Strategic Learning and Its Limits by H. Peyton Young*

William H. Sandholm†‡

November 6, 2007

While the cardinal role of game theory in economic analysis is no longer challenged, a fundamental question about the standard methods of game-theoretic prediction remains: why should we expect players' behavior to conform with some notion of equilibrium play? The traditional approach to this problem is to ask whether agents can arrive at equilibrium through a purely introspective process. By now, various epistemic conditions ensuring equilibrium play are available, but these conditions seem overly demanding for most applications.

The other leading approach to justifying equilibrium play is to ask whether agents can learn to behave as equilibrium concepts predict over the course of a repeated interaction. One can divide research on this question into two main strands. Models from *evolutionary game theory* study the behavior of large populations of agents whose decision rules condition on current strategic conditions, and ask whether or not the agents' aggregate behavior will come to resemble a Nash equilibrium.[1] Alternatively, models of *learning in games* consider whether a small group of players, one for each role in the game, ultimately will behave in accordance with equilibrium predictions.

The theory of learning in games is the subject of Peyton Young's delightful new monograph, *Strategic Learning and Its Limits*. In this slim volume, Young manages not only to survey the major developments in the field over the past fifteen years, but also to present in a clear and concise fashion some of the mathematical techniques that underpin this area of research. The learning literature has reached a level of maturity at which a taking of stock is in order, and Young provides a masterly synthesis of what has been achieved to

*xi + 165 pages, Oxford University Press, 2004.

†Department of Economics, University of Wisconsin, 1180 Observatory Drive, Madison, WI 53706, whs@ssc.wisc.edu, http://www.ssc.wisc.edu/~whs.

‡I thank Dean Foster, Drew Fudenberg, and Sergiu Hart for helpful comments, and gratefully acknowledge financial support under NSF Grant SES-0617753.

[1]See Hofbauer and Sigmund (2003) and Sandholm (2007) for recent surveys, and Sandholm (2008) for a fuller current treatment.

date. Anyone seeking a brief but substantive overview of what this literature has taught us will read this book with pleasure.

Chapters 1 through 6 of *Strategic Learning and Its Limits* offer a thorough presentation of models of *heuristic learning*, in which agents use simple myopic rules to learn to play repeated normal form games.[2] Young classifies these models in two groups according to the sorts of information agents consider when making decisions. Models in which players concentrate on their own *past payoffs* can be exceedingly simple: indeed, the players need not even be aware that they are playing a game. In contrast, when players aim to predict opponents' future behavior, it is natural for them to attend to opponents' *past play*. Past play models are often more sophisticated than past payoff models, as they require players to possess (possibly naive) theories about how opponents act.

To begin, Young considers *reinforcement learning*, the simplest sort of past payoff model. Here every player assigns each of his strategies a positive weight, a weight that is increased each time the strategy is played; his choices in each period are random, with probabilities determined by the strategy weights. While reinforcement learning schemes are optimal in stationary environments, there are few classes of games in which they always converge to equilibrium. But Young notes that these schemes share two basic properties with more successful learning procedures: namely, probabilistic choice and "sluggish adaptation" (that is, inertia in the updating of choice probabilities).

In Chapters 2 through 4, Young studies more sophisticated past payoff models based on the elimination of regret. Consider a single agent who faces a repeated decision problem: in each period, the agent makes a choice from a finite set of actions A and then obtains a payoff. As of period t, the agent's *regret* for (not having chosen) action a is defined as the difference between two terms: the average payoff he would have obtained had he chosen a in periods 1 through t, and the average payoff he actually obtained during those periods. A strategy for this repeated decision problem satisfies *no regret* (or is *consistent*) if it ensures that for any sequence of payoff realizations, the agent's regret for each of his actions becomes nonpositive as t approaches infinity. Returning to the context of repeated games, it is easy to show that if each player follows a no-regret strategy, then the time average of play must converge to the set of *coarse correlated equilibria*: this concept is the generalization of correlated equilibrium obtained when players' decisions about whether to follow the proposed correlated strategy are made at the ex ante stage.

