Belief Learning

In the study of learning in games, *belief learning* refers to models in which players are engaged in a dynamic game and each player optimizes, or ε optimizes, with respect to a *prediction rule*, which gives a forecast of future opponent behavior as a function of the current history. This article focuses on the most studied class of dynamic games, two player discounted repeated games with finite stage game action sets and perfect monitoring. An important example of a dynamic game that violates perfect monitoring and therefore falls outside this framework is Fudenberg and Levine (1993). For a more comprehensive survey of belief learning, see Fudenberg and Levine (1998).

The earliest and perhaps best known example of belief learning is the *best response dynamics* of Cournot (1838). In Cournot's model, each player predicts that her opponent will repeat next period whatever action the opponent chose in the previous period.

The most studied belief learning model is *Fictitious play*, Brown (1951). In fictitious play, each player predicts that the probability that her opponent will play, say, L next period, is a weighted sum of an initial probability on L and the frequency with which L has been chosen to date. The weight on the frequency is $t/(t+k)$ where t is the number of periods thus far and $k > 0$ is a parameter; the larger is k, the more periods the initial probability significantly affects forecasting.

The remainder of this article discusses four topics: (1) belief learning versus Bayesian learning, (2) convergence to equilibrium, (3) special issues in games with payoff types, and (4) sensible beliefs.

Belief learning versus Bayesian learning. Recall that, in a repeated game, a behavior strategy gives, for every history, a probability over the player's stage game actions next period. In a Bayesian model, each player chooses a behavior strategy that best responds to a *belief*, a probability distribution over the opponent's behavior strategies.

Player 1's prediction rule about player 2 is mathematically identical to a behavior strategy for player 2. Thus, any belief learning model is equivalent to a Bayesian model in which the player optimizes with respect to a belief that places probability one on her prediction rule, now reinterpreted as the opponent's behavior strategy.

Conversely, any Bayesian model is equivalent to a belief learning model. Explicitly, for any belief over player 2's behavior strategies there is a degenerate belief, assigning probability one to a particular behavior strategy, that is equivalent in the sense that both beliefs induce the same distributions over play in the game, no matter what behavior strategy player 1 herself adopts. This is a form of Kuhn's Theorem; Kuhn (1964). I refer to the behavior strategy used in the degenerate belief as a *reduced form* of the original belief. Thus, any Bayesian model is equivalent to a Bayesian model in which each player's belief places probability one on the reduced form, and any such Bayesian model is equivalent to a belief learning model.

As an example, consider fictitious play. I focus on stage games with just two actions, L and R. By an i.i.d. strategy for player 2, I mean a behavior strategy in which player 2 plays L with probability q, independent of history. Thus, if $q = 1/2$ then player 2 always randomizes 50:50 between L and R. Fictitious play is equivalent to a degenerate Bayesian model in which each player places probability one on the fictitious play prediction rule, and one can show that this is equivalent in turn to a non-degenerate Bayesian model in which the belief is represented as a Beta distribution over q. The uniform distribution over q, for example, corresponds to taking the initial probability of L to be $1/2$ and the parameter k to be 2.

There is a related but distinct literature in which players optimize with respect to *stochastic* prediction rules. In some cases (e.g., Foster and Young (2003)), these models have a quasi-Bayesian interpretation: most of the time, players optimize with respect to fixed prediction rules, as in a Bayesian model, but occasionally players switch to new prediction rules, implicitly abandoning their priors.

Convergence to Equilibrium. Within the belief learning literature, the investigation of convergence to equilibrium play splits into two branches. One branch investigates convergence within the context of specific classes of belief learning models. The best response dynamics, for example, converge to equilibrium if the stage game is solvable by the iterated deletion of strictly dominated strategies. See Bernheim (1984) and, for a more general class of models, Milgrom and Roberts (1991). For an ε optimizing variant of fictitious play, convergence to approximate equilibrium play obtains for all zero sum games, all games with an interior ESS, and all common interest games, in addition to all games that are strict dominance solvable, with the approximation closer the smaller is ε. Somewhat weaker convergence results are available for supermodular games. These claims follow from results in Hofbauer and Sandholm (2002).

