GAMES AND ECONOMIC BEHAVIOR 9, 161–171 (1995)

Stochastic Common Learning

DOV MONDERER

Faculty of Industrial Engineering and Management, Technion, Haifa, Israel

AND

DOV SAMET

Faculty of Management, Tel Aviv University, Tel Aviv, Israel

Received May 13, 1991

We present a general model in which agents learn by observing a stochastic process. We show that agents not only learn in such a model but also learn commonly. By this we mean that in the long run agents will form common beliefs concerning important facts and in particular they will commonly believe that their views regarding the future are similar. *Journal of Economic Literature* Classification Numbers: C72, C73, D82. © 1995 Academic Press, Inc.

1. INTRODUCTION

The purpose of this paper is to bridge a gap between two types of results concerning learning processes in which several agents gain new information and update their beliefs according to Bayes' rule. In the first type the emphasis is on convergence of the learning process to common knowledge; learning takes place by information exchange, which leads eventually to common knowledge of certain facts. In the other, agents are engaged in a repeated game and learn about each other's behavior by observing past moves. The purpose of the latter model is to show convergence to equilibrium, but the conclusions regarding the learning process are limited. The environment is usually stochastic and the players gain hardly any knowledge (i.e., facts which are assigned probability one) let alone common knowledge. We present here a general model in which agents learn by observing a stochastic process, and we show that even

0899-8256/95 $6.00

in this general model agents not only learn but also learn commonly. By this we mean that in the long run agents will form common beliefs concerning important facts and in particular they will commonly believe that their views regarding the future are similar.

A model of the first type is presented by Geanakoplos and Polemarchakis (1982). In this model two agents inform each other repeatedly about the probability they assign to a fixed event. After each period they update their posterior beliefs on the basis of what they learn about each other. The process converges after a finite time,and from then on the agents do not change the probability they assign to the fixed event. These probabilities become common knowledge and therefore by Aumann (1976) they are equal. In a similar process discussed by Sebenius and Geanakoplos (1983) the agents announce their consent to bet on some fixed event. Here again, the process reaches a point at which the posterior probability of the event becomes common knowledge and as a result one of the agents announces his refusal to bet. More elaborate exchanges of information in groups of more than two agents are discussed by Parikh and Krasucki (1990).

Models of mutual learning in games which leads to equilibrium have recently occupied the attention of several authors. (See Kalai and Lehrer (1993) for a comprehensive reference list). In Kalai and Lehrer (1993) learning is modeled by Bayesian updating. Each player starts with initial erroneous beliefs regarding the strategies of all other players. The authors show that if each player assigns a positive probability to the real strategy played by the others,[1] their beliefs about the future converge in the long run. The merging of players' opinions in this model is related to a result of Blackwell and Dubins (1962) concerning a single agent who learns about the probability distribution of the future by observing a stochastic process.[2]

In this paper we study a general model of mutual learning in a Bayesian model. This model includes repeated games with incomplete information as a special case. The agents start with the same prior distribution on a state space. In addition they have private information which is given by finite partitions. At every stage the outcome of a stochastic process is observed by the agents and becomes common knowledge among them. Each player updates his beliefs according to the observed outcome. We show that the agents' predictions regarding the future outcomes of the process merge in the long run. More importantly, with probability one this merging of opinions is common belief. In other words, the similarity of beliefs becomes almost common knowledge.

[1] This is the "grain of truth" assumption in Kalai and Lehrer (1993). In fact, they use a weaker assumption to derive their results.

[2] It is assumed in this model that the true probability distribution is absolutely continuous with respect to the agent's prior probability distribution.

To illustrate this result, imagine two stockbrokers who graduated from the same school and have the same view of the business world (this corresponds to the common prior assumption). The brokers are employed by different companies and therefore acquire different information about the world (this differential information is modeled by the individual partitions of the agents—the private information). In particular they evaluate differently the future of the stochastic process of stock prices. It turns out that the significance of the initial differential information diminishes in time. By observing the stock prices, which are common knowledge, the brokers learn something about the initial information of each other and about future stock prices. The learning process converges and in the long run the brokers have similar assessments of the distribution of future prices. Moreover, our result says that this similarity of assessments becomes common belief among the agents. By this we mean that they believe that their assessments are similar, they believe that they believe that the assessments are similar, etc., where "believe" stands for "assign high probability."