[2] A more narrowly focused overview of heuristic learning can be found in Sergiu Hart's 2003 Walras-Bowley lecture (Hart (2005)). For an earlier book-length treatment of the learning literature, one that also covers models of learning in extensive form games, see Fudenberg and Levine (1998). For more technical presentation of models of adaptive learning and prediction with many pointers to the computer science and statistics literatures, see Cesa-Bianchi and Lugosi (2006).

It is natural to ask next whether one can prove a stronger convergence result, under which average behavior converges the set of correlated equilibria. To do so, one must replace the no-regret criterion with something more demanding, so that agents' behavior will satisfy not only the ex ante notion of optimality posited by coarse correlated equilibrium, but also the interim notion posited by correlated equilibrium. This stronger criterion, called *conditional no regret* (or *conditional consistency*), requires that for any sequence of payoff realizations, the following statement is true: for each action a that the agent played with nonnegligible frequency, the agent would not have been better off having always played an alternative action a' in place of a. By construction, the connection between conditional no regret and correlated equilibrium is precisely analogous to that between no regret and coarse correlated equilibrium.

After finishing his treatment of no regret models, Young turns his attention to models of learning in which agents focus on past play, beginning in Chapter 5 with calibrated learning. Consider an agent who is about to view an infinite sequence of observations from the finite set O. Before each observation appears, the agent makes a forecast about its realization, a forecast that takes the form of a probability distribution on O. Roughly speaking, an agent's *forecast* over an infinite sequence of observations is *calibrated* if the following statement is true: if the agent's forecast is p in a nonnegligible number of periods, then the empirical distribution of outcomes in those periods is close to p. More demandingly, an agent's *forecasting procedure* is *calibrated* if it generates calibrated forecasts for any possible sequence of observations. If all players in a repeated game choose myopic best responses to calibrated forecasts, their time-averaged behavior converges to the set of correlated equilibria.

Young's presentation of these ideas, including both the early analyses of Hannan (1957) and Blackwell (1956a,b) and the more recent ones of Foster and Vohra (1993, 1997, 1998), Fudenberg and Levine (1995, 1999) and Hart and Mas-Colell (2000, 2001), is clear, concise, and complete. The cornerstone of Young's presentation is *Blackwell's Approachability Theorem*; this generalization of the Minmax Theorem to games with vector-valued payoffs is of cardinal importance throughout the theory of heuristic learning in games.

In Chapter 6, Young considers models of *fictitious play*. In the basic model of Brown (1951), each player selects myopic best responses to his beliefs about his opponents' strategies, where these beliefs specify that each opponent will play the mixed action defined by his time-averaged behavior. If one assumes that payoffs are subjected to small random shocks, as in the *stochastic fictitious play* model of Fudenberg and Kreps (1993), then period-by-period behavior becomes stochastic, introducing the possibilities of both convergence of period-by-period behavior and elimination of regret.

The last two chapters of *Strategic Learning and Its Limits* offer tastes of recent work in other branches of the learning literature. In Chapter 7, Young considers *rational learning*, the branch that comes closest to traditional economic modeling. Rational learning models ask how Bayesian rational players—that is, players who form prior beliefs, update them in the face of past experience using Bayes' rule, and choose strategies in a dynamically optimal fashion—might learn to play a game. Young reviews the two central results in this literature, the Nash convergence theorem of Kalai and Lehrer (1993) and the impossibility theorem of Nachbar (1997), and then presents more recent work on the impossibility of rational learning of mixed equilibria. In Chapter 8, Young turns to models of *random search with independent verification*. Here each agent begins with a hypothesis about the stage game mixed strategy profile used by his opponents. The agent plays a myopic best response to this hypothesis until the evidence provided by past play leads him to reject it, in which case another hypothesis is formed at random. Such models allow agents to coordinate on stage game Nash equilibria in a high proportion of periods over long enough time spans. But as the initial period of play involves a random exploration of the set of mixed strategy profiles, random search models have limited predictive power when the time span of interest is of moderate duration.