In either the best response dynamics or ε fictitious play, convergence is to repeated play of a single stage game Nash equilibrium; in the case of ε fictitious play, this equilibrium may be mixed. There is a large body of work on convergence that is weaker than what I am considering here. In particular, there has been much work on convergence of the empirical marginal or joint distributions. For mixed strategy equilibrium, it is possible for empirical distributions to converge to equilibrium even though play does not resemble repeated equilibrium play: play may exhibit obvious cycles, for example. The study of convergence to equilibrium play is relatively recent and was catalyzed by Fudenberg and Kreps (1993).

Hart and Mas-Colell (2003) and Hart and Mas-Colell (2004) (hereafter HM) study convergence to equilibrium play in learning models, including but not limited to belief learning models, that are (a) decoupled, meaning that player 1's behavior does not depend directly on player 2's stage game payoffs, and (b) satisfy a memory bound. They find that universal convergence is impossible for any such model: for any such model there exist stage games for which play fails to converge to equilibrium play. For a (continuous time) learning dynamic that is decoupled but violates the

memory bound and exhibits convergence for all finite stage games, see Shamma and Arslan (2005).

The second branch of the literature, for which Kalai and Lehrer (1993a) (hereafter KL) is the central paper, takes a Bayesian perspective and asks what conditions on beliefs are sufficient to give convergence to equilibrium play. I find it helpful to characterize this literature in the following way. Say that a belief profile (giving a belief for each player) has the *learnable best response property* (LBR) if there is a profile of best response strategies (the LBR strategies) such that, if the LBR strategies are played, then each player learns to predict the play path.

A player *learns to predict the play path* if her prediction of next period's play is asymptotically as good as if she knew her opponent's behavior strategy. If the behavior strategies call for randomization then players accurately predict the distribution over next period's play rather than the realization of next period's play. For example, consider a 2×2 game in which player 1 has stage game actions T and B and player 2 has stage game actions L and R. If player 2 is randomizing 50:50 every period and player 1 learns to predict the path of play then for every ε there is a time, which depends on the realization of player 2's strategy, after which player 1's next period forecast puts the probability of L within ε of $1/2$. (This statement applies to a set of play paths that arises with probability one with respect to the underlying probability model; I gloss over this sort of complication both here and below.) For a more complicated example, suppose that in period t player 2 plays L with probability $1 - \alpha$, where α is the frequency that the players have played the profile (B, R). If player 1 learns to predict the play path then for any ε there is a time, which now depends on the realization of both players' strategies, after which player 1's next period forecast puts the probability of L within ε of $1 - \alpha$.

Naively, if LBR holds, and players are using their LBR strategies, then, in the continuation game, players are optimizing with respect to posterior beliefs that are asymptotically correct and so continuation behavior strategies should asymptotically be in equilibrium. This intuition is broadly correct but there are three qualifications.

First, in general, convergence is to Nash equilibrium play in the *repeated* game, not necessarily to repeated play of a single stage game equilibrium. If players are myopic (meaning that players optimize each period as though their discount factors were zero), then the set of equilibrium play paths comprise all possible sequences of stage game Nash equilibria, which is a very large set if the stage game has more than one equilibrium. If players are patient then the folk theorem implies that the set of possible equilibrium paths is typically even larger.

Second, convergence is to an equilibrium play path, not necessarily to an equilibrium of the repeated game. The issue is that LBR implies accurate forecasting only along the play path. A player's predictions about how her opponent would respond to deviations may be grossly in error, forever. Therefore, posterior beliefs need *not* be asymptotically correct and, unless players are myopic, continuation behavior strategies need *not* be asymptotically in equilibrium. Kalai and Lehrer

(1993b) shows that behavior strategies can be doctored at information sets off the play path so that the modified behavior strategies are asymptotically in equilibrium yet still generate the same play path. This implies that the play path of the original strategy profile was asymptotically an equilibrium play path.

Third, the exact sense in which play converges to equilibrium play depends on the strength of learning. See KL and also Sandroni (1998).

KL shows that a strong form of LBR holds if beliefs satisfy an absolute continuity condition: each player assigns positive probability to any (measurable) set of play paths that has positive probability given the players' actual strategies. A sufficient condition for this is that each player assigns positive, even if extremely low, probability to her opponent's actual strategy, a condition that KL call *grain of truth*. Nyarko (1998) provides the appropriate generalization of absolute continuity for games with type space structures, including the games with payoff uncertainty discussed below.

