Evolution and Learning in Games: Overview*

William H. Sandholm†

January 19, 2007

1. Introduction

The theory of evolution and learning in games provides models of disequilibrium behavior in strategic settings. Much of the theory focuses on whether and when disequilibrium behavior will resolve in equilibrium play, and, if so, on predicting which equilibrium will be played. But the theory also offers techniques for characterizing perpetual disequilibrium play.

2. A Taxonomy

Models from *evolutionary game theory* consider the behavior of large populations in strategic environments. In the biological strand of the theory, agents are genetically programmed to play fixed actions, and changes in the population's composition are the result of natural selection and random mutations. In economic approaches to the theory, agents actively choose which actions to play using simple myopic rules, so that changes in aggregate behavior are the end result of many individual decisions. *Deterministic evolutionary dynamics*, usually taking the form of ordinary differential equations, are used to describe behavior over moderate time spans, while *stochastic evolutionary dynamics*, modeled using Markov processes, and more commonly employed to study behavior over very long time spans.

Models of *learning in games* focus on the behavior of small groups of players, one of whom fills each role in a repeated game. These models too can be partitioned into two

*I thank John Nachbar for a number of helpful conversations, and for sharing his expertise on coordinated Bayesian learning. Financial support under NSF Grants SES-0092145 and SES-0617753 is gratefully acknowledged.

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

categories. Models of *heuristic learning* (or *adaptive learning*) resemble evolutionary models, in that their players base their decisions on simple myopic rules.[1] In models of *coordinated Bayesian learning* (or *rational learning*) each player forms explicit beliefs about the repeated game strategies employed by other players, and plays a best response to those beliefs in each period. The latter models assume a degree of coordination of players' prior beliefs that is sufficient to ensure that play converges to Nash equilibrium.[2]

3. Evolutionary Game Theory

The roots of evolutionary game theory lie in mathematical biology. Maynard Smith and Price (1973) introduced the equilibrium notion of an *evolutionarily stable strategy* (or *ESS*) to capture the possible stable outcomes of a dynamic evolutionary process by way of a static definition. Later, Taylor and Jonker (1978) offered the *replicator dynamic* as an explicitly dynamic model of the natural selection process. The decade that followed saw an explosion of research on the replicator dynamic and related models of animal behavior, population ecology, and population genetics—see Hofbauer and Sigmund (1988).

In economics, evolutionary game theory studies the behavior of populations of strategically interacting agents who actively choose among the actions available to them. Agents decide when to switch actions and which action to choose next using simple myopic rules known as *revision protocols*.[3] A population of agents, a game, and a revision protocol together define a stochastic process—in particular, a Markov process—on the set of population states.

3.1 Deterministic Evolutionary Dynamics

How the analysis proceeds depends on the time horizon of interest. Suppose that for the application in question, our interest is in moderate time spans. Then if the population size is large enough, the idiosyncratic noise in agent's choices is averaged away, so that the evolution of aggregate behavior follows an almost deterministic path (Benaïm and Weibull (2003)). This path is described by a solution to an ordinary differential

[1]Evolutionary models and heuristic learning models are sometimes distinguished by the inputs to the agents' decision rules. In both the stochastic evolutionary model of Kandori et al. (1993) and the heuristic learning model of Young (1993), agents' decisions take the form of noisy best responses. But in the former model agents evaluate each action by its performance against the population's current behavior, while in the latter they consider performance against the time averages of opponents' past play.

[2]By dropping this coordination assumption, one obtains the more general class of *Bayesian learning* (or *belief learning*) models. Since such models can entail quite naive beliefs, belief learning models overlap with heuristic learning models—see Section 4 below.

[3]See Sandholm (2006).

equation. For example, Björnerstedt and Weibull (1996) and Schlag (1998) show that if agents use certain revision protocols based on imitation of successful opponents, then the population's aggregate behavior follows a solution to Taylor and Jonker's (1978) replicator dynamic. This argument provides an alternative, economic interpretation of this fundamental evolutionary model.