While Young classifies learning models in terms of their reliance on past payoffs or past play, it is more fruitful to divide his book somewhat differently, into learning models that can be analyzed using approachability theory, and other models of learning games. Given Young's aim of finding learning models that converge to equilibrium in all games, this division is natural. The greatest successes in attaining this goal rely on repeated game strategies derived from worst-case analyses of repeated multicriterion decision problems; approachability theory allows one to conduct these analyses in an astonishingly uncomplicated way.

In his early chapters, where he interweaves the theories of approachability and learning in games, Young achieves an impressive balance between readability and precision, and I expect this exposition to become a standard reference on its subject. Given their page counts, Young's treatments of other models in the later chapters cannot provide a similar level of detail. Therefore, rather than surveying all of the relevant models in a cursory way, Young wisely chooses to present a few representative models in depth. In so doing, he is able to provide the reader with an appreciation for a wide swath of the literature. All told, *Strategic Learning and Its Limits* offers an exemplary introduction to recent work on learning in games.

References

Blackwell, D. (1956a). An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6:1–8.

Blackwell, D. (1956b). Controlled random walks. In Gerretsen, J. C. H. and Groot, J. D., editors, *Proceedings of the International Conference of Mathematicians 1954*, volume 3, pages 336–338. North Holland, Amsterdam.

Brown, G. W. (1951). Iterative solutions of games by fictitious play. In Koopmans, T. C. et al., editors, *Activity Analysis of Production and Allocation*, pages 374–376. Wiley, New York.

Cesa-Bianchi, N. and Lugosi, G. (2006). *Prediction, Learning, and Games*. Cambridge University Press, Cambridge.

Foster, D. P. and Vohra, R. (1993). A randomized rule for selecting forecasts. *Operations Research*, 41:704–709.

Foster, D. P. and Vohra, R. (1997). Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, 21:40–55.

Foster, D. P. and Vohra, R. (1998). Asymptotic calibration. *Biometrika*, 85:379–390.

Fudenberg, D. and Kreps, D. M. (1993). Learning mixed equilibria. *Games and Economic Behavior*, 5:320–367.

Fudenberg, D. and Levine, D. K. (1995). Consistency and cautious fictitious play. *Journal of Economic Dynamics and Control*, 19:1065–1089.

Fudenberg, D. and Levine, D. K. (1998). *Theory of Learning in Games*. MIT Press, Cambridge.

Fudenberg, D. and Levine, D. K. (1999). Conditional universal consistency. *Games and Economic Behavior*, 29:104–130.

Hannan, J. (1957). Approximation to Bayes risk in repeated play. In Dresher, M., Tucker, A. W., and Wolfe, P., editors, *Contributions to the Theory of Games III*, volume 39 of *Annals of Mathematics Studies*, pages 97–139. Princeton University Press, Princeton.

Hart, S. (2005). Adaptive heuristics. *Econometrica*, 73:1401–1430.

Hart, S. and Mas-Colell, A. (2000). A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68:1127–1150.

Hart, S. and Mas-Colell, A. (2001). A general class of adaptive strategies. *Journal of Economic Theory*, 98:26–54.

Hofbauer, J. and Sigmund, K. (2003). Evolutionary game dynamics. *Bulletin of the American Mathematical Society (New Series)*, 40:479–519.

Kalai, E. and Lehrer, E. (1993). Rational learning leads to Nash equilibrium. *Econometrica*, 61:1019–1045.

Nachbar, J. H. (1997). Prediction, optimization, and learning in games. *Econometrica*, 65:275–309.

Sandholm, W. H. (2007). Evolutionary game theory. In Meyers, R. A. et al., editors, *Encyclopedia of Complexity and System Science*. Forthcoming, Springer, Heidelberg.

Sandholm, W. H. (2008). *Population Games and Evolutionary Dynamics*. Forthcoming, MIT Press, Cambridge.