \in \operatorname{supp}(F^i_{\hat{\theta}})$, $\theta^i(\rho) = \hat{\theta}$ as in part (i). Then, whenever $r_n(s)/n \to \rho(s) \in \operatorname{supp}(F^i_{\hat{\theta}})$, by Lemma 3, $\phi^i_{\hat{\theta},n}(s) \to 1$ and $\phi^j_{\hat{\theta},n}(s) \to 1$. Consequently, $|\phi^1_{\theta,n}(s) - \phi^2_{\theta,n}(s)| \to 0$ for each $\theta \in \Theta$. Therefore, by (24), for each $\theta \in \Theta$,

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

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

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

PROOF OF PART (ii): Necessity. Suppose that for some $\hat{\theta}$ and $\hat{\rho} \in \operatorname{supp}(F^i_{\hat{\theta}})$, $P^j(\hat{\rho}) \nsubseteq \operatorname{supp}(F^j_{\hat{\theta}})$, 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 $\rho(s) \in \hat{V} \cap \operatorname{supp}(F_{\hat{\theta}}^i)$, Lemma 3 implies that $\phi_{\hat{\theta},n}^i(s) \to 1$ and $\phi_{\theta^j(\hat{\rho}),n}^j(s) \to 1$, so that $|\phi_{\hat{\theta},n}^1(s) - \phi_{\hat{\theta},n}^2(s)| \to 1$. Since $\hat{V}$ is an open neighborhood of $\hat{\rho} \in \operatorname{supp}(F_{\hat{\theta}}^i)$, $\operatorname{Pr}^i(\rho(s) \in \hat{V} \cap \operatorname{supp}(F_{\hat{\theta}}^i)|\hat{\theta}) > 0$, showing that

$$\operatorname{Pr}^i\big(|\phi_{\hat{\theta},n}^1(s) - \phi_{\hat{\theta},n}^2(s)| \to 1\big) > 0,$$

and thus contradicting asymptotic agreement.   □

References

Acemoglu, Daron, Victor Chernozhukov, and Muhamet Yildiz (2006), "Learning and disagreement in an uncertain world." Report. [187]

Aumann, Robert (1998), "Common priors: A reply to Gul." *Econometrica*, 66, 929–938. [191]

Aumann, Robert J. (1987), "Correlated equilibrium as an expression of Bayesian rationality." *Econometrica*, 55, 1–18. [191]

Berk, Robert (1966), "Limiting behavior of posterior distributions when the model is incorrect." *The Annals of Mathematical Statistics*, 37, 51–58. [193]

Billingsley, Patrick (1995), *Probability and Measure*, third edition. Wiley, New York. [194]

Blackwell, David and Lester Dubins (1962), "Merging of opinions with increasing information." *The Annals of Mathematical Statistics*, 33, 882–886. [192, 213]

Burnham, Kenneth and David Anderson (1989), *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, New York. [215]

Cripps, Martin, Jeffrey C. Ely, George J. Mailath, and Larry Samuelson (2008), "Common learning." *Econometrica*, 76, 909–933. [193]

de Haan, Laurens (1970), *On Regular Variation and Its Application to the Weak Convergence of Sample Extremes*, volume 32 of Mathematical Centre Tract. Mathematisch Centrum, Amsterdam. [208]

Diaconis, Persi and David Freedman (1986), "On inconsistent Bayesian estimates of location." *The Annals of Statistics*, 14, 68–87. [192]

Diaconis, Persi and David Freedman (1990), "On the uniform consistency of Bayes estimates for multinomial probabilities." *The Annals of Statistics*, 18, 1317–1327. [217]

Doob, Joseph L. (1949), "Application of the theory of martingales." In *Le calcul des probabilities et ses applications*, 23–27, CNRS, Paris. [217]

Freedman, David A. (1963), "On the asymptotic behavior of Bayes estimates in the discrete case. I." *The Annals of Mathematical Statistics*, 34, 1386–1403. [192]

Freedman, David (1965), "On the asymptotic behavior of Bayes estimates in the discrete case. II." *The Annals of Mathematical Statistics*, 36, 454–456. [192]

Fudenberg, Drew and David Kreps (1995), "Learning in extensive games. I. Self-confirming equilibria." *Games and Economic Behavior*, 8, 20–55. [194]

Fudenberg, Drew and David K. Levine (1993), "Self-confirming equilibria." *Econometrica*, 61, 523–545. [194]

Fudenberg, Drew and David K. Levine (2006), "Superstition and rational learning." *American Economic Review*, 96, 630–651. [194]

Gul, Faruk (1998), "A comment on aumann's Bayesian view." *Econometrica*, 66, 923–927. [191]

Harrison, J. Michael and David M. Kreps (1978), "Speculative investor behavior in a stock market with heterogenous expectations." *Quarterly Journal of Economics*, 92, 323–336. [213]

Jackson, Matthew O., Ehud Kalai, and Rann Smorodinsky (1999), "Bayesian representation of stochastic processes under learning: De Finetti revisited." *Econometrica*, 67, 875–893. [194]

Kurz, Mordecai (1994), "On the structure and diversity of rational beliefs." *Economic Theory*, 4, 877–900. [193]

Kurz, Mordecai (1996), "Rational beliefs and endogenous uncertainty: An introduction." *Economic Theory*, 8, 383–397. [193]

Miller, Ronald I. and Chris W. Sanchirico (1999), "The role of absolute continuity in 'merging of opinions' and 'rational learning'." *Games and Economic Behavior*, 29, 170–190. [192]

Morris, Stephen (1996), "Speculative investor behavior and learning." *Quarterly Journal of Economics*, 111, 1111–1133. [213]

Savage, Leonard J. (1954), *The Foundations of Statistics*. Wiley, New York. Reprinted in 1972 by Dover, New York. [188, 194, 213, 216]

Seneta, Eugene (1976), *Regularly Varying Functions*, volume 508 of Lecture Notes in Mathematics. Springer, Berlin. [208]

Stinchcombe, Maxwell (2005), "The unbearable flightness of Bayesians: Generically erratic updating." Report, University of Texas, Austin. [192]

Van der Vaart, Aad W. (1998), *Asymptotic Statistics*, volume 3 of Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge. [197]