Much of the literature on deterministic evolutionary dynamics focuses on connections with traditional game theoretic solution concepts. For instance, under a wide range of deterministic dynamics, all Nash equilibria of the underlying game are rest points. While some dynamics (including the replicator dynamic) have additional non-Nash rest points, there are others under which rest points and Nash equilibria are identical (Brown and von Neumann (1950); Smith (1984); Sandholm (2006)).

A more important question, though, is whether Nash equilibrium will be approached from arbitrary disequilibrium states. For certain specific classes of games, general convergence results can be established (Hofbauer (2000); Sandholm (2007)). But beyond these classes, convergence cannot be guaranteed. One can construct games under which no reasonable deterministic evolutionary dynamic will converge to equilibrium—instead, the population cycles through a range of disequilibrium states forever (Hofbauer and Swinkels (1996); Hart and Mas-Colell (2003)). More surprisingly, one can construct games in which nearly all deterministic evolutionary dynamics not only cycle forever, but also fail to eliminate strictly dominated strategies (Hofbauer and Sandholm (2006)). If we truly are interested in modeling the dynamics of behavior, these results reveal that our predictions cannot always be confined to some set of static equilibria; rather, more complicated limit phenomena like cycles and chaotic attractors must also be permitted as predictions of play.

3.2 Stochastic Evolutionary Dynamics

If we are interested in behavior over very long time horizons, deterministic approximations are no longer valid, and we must study our original Markov process directly. Under certain nondegeneracy assumptions, the long run behavior of this process is captured by its unique stationary distribution, which describes the proportion of time the process spends in each population state.

While stochastic evolutionary processes can be more difficult to analyze than their deterministic counterparts, they also permit us to make surprisingly tight predictions. By making the amount of noise in agents' choice rules vanishingly small, one can often ensure that all mass in the limiting stationary distribution is placed on a single population state. This *stochastically stable state* provides a unique prediction of play even in games with multiple strict equilibria (Foster and Young (1990); Kandori et al. (1993)).

The most thoroughly studied model of stochastic evolution considers agents who usually play a best response to the current population state, but who occasionally choose a strategy at random. Kandori et al. (1993) show that if the agents are randomly matched to play a symmetric 2 x 2 coordination game, then taking the "mutation" probability to zero leads to the selection of the *risk dominant equilibrium* as the unique stochastically stable state.[4] Selection results of this sort have since been extended to cases in which the underlying game has an arbitrary number of strategies, as well as to settings in which agents are positioned on a fixed network, interacting only with neighbors (see Kandori and Rob (1995); Blume (2003); Ellison (1993, 2000)). Stochastic stability has also been employed in contexts where the underlying game has a nontrivial extensive form; these analyses have provided support for notions of backward induction (e.g., subgame perfection) and forward induction (e.g., signaling game equilibrium refinements)—see Nöldeke and Samuelson (1993) and Hart (2002).

Still, these selection results must be interpreted with care. When the number of agents is large or the rate of "mutation" is small, states that fail to be stochastically stable can be coordinated upon for great lengths of time (Binmore et al. (1995)). Consequently, if the relevant time span for the application at hand is not long enough, the stochastically stable state may not be the only reasonable prediction of behavior.

4. Learning in Games

4.1 Heuristic Learning

Learning models study disequilibrium adjustment processes in repeated games. Like evolutionary models, heuristic learning models assume that players employ simple myopic rules in deciding how to act. In the simplest of these models, each player decides how to act by considering the payoffs he has earned in the past. For instance, under reinforcement learning (Börgers and Sarin (1997); Erev and Roth (1998)), agents choose each strategy with probability proportional to the total payoff that the strategy has earned in past periods.

