An Economist's Perspective on Multi-Agent Learning

by Drew Fudenberg and David K. Levine

October 5, 2006

In their wide-ranging and provocative discussion, Shoham, Powers and Grenager (SPG) survey several large literatures from computer science and game theory, and identify five categories of questions about multi-agent learning (MAL) that these literatures seem to address. Their unified framework for interpreting models of MAL provides a useful bridge between the economics and AI communities. To reinforce that bridge, we would like to comment on the relevance and relative importance of the five categories for economics, emphasize some modeling issues that SPG do not highlight, and to correct what seem to us to be some minor imprecisions in their discussion.

I. Five Categories of Research on Multi-Agent Learning

The five SPG categories of MAL research are computational, descriptive (defined as "how natural agents learn in the context of other learners"), "normative" (defined as the study of whether rules are in equilibrium with each other), prescriptive cooperative, and prescriptive non-cooperative. Not surprisingly, computational issues are of more central concern to computer scientists than economists.

From the perspective of economists and other social scientists, description and the related goal of *prediction* are the most central objective of game theory and hence of the study of learning in games. By "prediction" here we mean not only the narrow issue of matching the data on period-by-period learning in experiments,[1] but also the larger and more important question of when and whether we should expect play in a given game to resemble an equilibrium, and the related questions of what to expect when the learning

[1] Since many plausible learning models behave in roughly similar ways in simple settings, it can be difficult to distinguish them empirically; Salmon [2001] argues that the prevailing tests are too weak to do so with the sort of data that is typically available.

process does not converge, and if it does converge to an equilibrium, is any particular subset of the equilibria selected?

The “normative” question of which rules are in equilibrium with one another has not been of much interest in economics, and indeed as SPG note this question has been explicitly critiqued by Fudenberg and Kreps [1993] among others. Since SPG do not elaborate that critique, we will do it here: From the economist’s viewpoint, the main agenda of the learning in games literature is determining when and whether we should expect play in a given game to look like an equilibrium, so assuming that the learning rules start out in an equilibrium in some “learning rule game” begs the question of why this should be the case. In computer science terminology, equilibrium corresponds to a set of joint restrictions on the initial state of the various learning rules, and there is no reason to think that the system would be initialized in this way.

Like computational issues, prescriptive cooperative learning has not received a great deal of attention from economists, but the theory of mechanism design might well benefit from results in this direction. Finally, prescriptive non-cooperative learning models are of interest to economists for two reasons: These models may be useful for giving people advice about how to play in games, and they may also help us make better predictions. That is, because learning rules for games have evolved over a long period of time, there is some reason to think that rules that are good rules from a prescriptive point of view may in fact be good from a descriptive point of view. This highlights the fact that the five different categories may complement each other as well as representing distinct direction.

II. Modeling Issues

The nature of multi-agent learning depends not only on the strategies and payoff functions of the game, but also on the context in which a game takes place. This context includes whether or not players observe one another’s actions each period, whether the players have played this or a similar game in the past, and on what other information they may have that help them understand their own incentives and predict their opponents’ motivations.

One example of the role of context is the difference between an environment in which a fixed pair of agents plays each other in every period, and environments with a large population of roughly similar agents. When the same two agents play each other every period, they may try to influence each other's future play, so that the natural benchmark solution concept is that of equilibrium in the repeated game. However, the repeated game equilibrium is not applicable in some environments with a large population of agents.

To see this, consider first the case of games with anonymous random matching: Each period, all players are randomly matched to play a one-shot simultaneous-move game, and at the end of each round each player observes only the play in his own match. The way a player acts today will influence the way his current opponent plays tomorrow, but if the population is sufficiently large compared to the discount factor then the player is unlikely to be matched with his current opponent or anyone who has met the current opponent for a long time. An important implication of this for learning theory is that myopic play is approximately optimal if the population is finite but sufficiently large. [2]

As a second example, consider a model with aggregate statistics. Again, each period, all players are randomly matched to play a one-shot game. At the end of the round, the population aggregates are announced. If the population is large, each player has little influence on the population aggregates, and consequently little influence on future play, so players have no reason to depart from myopic play. Some, but too few, experiments use this design.

Each of the above environments has the conceptual advantage that we can suppose that players are only trying to learn their optimal strategy, and not to influence the future course of the overall system. In these environments simple behavior rules like smooth fictitious play can be justified, and there is no need to consider the sort of "teaching" that SPG mention in Sections 3 and 4.3. However, many of the game theory papers that have considered interacting learning rules have simplified by assuming either that there is a single agent on each side, or that the entire population has the same information and beliefs. In most of the intended applications of the theory it seems likely

[2] Note also that myopic play need not be optimal for a rational player even in a large population if the discount factor is close enough to 1 (Ellison [1995]).

that agents have heterogeneous beliefs, and it is very important for the literature to take this into account.[3]