The notion of "common belief," as expressed in the previous sentence, is similar to the notion of common knowledge. While relaxing the level of certitude, by replacing "know" with "believe," the concept of common belief preserves the sense of the way in which information is shared among the agents.[3] This work demonstrates the ability of common belief to convey the extent to which information is shared for cases in which common knowledge does not exist. Other examples where common belief serves as a good proxy to common knowledge are discussed in Monderer and Samet (1989, 1990).

2. The Common Learning Theorem

Let (Ω, Σ, μ) be a probability space. By a *partition* we mean a countable measurable partition of Ω. A finite set I is the set of agents. Information of the agents in each period $t = 0, 1, 2, \ldots$ is given by partitions which we describe as follows. At time $t = 0$ agent i has a finite[4] partition Π_i^0 with M_i elements. For $t = 1, 2, \ldots$ let f^t be a discrete[5] random variable, and let Π^t be the partition generated by $f^1, \ldots, f^t$. The agents observe

[3] This special way in which information is shared, which is hinted at in the word "common," is expressed by the infinite hierarchy of information about information. There are also alternative ways to describe it (see Monderer and Samet (1989)).

[4] The finiteness of the initial information is crucial for our results. We discuss it at the end of Section 2.

[5] Our results hold without this assumption, but the proofs and notations may become more cumbersome.

the value of f^t at each period $t \geq 1$ and as a result agent i's information at time $t \geq 1$ is given by the partition Π_i^t which is the join[6] $\Pi_i^0 \vee \Pi^t$. We will assume that all sets in all the involved partitions have positive probability.[5]

For a given event E and $0 \leq p \leq 1$, let $B_{t,i}^p(E)$ be the event that agent i *p-believes E at time t*, i.e., the event that at this time he assigns a probability of at least p to E. Thus

$$B_{t,i}^p(E) = \{\omega \mid \mu(E \mid \Pi_i^t(\omega)) \geq p\},$$

where $\Pi_i^t(\omega)$ is the element of Π_i^t which contains ω. We denote by $B_t^p(E)$ the event that all agents p-believe E, that is,

$$B_t^p(E) = \cap_{i \in I} B_{t,i}^p(E).$$

The event that E is *common p-belief* is

$$C_t^p(E) = \cap_{n=1}^{\infty} (B_t^p)^n(E).$$

For further details concerning common beliefs see Monderer and Samet (1989).

The results of this paper are the consequence of the following principle.

THE COMMON LEARNING THEOREM.[7] *Let $(S_t)_{t \geq 1}$ be a nondecreasing sequence of events such that $\lim_{t \to \infty} \mu(S_t) = 1$ and let $0 \leq p < 1$. Then for almost all ω there exists time $T = T(\omega)$ such that for each time $t \geq T$, the agents commonly p-believe S_t, at time t in ω, i.e., $\omega \in C_t^p(E)$. Equivalently,*

$$\mu(\cup_{T=1}^{\infty} \cap_{t \geq T} C_t^p(S_t)) = 1.$$

To prove this theorem we use the following two lemmata.

LEMMA 1. *Let $M = \sum_{i \in I} M_i$. Then for any event E, time t, and probability p,*

[6] The join of two partitions is the coarsest common refinement of the partitions. In terms of the fields generated by the partitions, the join corresponds to the smallest field which contains both fields.

[7] Fudenberg and Tirole (1991) prove a theorem of the same nature for the case in which the information given to the agents in each period is trivial. In their theorem the set S is also fixed in time but the probability distribution changes such that the probability of S approaches 1.

$$C_t^p(E) = (B_t^p)^M(E).$$

Proof. Note that for $t \geq 1$ each element of Π^t is a union of elements of Π_i^t for all $i \in I$. This implies that for any event E and each agent i, $B_{t,i}^p(E) = \cup_{j\geq 1} B_{t,i}^p(E_j)$, where $(E_j)_{j\geq 1}$ is the partition of E induced by Π^t. The partition that Π^t induces on $B_{t,i}^p(E)$ is precisely $(B_{t,i}^p(E_j))_{j\geq 1}$. Therefore for each n, $(B_t^p)^n(E) = \cup_{j\geq 1}(B_t^p)^n(E_j)$, and also $C_t^p(E) = \cup_{j\geq 1} C_t^p(E_j)$.