There is no belief learning model that is decoupled (in the sense of HM, cited above) for which LBR holds for all stage games; one can show this by a direct diagonalization argument, without appealing to the non-convergence results in HM. In effect, LBR requires that players take each other's payoffs into account.

Games with Payoff Uncertainty. Suppose that, at the start of the repeated game, each player is privately informed of his or her stage game payoff function, which remains fixed throughout the course of the repeated game. Refer to player i's stage game payoff function as her *payoff type*. Assume that the joint distribution over payoff functions is independent (to avoid correlation issues that are not central to my discussion) and commonly known.

Each player can condition her behavior strategy in the repeated game on her realized payoff type. A mathematically correct way of representing this conditioning is via distributional strategies; see Milgrom and Weber (1985).

For any belief about player 2, now a probability distribution over player 2's distributional strategies, and given the probability distribution over player 2's payoff types, there is a behavior strategy for player 2 in the repeated game that is equivalent in the sense that it generates the same distribution over play paths. Again, this is essentially Kuhn's theorem. And again, I refer to this behavior strategy as a *reduced form*.

Say that a player *learns to predict the play path* if her forecast of next period's play is asymptotically as good as if she knew the reduced form of her opponent's distributional strategy. This definition specializes to the previous one if the distribution over types is degenerate. If distributional strategies are in equilibrium then, in effect, each player is optimizing with respect to a degenerate belief that puts probability one on her opponent's actual distributional strategy and in this case players trivially learn to predict the path of play.

One can define LBR for distributional strategies and, much as in the payoff certainty case, one can show that LBR implies convergence to equilibrium play in

the repeated game with payoff types. More interestingly, there is a sense in which play converges to equilibrium play of the *realized* repeated game – the repeated game determined by the realized type profile. The central paper is Jordan (1991). Other important papers include KL (cited above), Jordan (1995), Nyarko (1998), and Jackson and Kalai (1999) (which studies recurring rather than repeated games).

Suppose first that the realized type profile has positive probability. In this case, if a player learns to predict the play path then, as shown by KL, her forecast is asymptotically as good as if she knew both her opponent's distributional strategy *and* her opponent's realized type. LBR then implies that actual play, meaning the play generated by the realized behavior strategies, converges to equilibrium play of the realized repeated game. For example, suppose that the type profile for matching pennies gets positive probability. In the unique equilibrium of repeated matching pennies, players randomize 50:50 in every period. Therefore, LBR implies that if the matching pennies type profile is realized then each player's behavior strategy involves 50:50 randomization asymptotically.

In contrast, if the distribution over types admits a continuous density, so that no type profile receives positive probability, then the form of convergence is more subtle. Consider an outside observer who knows the profile of distributional strategies but *not* the realized type profile. For either a discrete or a continuous type distribution, LBR implies that this outside observer's posterior over play paths converges to that of an equilibrium of the realized repeated game. This implies that to an observer who knows the realized payoff types but *not* the distributional strategies, realized play looks asymptotically like equilibrium play.

If the type distibution is continuous, however, actual play (again meaning play generated by the realized behavior strategies) may not converge to equilibrium play of the realized repeated game. Indeed, if the realized stage game is like matching pennies, with a unique and fully mixed equilibrium, and if players optimize (rather than ε optimize) then actual play cannot converge to equilibrium play asymptotically, even if the distributional strategies constitute an equilibrium of the type space game. See Foster and Young (2001). Instead, convergence involves a form of purification in the sense of Harsanyi (1973), a point that has been emphasized by Nyarko (1998) and Jackson and Kalai (1999). For simplicity, suppose that players are myopic. Suppose further that LBR holds and that the realized stage game has a unique and fully mixed equilibrium. With these assumptions, the unique equilibrium of the realized repeated game calls for repeated play of the stage game equilibrium. It is not hard to show, in contrast, that optimization in the type space game calls for each player to play a pure strategy as a function of her realized type. In the type space game, therefore, actual play is pure but it looks random to an opponent who knows the distributional strategy but not the realized type, or to an outside observer who knows the realized type but not the distributional strategy. As play proceeds, each player in effect learns more about her opponent's realized type, but (in contrast to the case in which the realized type profile gets positive probability) never enough

to zero in on her opponent's actual play.