The role of populations is also important in understanding evolutionary models. Evolutionary models are population models, but the converse is not generally true. Evolutionary models are what we may more broadly describe as aggregate models. An aggregate model starts with a description of aggregate behavior of a population of agents. An example of this is the "replicator dynamic" mentioned in SPG. Here a fraction of the population playing a strategy increases if the utility received from that strategy is above average. There are two main reasons that economists are interested in the replicator and related models. One is that, as shown by Borgers and Sarin [2000], the replicator dynamic can approximate the evolution of mixed strategies used by human agents who follow a particular sort of reinforcement learning. A second is the possibility that learning is "social" in the sense of players copying the successes of other players. Examples of this are Binmore and Samuelson [1995], Bjonerstedt and Weibull [1995], or Schlag [1994]. There is also a class of non-equilibrium models of "social learning," where players are trying to learn what technology, crop or brand is best, such as Kirman [1994] and Ellison and Fudenberg [1993 ,1995]. These models typically lead to aggregate behavior that is "replicator-like", that is, shares some of the qualitative properties of the replicator dynamic; a series of papers starting with Samuelson-Zhang [1992] has investigated what can be said about the long-run behavior of such systems.

While there is a dizzying array of possibilities, one fortunate fact is that often quite different contexts lead to similar mathematics. An important example is the connection between population partial best-response dynamics, where a fraction of the population adopts the best response to the current population play, and fictitious play where players adopt a best response to historical averages. The path of play in population partial best-response is asymptotically the same as that of fictitious play with time measured on a logarithmic rather than linear scale.

[3] Fudenberg and Levine [1993] allows heterogeneous beliefs in a model with a continuum of agents but only study the steady states, Hopkins [1999] models the dynamics of a system with a continuum of fictitious-play learners who have heterogeneous beliefs. The next step is to extend the analysis of heterogeneous beliefs to the dynamics of systems with large but finite populations.

Another modeling point relates to the distinction between rational and irrational play on the one hand and the distinction between learning contexts on the other. As noted by Fudenberg and Kreps [1993], fictitious play corresponds to rational Bayesian learning by an agent who is convinced that the opposing player's actions are drawn from a fixed but unknown distribution, provided that the learner's prior has a specific functional form. Fudenberg and Levine [1993] extend this rational Bayesian approach to agents who are learning to play a general extensive-form game in the setting of anonymous random matching. Thus, in contrast to the discussion in Section 4.1.1 of SGP, we see the distinction between fictitious play on the one hand and the analysis of Kalai and Lehrer [1993] on the other as not being about "rational" vs. some other sort of play but on the environments in which the rules are analyzed.

We should emphasize that we agree with SPG about much of what they say. Clearly there has been a rush to equilibrium., and we share their concern about the "default blind adoption of equilibria as the driving concept in complex games." It would be wonderful if there were a way to use field data to understand when equilibrium analysis is justified, but once one leaves the controlled laboratory environment it seems very difficult to identify equilibrium play. If one is certain that payoffs are constant over time, then any variation in play at all shows that agents are not playing a static equilibrium, but this leaves open both the possibility that payoff functions vary and that play corresponds to the equilibrium of some dynamic game. So what is needed is a plausible set of identifying restrictions on the nature of payoffs and strategies, and a model of non-equilibrium play that can be econometrically implemented when the actual payoff functions of the players are unknown to the analyst.[4]

SPG also say that "…in the context of complex games, so-called "bounded rationality", or the deviation from the ideal behavior of omniscient agents, is not an esoteric phenomenon to be brushed aside." We strongly agree with the idea that deviations from equilibrium should be taken seriously, but we would like to point out that most game theorists do not identify equilibrium with the "ideal behavior of omniscient

[4] See Fudenberg [2006] for a discussion of some of the related work and issues.

agents," and that a long literature emphasizes that common knowledge of rationality is not sufficient to produce equilibrium outcomes.[5]

We also strongly support SPG's statement that it is pointless to analyze the convergence properties of arbitrary learning rules. Moreover, it does not seem sensible to make convergence to equilibrium the main factor that is used to justify interest in a given rule. Instead, one should have some reason to think that the rules are a plausible approximation of behavior in a case of interest. This was, for example, the motivation for our own consideration of properties such as universal consistency. The consideration of this "prescriptive non-cooperative" property suggested to us that smoothed, as opposed to exact, fictitious play would do a better job of description and prediction.

Let us briefly comment on a few of the details of SPG. First, the discussion of rock-paper-scissors may not be ideal, as there is no reason to think that the ex-post "winners" will be playing an equilibrium strategy. The relevant question is whether the population as a whole resembles equilibrium. Consider a population all playing pure strategies, such as "play the opposite of what the opponent did last period." The actual payoff of any agent will depend on whom he is matched with. In order for this not to be an equilibrium, we would need to identify a strategy that in expectation does better than the proposed equilibrium against the prevailing population distribution.