Observe that each element of Π^t is a union of at most M_i elements of Π_i^t. It is easy to see that for any event X the sequence $((B_t^p)^n(X))_{n\geq 1}$ is decreasing. Thus for each j, $(B_t^p)^n(E_j)$ is constant for $n \geq M$ and therefore $C_t^p(E_j) = (B_t^p)^M(E_j)$. We conclude that

$$C_t^p(E) = \cup_{j\geq 1} C_t^p(E_j) = \cup_{j\geq 1}(B_t^p)^M(E_j) = (B_t^p)^M(E),$$

as we wanted to prove. ∎

LEMMA 2. *Let $(S_t)_{t\geq 1}$ be a nondecreasing sequence of events such that* $\lim_{t\to\infty} \mu(S_t) = 1$. *Then for all* $0 \leq p < 1$,

$$\mu(\cup_{T=1}^{\infty} \cap_{t\geq T} B_t^p(S_t)) = 1.$$

Equivalently, for almost all ω there exists $T = T(\omega)$ such that for all $t \geq T$, $\omega \in B_t^p(S_t)$.

Proof.[8] It is enough to show that the lemma holds for $B_{t,i}^p$ (rather than B_t^p) for each agent i. Denote $B_t = B_{t,i}^p(S_t)$ and let $\overline{B}_t$ be the complement of B_t. It suffices to show that $\lim_{T\to\infty} \mu(\cup_{t\geq T}\overline{B}_t) = 0$. Note that for each t and for each $S \in \Pi_t$ which is a subset of $\overline{B}_t$, $\mu(\overline{S}_t | S) \geq 1 - p$, where $\overline{S}_t$ is the complement of S_t. Therefore, also for $T \leq t$, $\mu(\overline{S}_T | S) \geq 1 - p$, since $(S_t)_{t\geq 1}$ is nondecreasing.

Thus, for every $T \geq 1$,

$$\mu(\overline{S}_T | \cup_{t>T}\overline{B}_t) \geq 1 - p,$$

which implies that $\mu(\overline{S}_T) \geq (1 - p)\mu(\cup_{t\geq T}\overline{B}_t)$. As $\lim_{T\to\infty} \mu(\overline{S}_T) = 0$ the result follows. ∎

Proof of the Common Learning Theorem. We prove by induction on m that

$$\lim_{T\to\infty} \mu(\cap_{t\geq T}(B_t^p)^m(S_t)) = 1.$$

[8] Kalai and Lehrer (1993) proved a special case of Lemma 2. We use a similar proof.

By Lemma 1, the case $m = M$ is what we need to prove. The case $m = 1$ is proved in Lemma 2. Suppose we proved for m. Denote

$$\hat{S}_T = \cap_{t \geq T} (B_t^p)^m (S_t).$$

Note that since $\hat{S}_t \subseteq B_t^p(S_t)$, it follows that $B_t^p(\hat{S}_t) \subseteq (B_t^p)^{m+1}(S_t)$. Now, since $(\hat{S}_t)_{t \geq 1}$ is nondecreasing and by the induction hypothesis it follows that $\lim_{t \to \infty} \mu(\hat{S}_t) = 1$. Thus by Lemma 2,

$$1 = \lim_{T \to \infty} \mu(\cap_{t \geq T} B_t^p(\hat{S}_t)) \leq \lim_{T \to \infty} \mu(\cap_{t \geq T} (B_t^p)^{m+1}(S_t)),$$

which proves our claim for $m + 1$. ■

The finiteness of the initial partitions is crucial in the Common Learning Theorem. A counterexample is easy to construct when these partition are infinite. Suppose there are two agents, 1 and 2. Let

$$\Omega = \{0, 1, \ldots, n, \ldots\}.$$

The partition of agent 1 is $\{\{0\}, \{1, 2\}, \{3, 4\}, \ldots, \{2n - 1, 2n\} \ldots\}$ and that of agent 2 is $\{\{0, 1\}, \{2, 3\}, \ldots, \{2n, 2n + 1\} \ldots\}$. Let the probability distribution over Ω be $\mu(n) = (1 - q)q^n$, for $0 < q < 1$. Suppose the signals the agents receive in each period are trivial, i.e., $\Pi^t = \Omega$ for each t. Thus for all t, $B_{t,i}^p = B_{0,i}^p$. Let $S_t = \{1, \ldots, t\}$. Consider an odd t. Then for any $p > 1/(1 + q)$, $t \notin B_{0,i}^p(S_t)$. Therefore also $(t - 1) \notin B_{0,2}^p(B_{0,1}^p(S_t))$. Continuing down to $t = 0$ we see that $C_t^p(S_t) = \varnothing$. This can be easily shown also for even t's and thus the Common Learning Theorem does not hold.