Sensible Beliefs. A number of papers investigate classes of prediction rules that are sensible in that they exhibit desirable properties, such as the ability to detect certain kinds of patterns in opponent behavior. See Aoyagi (1996), Fudenberg and Levine (1995), Fudenberg and Levine (1999), and Sandroni (2000).

Nachbar (2005) instead studies the issue of sensible beliefs from a Bayesian perspective. For simplicity, focus on learning models with known payoffs. Fix a belief profile, fix a subset of behavior strategies for each player, and consider the following criteria for these subsets.

- *Learnability* – given beliefs, if players play a strategy profile drawn from these subsets then they learn to predict the play path.

- *CSP* – a diversity or richness condition. Informally (the formal statement is tedious), CSP requires that if a behavior strategy is included in one of the strategy subsets then certain variations on that strategy must be included as well. CSP is satisfied automatically if the strategy subsets consist of all strategies satisfying a standard complexity bound, the same bound for both players. Thus CSP holds if the subsets consist of all strategies with k-period memory, or all strategies that are automaton implementable, or all strategies that are Turing implementable, and so on.

- *Consistency* – each player's subset contains a best response to her belief.

The motivating idea is that beliefs that are probability distributions over strategy subsets satisfying learnability, CSP, and consistency are sensible beliefs, or at least are candidates for being considered sensible. Nachbar (2005) studies whether any such beliefs exist.

Consider, for example, the Bayesian interpretation of fictitious play in which beliefs are probability distributions over the i.i.d. strategies. The set of i.i.d. strategies satisfies learnability and CSP. But for any stage game in which neither player has a weakly dominant action, the i.i.d. strategies violate consistency: any player who is optimizing will not be playing i.i.d.

Nachbar (2005) shows that this observation about Bayesian fictitious play extends to all Bayesian learning models. For large classes of repeated games, for *any* belief profile there are *no* strategy sets that simultaneously satisfy learnability, CSP, and consistency. Thus for example, if each player believes the other is playing a strategy that has a k-period memory then one can show that learnability and CSP hold but consistency fails: best responding in this setting requires using a strategy with a memory of more than k periods. The impossibility result generalizes to ε optimization and ε consistency, for ε sufficiently small. The result also generalizes to games with payoff uncertainty (with learnability, CSP, and consistency now defined in terms of distributional strategies); see Nachbar (2002).

I conclude with four remarks. First, since the set of all strategies always satisfies CSP and consistency, it follows that the set of all strategies is not learnable for *any* beliefs: for any belief profile there is a strategy profile that the players will not learn to predict. This can be also be shown directly by a diagonalization argument along the lines of Oakes (1985) and Dawid (1985). The impossibility result of Nachbar (2005) can be viewed as a game theoretic version of Dawid (1985). For a description of what sets *are* learnable, see Noguchi (2005).

Second, if one constructs a Bayesian learning model satisfying learnability and consistency then LBR holds and, if players play their LBR strategies, play converges to equilibrium play. This identifies a potentially attractive class of Bayesian models in which convergence obtains. The impossibility result says, however, that if learnability and consistency hold then player beliefs must be partially equilibrated in the sense of, in effect, excluding strategies required by CSP.

Third, consistency is not *necessary* for LBR or convergence. For example, for many stage games, variants of fictitious play satisfy LBR and converge even though these learning models are inconsistent. The impossibility result is a statement about the ability to construct Bayesian models with certain properties; it is not a statement about convergence *per se*.

Lastly, it may be that learnability, CSP, and consistency are too strong to be taken as a necessary for beliefs to be sensible. It is an open question whether one can construct Bayesian models satisfying conditions that are weaker but still strong enough to be interesting.

Elsewhere in Palgrave: Repeated Games. Adaptive Learning.

References

Aoyagi, M. (1996): "Evolution of Beliefs and the Nash Equilibrium of Normal Form Games," *Journal of Economic Theory*, 70, 444–469.

Bernheim, B. D. (1984): "Rationalizable Strategic Behavior," *Econometrica*, 52(4), 1007–1028.

Brown, G. W. (1951): "Iterative Solutions of Games By Fictitious Play," in *Activity Analysis of Production and Allocation*, ed. by T. J. Koopmans, pp. 374–376. John Wiley, New York.

Cournot, A. (1838): *Researches into the Mathematical Principles of the Theory of Wealth*. Kelley, New York, Translation from the French by Nathaniel T. Bacon. Translation publication date: 1960.