SPG also identify several different types of learning models. It is worth emphasizing in particular that both no-regret and smooth fictitious play models are closely connected, in the sense of having similar asymptotic properties, and can both be adapted to situations in which a player observes only their own payoffs. From the perspective of economists, q-learning and other procedures that use generalizations of reinforcement learning to estimate value functions in environments with a state variable have not been well-studied. The issue here, as with other sorts of reinforcement learning models, is in deciding "what is re-enforced," that is, what the state variable is. This is the analog of the problem of specifying a family of prior distributions for a Bayesian learner. Once the map from state variables in a game to implicit models of opponents' play is better understood, the results on q-learning may be very useful for economists. It may be

[5] Von Neumann and Morgenstern did advance this interpretation of non-cooperative game theory, but as they realized, that interpretation is only helpful in two-player zero-sum games; Nash explicitly realized that

that there considering q-learning in the multiple-agent case where players simultaneously try to calculate value function will lead to important new insights.

This brings us to our final point, the distinction between passive and active learning. Passive learning is learning by observing whatever happens to be observable. Active learning, on the other hand, involves actively trying things to discover their consequences. Most work on active learning studies the single-agent case, as in the two-armed bandit problem, but there are modest literatures in economics about active multi-agent learning in both the equilibrium (Bolton and Harris [1999], Cripps, Keller and Rady [2005])) and non-equilibrium setting (Fudenberg and Levine [1993], Fudenberg and Kreps [1995], Dubey and Haimanko [2004], Jehiel and Samet [2005].) One issue is that active learning may fail or be inadequate. In non-equilibrium learning models, this leads to equilibrium concepts such as self-confirming equilibrium, demanding a high degree of knowledge of the equilibrium path, which is necessarily observed, but little knowledge of off-the equilibrium path, which is not. This weakening of Nash equilibrium has yielded fruitful insights to economist on a range of descriptive issues, including understanding play in the experimental context.

in general games equilibrium requires some way to coordinate the expectations of the players.

REFERENCES

Benaim, M. and M. Hirsch [1999] “Mixed Equilibria Arising from Fictitious Play in Perturbed Games,” *Games and Economic Behavior* 29, 36-72

Binmore, K. and L. Samuelson [1992]: "Evolutionary Stability in Repeated Games Played by Finite Automata," *Journal of Economic Theory*, 57: 278-305

Bolton, P. and C. Harris "Strategic Experimentation", *Econometrica*, 67, 1999, 349-374.

Börgers, T. and R. Sarin [2000] “Naive Reinforcement Learning With Endogenous Aspirations,” *International Economic Review* 41: 921-950.

Bjornerstedt, J. and J. Weibull [1995]: "Nash Equilibrium and Evolution by Imitation," in *The Rational Foundations of Economic Behavior*, ed. K. Arrow et al, Macmillan: London

Camerer, C., and T.-H. HO (1999): “Experience-Weighted Attraction Learning in Normal Form Games,” *Econometrica*, 67: 827–874

Cripps, M., G. Keller, and S. Rady [2005] “Strategic Experimentation with Exponential Bandits *Econometrica* 73

Dubey, P. and O. Haimanko [2004] “Learning with Perfect Information,” *Games and Economic Behavior* 46[2], 304-324.

Ellison, G. [1995] “Cooperation in the Prisoner's Dilemma with Anonymous Random Matching,” *The Review of Economic Studies*

Ellison, G. and D. Fudenberg [1993] “Rules of Thumb for Social Learning” *Journal of Political Economy*, 101 , 612-643.

Ellison, G. and D. Fudenberg [1995] “Word of Mouth Communication and Social Learning” *Quarterly Journal of Economics*, 110 93-126.

Fudenberg, D. [2006] “Advancing on ‘Advances in Behavioral Economics,” *Journal of Economic Literature* 44: 604-711.

Fudenberg, D. and D. Kreps [1993] “Learning Mixed Equilibria,” *Games and Economic Behavior*, 5, 320-367.

Fudenberg, D. and D. Kreps [1996] “Learning in Extensive Form Games, II: Experimentation and Nash Equilibrium,” mimeo.

Fudenberg, D. and D. K. Levine [1993] "Steady State Learning and Nash Equilibrium," *Econometrica* 61, 547-573.

Hopkins, E. [1999] “Learning, Matching and Aggregation,” *Games and Economic Behavior* 26, 79-110.

Jehiel, P. and D. Samet [2005] “Learning to Play Games in Extensive Form by Valuation,” *Journal of Economic Theory* 124, 129-148.

Kalai, E. and E. Lehrer [1993] “Rational Learning Leads to Nash Equilibrium,” *Econometrica* 61, 1019-1045.

Salmon, T. C. [2001] “An Evaluation of Econometric Models of Adaptive Learning,” *Econometrica*, 69: 1597–1628

Samuelson, L. and J. Zhang [1992] “Evolutionary stability in Asymmetric games,” *Journal of Economic Theory*, 57:363-91.

Schlag, K. [1994]: "Why Imitate, and if so, How? Exploring a Model of Social Evolution," *Universitat Bonn*, pp b-296.