Note that the Common Learning Theorem is stated for any $p < 1$ but fails to hold for $p = 1$. Indeed, change in the previous example the initial partition for both players to be $\{\Omega\}$. Then $B_{t,i}^1(S_t) = \varnothing$ for all t and $i = 1$, 2 and therefore $C_t^p(S_t) = \varnothing$.

3. Common Learning of the Initial Information

At time 0 each of the agents has a posterior probability regarding the initial information of the other agents. As the stochastic process unfolds, each agent updates this posterior probability. In the long run the posteriors converge, which means that there is not much to learn anymore. Not much can be said in general about the limit of this converging learning. Clearly the agents do not necessarily learn the true initial information of each other and their beliefs do not even have to merge in the long run. We show in this section, however, that the fact that the agents have already learnt whatever there was to be learnt becomes almost common knowledge or more precisely common belief.

For agent i we will denote by $\mu_i^t(\omega)$ the posterior probability distribution of i in state ω at time t. That is, for all $S \in \Sigma$,

$$\mu_i^t(\omega)(S) = \mu(S \mid \Pi_i^t(\omega)).$$

Let Π^0 be the join of all agents' initial partitions, that is, $\Pi^0(\omega) = \cap_{i\in I}\Pi_i^0(\omega)$, for each ω. The numbers $(\mu_i^t(\omega)(S))_{S\in\Pi^0}$ describe agent i's posteriors of the initial information of all agents. By the Martingale Convergence Theorem these posteriors converge almost surely. Set for each $S \in \Pi^0$,

$$x_i^\infty(\omega)(S) = \lim_{t\to\infty} \mu_i^t(\omega)(S).$$

For $\varepsilon > 0$ and for $T \geq 1$ let F_T^ε be the event that the posteriors of the initial information of each agent have converged by the time T to an ε-neighborhood of their limit, i.e.,

$$F_T^\varepsilon = \cap_{i\in I} \cap_{t\geq T} \cap_{S\in\Pi^0} \{\omega \in \Omega : \mid \mu_i^t(\omega)(S) - x_i^\infty(\omega)(S) \mid < \varepsilon\}.$$

THEOREM 1. *Let $\varepsilon > 0$ and $0 \leq p < 1$. Then for almost all ω there exists a time $T = T(\omega)$ such that for all $t \geq T$ it is common p-belief at t that the posteriors of the agents, concerning their initial information, have converged to an ε-neighborhood of their limit.*

Proof. Note that the sequence $(\mu(F_t^\varepsilon))_{t\geq 1}$ converges to 1 by the Martingale Convergence Theorem, and apply the Common Learning Theorem to the sequence $(F_t^\varepsilon)_{t\geq 1}$. ■

The following lemma will be used in the next section.

LEMMA 3. *For almost all ω and for all i,*

$$x_i^\infty(\omega)(\Pi^0(\omega)) > 0.$$

Proof. As I is finite it suffices to prove the claim for a fixed agent i. Let $S \in \Pi_0$ and set $D = \{\omega \in \Omega \mid x_i^\infty(\omega)(S) = 0\}$. It is enough to show that $\mu(S \cap D) = 0$. By the Martingale Convergence Theorem, $x_i^\infty(\omega)(S) = E(1_S \mid \Sigma_i^\infty)(\omega)$, where Σ_i^∞ is the σ-field generated by $\Pi_i^0 \cup (\cup_{t=1}^\infty \Pi^t)$. As $D \in \Sigma_i^\infty$,

$$\int_D E(1_S \mid \Sigma_i^\infty)(\omega)\, d\mu(\omega) = \int_D 1_S(\omega)\, d\mu(\omega).$$

Because $E(1_S \mid \Sigma_i^\infty) = 0$ on D, $\mu(S \cap D) = 0$. ∎

4. Common Learning of the Future

At each point in time each agent anticipates the future of the stochastic process according to a probability distribution which is based on the information he has accumulated. In the long run the distributions of the agents become similar and moreover they are similar to the distribution that could have been formed by the agents had they shared their information. We show that this similarity of distributions is commonly believed by the agents.