Dawid, A. P. (1985): "The Impossibility of Inductive Inference," *Journal of the American Statistical Association*, 80(390), 340–341.

Foster, D., and P. Young (2001): "On the Impossibility of Predicting the Behavior of Rational Agents," *Proceedings of the National Academy of Sciences*, 98, 12848–12853.

——— (2003): "Learning, Hypothesis Testing, and Nash Equilibrium," *Games and Economic Behavior*, 45, 73–96.

Fudenberg, D., and D. Kreps (1993): "Learning Mixed Equilibria," *Games and Economic Behavior*, 5(3), 320–367.

Fudenberg, D., and D. Levine (1993): "Steady State Learning and Nash Equilibrium," *Econometrica*, 61(3), 547–574.

——— (1995): "Universal Consistency and Cautious Fictitious Play," *Journal of Economic Dynamics and Control*, 19, 1065–1089.

——— (1998): *Theory of Learning in Games*. MIT Press, Cambridge, MA.

——— (1999): "Conditional Universal Consistency," *Games and Economic Behavior*, 29, 104–130.

Harsanyi, J. (1973): "Games with Randomly Disturbed Payoffs: A New Rationale for Mixed-Strategy Eequilibrium Points," *International Journal of Game Theory*, 2, 1–23.

Hart, S., and A. Mas-Colell (2003): "Uncoupled Dynamics do Not Lead to Nash Equilibrium," *American Economic Review*, 93, 1830–1836.

——— (2004): "Stochastic Uncoupled Dynamics and Nash Equilibrium," The Hebrew University of Jerusalem.

Hofbauer, J., and W. Sandholm (2002): "On the Global Convergence of Stochastic Fictitious Play," *Econometrica*, 70(6), 2265–2294.

Jackson, M., and E. Kalai (1999): "False Reputation in a Society of Players," *Journal of Economic Theory*, 88(1), 40–59.

Jordan, J. S. (1991): "Bayesian Learning in Normal Form Games," *Games and Economic Behavior*, 3, 60–81.

——— (1995): "Bayesian Learning in Repeated Games," *Games and Economic Behavior*, 9, 8–20.

Kalai, E., and E. Lehrer (1993a): "Rational Learning Leads to Nash Equilibrium," *Econometrica*, 61(5), 1019–1045.

——— (1993b): "Subjective Equilibrium in Repeated Games," *Econometrica*, 61(5), 1231–1240.

Kuhn, H. W. (1964): "Extensive Games and the Problem of Information," in *Contributions to the Theory of Games, Volume II*, ed. by M. Dresher, L. S. Shapley, and A. W. Tucker, pp. 193–216. Princeton University Press, Annals of Mathematics Studies, 28.

Milgrom, P., and J. Roberts (1991): "Adaptive and Sophisticated Learning in Repeated Normal Form Games," *Games and Economic Behavior*, 3, 82–100.

Milgrom, P., and R. Weber (1985): "Distributional Strategies for Games with Incomplete Information," *Mathematics of Operations Research*, 10, 619–632.

Nachbar, J. H. (2002): "Basic Non-Cooperative Game Theory: The Unrated Version," Washington University, St. Louis.

——— (2005): "Beliefs in Repeated Games," *Econometrica*, 73, 459–480.

Noguchi, Y. (2005): "Merging with a Set of Probability Measures: A Characterization," Kanto Gakuin University.

Nyarko, Y. (1998): "Bayesian Learning and Convergence to Nash Equilibria Without Common Priors," *Economic Theory*, 11(3), 643–655.

Oakes, D. (1985): "Self-Calibrating Priors Do Not Exist," *Journal of the American Statistical Association*, 80(390), 339.

Sandroni, A. (1998): "Necessary and sufficient conditions for convergence to Nash equilibrium: the almost absolute continuity hypothesis," *Games and Economic Behavior*, 22, 121147.

——— (2000): "Reciprocity and Cooperation in Repeated Coordination Games: The Principled-Player Approach," *Games and Economic Behavior*, 32(2), 157–182.

Shamma, J., and G. Arslan (2005): "Dynamic Fictitious Play, Dynamic Gradient Play, and Distributed Convergence to Nash Equilibria," *Transactions on Automatic Control*, pp. 312–327.