We start by defining similarity of distributions. Let ξ_1 and ξ_2 be two probability measures on (Ω, Σ). We say that ξ_1 and ξ_2 are δ-close (with respect to the stochastic process $f^1, f^2, \ldots,$) if there exists an event B in the σ-algebra generated by the process such that the following two conditions are satisfied:

(a) $\xi_i(B) \geq 1 - \delta$ for $i = 1, 2$, and

(b) for almost all $u \in B$, for all $t \geq 1$,

$$e^{-\delta} < \frac{\xi_1(\Pi^t(u))}{\xi_2(\Pi^t(u))} < e^{\delta}.$$

The last condition can also be written in a way that reveals the symmetry of ξ_1 and ξ_2, as

$$|\log \xi_1(\Pi^t(u)) - \log \xi_2(\Pi^t(u))| < \delta.$$

Denote by A_T^δ the set of all states u for which for all $t \geq T$ the posterior probability distributions $(\mu_i^t(u))_{i \in I}$ are pairwise δ-close. That is,

$$A_T^\delta = \cap_{t \geq T} \cap_{i,j} \{u \mid \mu_i^t(u) \text{ and } \mu_j^t(u) \text{ are } \delta\text{-close}\}.$$

Theorem 2. *Let $0 \leq p < 1$ and $\delta > 0$. Then for almost all ω there is a time $T = T(\omega)$, such that for all $t \geq T$ the following holds:*

(a) *$\mu_i^t(\omega)$ and $\mu_j^t(\omega)$ are δ-close for all $i, j \in I$, and*

(b) *it is common p-believed in ω at time t that from time t, the* *posterior probability distributions of all agents will be pairwise δ-close forever. That is,*

$$\mu(\cup_{T=1}^{\infty}(A_T^{\delta} \cap (\cap_{t\geq T} C_t^p(A_t^{\delta})))) = 1.$$

Proof. Since $(A_t^{\delta})_{t\geq 1}$ is nondecreasing, it suffices to show, by the Common Learning Theorem, that $\lim_{T\to\infty} \mu(A_T^{\delta}) = 1$.

Let $B_{t,I}^q$ be the q-belief operator of the join of all agents in I. That is, for each event S,

$$B_{t,I}^q(S) = \{\omega \mid \mu(S \mid \Pi^t(\omega) \cap \Pi^0(\omega)) \geq q\}.$$

Also, set

$$D_\eta = \cap_{i\in I}\{\omega : x_i^{\infty}(\omega)(\Pi^0(\omega)) > \eta\}.$$

Note that by Lemma 3, $\lim_{\eta\to 0} \mu(D_\eta) = 1$.

We will show that for all $T \geq 1$ and $\eta \geq 0$,

$$D_\eta \cap F_T^{\varepsilon} \cap (\cap_{t\geq T} B_{t,I}^q(F_t^{\varepsilon})) \subseteq A_T^{\delta} \tag{1}$$

for sufficiently small ε and big q. As by Lemma 2,

$$\lim_{T\to\infty} F_T^{\varepsilon} \cap (\cap_{t\geq T} B_{t,I}^q(F_t^{\varepsilon})) = 1,$$

by (1),

$$\mu(D_\eta) \leq \lim_{T\to\infty} \mu(A_T^{\delta}).$$

Since $\lim_{\eta\to 0} \mu(D_\eta) = 1$, the result will follow.

To prove (1) consider $\omega \in D_\eta \cap F_T^{\varepsilon} \cap (\cap_{t\geq T} B_{t,I}^q(F_t^{\varepsilon}))$. We proceed to show that if ε is close enough to 0 and q close enough to 1 then for $t \geq T$, and for all $i, j \in I$, $\mu_i^t(\omega)$ and $\mu_j^t(\omega)$ are δ-close.

Set

$$A = \Pi^t(\omega) \cap \Pi^0(\omega) \cap F_t^{\varepsilon},$$

and for every $s \geq 1$ denote

$$B_s = \cup_{u \in A} \Pi^{t+s}(u)$$

and

$$B = \cap_{s \geq 1} B_s.$$

The set B_s includes each point in Ω which shares the same history of the stochastic process, until time $t + s$, with some point in A. We complete the proof by showing that for appropriate ε and q, for all $v \in B$ and for all $s \geq 1$,

$$e^{-\delta} < \frac{\mu_i^t(\omega)(\Pi^{t+s}(v))}{\mu_j^t(\omega)(\Pi^{t+s}(s))} < e^{\delta} \tag{2}$$

and

$$\mu_i^t(\omega)(B) \geq 1 - \delta. \tag{3}$$

To evaluate the middle term in (2) we consider the ratio

$$Z_i = \frac{\mu_i^t(\omega)(\Pi^{t+s}(v))}{\mu_l^t(\omega)(\Pi^{t+s}(v))}.$$

As $v \in B_s$ there is $u \in A$ with $\Pi^{t+s}(u) = \Pi^{t+s}(v)$. Since also $u \in \Pi_i^t(\omega)$ it follows that $\mu_i^t(\omega) = \mu_i^t(u)$. Substituting u for both v and ω in (2) and developing the numerator and denominator of Z_i we find that

$$\begin{aligned} Z_i &= \frac{\mu(\Pi_i^0(u) \cap \Pi^t(u))}{\mu(\Pi_i^t(u))} : \frac{\mu(\Pi^{t+s}(u) \cap \Pi^0(u)\Pi^t(u))}{\mu(\Pi^{t+s}(u) \cap \Pi_i^0(u))} \\ &= \frac{\mu(\Pi^0(u) \cap \Pi_i^t(u))}{\mu(\Pi_i^t(u))} : \frac{\mu(\Pi_i^{t+s}(u) \cap \Pi^0(u))}{\mu(\Pi_i^{t+s}(u))} \\ &= \frac{\mu_i^t(u)(\Pi^0(u))}{\mu_i^{t+s}(u)(\Pi^0(u))}. \end{aligned}$$

Since $u \in F_t^\varepsilon$ it follows that $|\mu_i^t(u)(\Pi^0(u)) - \mu_i^{t+s}(u)(\Pi^0(u))| \leq 2\varepsilon$. Since $\omega \in F_t^\varepsilon \cap D_\eta$, we have $\mu_i^t(\omega)(\Pi^0(w)) \geq \eta - \varepsilon$. Remembering that $\mu_i^t(u) = \mu_i^t(\omega)$ we conclude that $\mu_i^{t+s}(\Pi^0(w)) \geq \eta - 3\varepsilon$, which is positive for small enough ε. Thus, $|Z_i - 1| < 2\varepsilon/(\eta - 3\varepsilon)$. This, for small ε, guarantees that

$$e^{-\delta/2} < Z_i < e^{\delta/2}. \tag{4}$$

If we repeat the argument with j instead of i it yields

$$e^{-\delta/2} < Z_j < e^{\delta/2}.$$

Taking the ratio of Z_i and Z_j we prove (2). To show (3) we note that $B_s \supseteq F_t^\varepsilon$ and $\omega \in B_{t,I}^q(F_t^\varepsilon)$ and therefore $\mu_I^t(\omega)(B_s) \geq q$. In addition by (4), $\mu_i^t(\omega)(\Pi^{t+s}(v)) \geq e^{-\delta/2}\mu_I^t(\omega)(\Pi^{t+s}(v))$, for each $v \in B_s$. Thus, $\mu_i^t(\omega)(B_s) \geq qe^{-\delta/2} \geq 1 - \delta$, if q is close enough to 1. As $(B_s)_{s\geq 1}$ is decreasing, (2) is proved. ∎

REFERENCES

AUMANN, R. (1976). "Agreeing to Disagree," *Ann. Statist.* **4**, 1236–1239.

BLACKWELL, D., AND DUBINS, L. (1962). "Merging of Opinions with Increasing Information," *Ann. Math. Statist.* **38**, 882–886.

FUDENBERG, D., AND TIROLE, J. (1991). *Game Theory*, Cambridge, MA: The MIT Press.

GEANAKOPLOS, J., AND POLEMARCHAKIS, H. (1982). "We Can't Disagree Forever," *J. Econ. Theory* **28**, 192–200.

KALAI, E., AND LEHRER, E. (1993). "Rational Learning Leads to Nash Equilibrium," *Econometrica* **61**, 1019–1045.

MONDERER, D., AND SAMET, D. (1989). "Approximating Common Knowledge with Common Beliefs," *Games Econ. Behav.* **1**, 170–190.

MONDERER, D., AND SAMET, D. (1990). "Proximity of Information in Games with Incomplete Information," Working Paper No. 63/90, Faculty of Management, Tel Aviv University.

PARIKH, R., AND KRASUCKI, P. (1990). "Communication, Consensus, and Knowledge," *J. Econ. Theory* **52**, 178–189.

SEBENIUS, J., AND GEANAKOPLOS, J. (1983). "Don't Bet on It: Contingent Agreements with Assymetric Information," *J. Amer. Statist. Assoc.* **78**, 424–426.