By considering rules that look not only at payoffs earned, but also at payoffs foregone, one can obtain surprisingly strong convergence results. Define a player's *regret* for (not having played) action a to be the difference between the average payoff he would have

[4] In the risk dominant equilibrium, all agents play the action that is optimal against an opponent who is equally likely to choose each action.

earned had he always played *a* in the past, and the average payoff he actually received. Under *regret matching*, each action whose regret is positive is chosen with probability proportional to its regret. Hart and Mas-Colell (2000) show that regret matching is a *consistent* repeated game strategy: it forces a player's regret for each action to become nonpositive. If used by all players, regret matching ensures that their time-averaged behavior converges to the set of coarse correlated equilibria of the underlying game.[5]

Some of the most striking convergence results in the evolution and learning literature establish a stronger conclusion: namely, convergence of time-averaged behavior to the set of *correlated equilibria*, regardless of the game at hand. The original result of this sort is due to Foster and Vohra (1997, 1998), who prove the result by constructing a calibrated procedure for forecasting opponents' play.[6] Hart and Mas-Colell (2000) show that simpler procedures also ensure convergence to correlated equilibrium—in particular, procedures that define conditionally consistent repeated game strategies.[7,8]

In some heuristic learning models, players use simple rules to predict how opponents will behave, and then respond optimally to those predictions. The leading examples of such models are *fictitious play* and its stochastic variants (Brown (1951); Fudenberg and Kreps (1993)): in these models, the prediction about an opponents' next period play is given by the empirical frequencies of his past plays. Beginning with Robinson (1951), many authors have proved convergence results for standard and stochastic fictitious play in specific classes of games.[9] But as Shapley (1964) and others have shown, these models do not lead to equilibrium play in all games.

4.2 Coordinated Bayesian Learning

The prediction rule underlying two-player fictitious play can be described by a belief about the opponent's repeated game strategy that is updated using Bayes' rule in the face of observed play. In particular, this belief specifies that the opponent chooses his stage

[5]*Coarse correlated equilibrium* is a generalization of correlated equilibrium under which players' incentive constraints must be satisfied at the ex ante stage rather than at the interim stage—see Young (2004).

[6]A *forecasting procedure* produces probabilistic forecasts of how opponents will act. The procedure is *calibrated* if in those periods in which the forecast is given by the probability vector p, the empirical distribution of opponents' play is approximately p.

[7]A repeated game strategy is *conditionally consistent* if for each frequently played action a, the agent would not have been better off had he always played an alternative action a' in place of a.

[8]Another variety of heuristic learning models, based on *random search and independent verification*, ensures a stochastic form of convergence to Nash equilbrium regardless of the game being played (Foster and Young (2003)). Since in these models the time required before equilibrium is first reached is quite long, they are most relevant to applications with especially long time horizons.

[9]For an overview of these results, see Hofbauer and Sandholm (2002).

game actions in an i.i.d. fashion, conditional on the value of an unknown parameter.[10] Evidently, each player's beliefs about his opponent are wrong: player 1 believes that player 2 chooses actions in an i.i.d. fashion, whereas player 2 actually plays optimally in response to his own (i.i.d.) predictions about player 1's behavior. It is therefore not surprising that fictitious play processes do not converge in all games.

In models of *coordinated Bayesian learning* (or *rational learning*), it is not only supposed that players form and respond optimally to beliefs about the opponent's repeated game strategy; it is also assumed that the players' initial beliefs are coordinated in some way. The most studied case is one in which prior beliefs satisfy an absolute continuity condition: if the distribution over play paths generated by the players' actual strategies assigns positive probability to some set of play paths, then so must the distribution generated by each player's prior. A strong sufficient condition for absolute continuity is that each player's prior assigns a positive probability to his opponent's actual strategy.

The fundamental result in this literature, due to Kalai and Lehrer (1993), shows that under absolute continuity, each player's forecast along the path of play is asymptotically correct, and the path of play is asymptotically consistent with Nash equilibrium play in the repeated game. Related convergence results have been proved for more complicated environments in which each player's stage game payoffs are private information (Jordan (1995); Nyarko (1998)). If these type distributions are continuous, then the sense in which play converges to equilibrium can involve a form of purification: while actual play is pure, it appears random to an outside observer.

How much coordination of prior beliefs is needed to prove convergence to equilibrium play? Nachbar (2005) proves that for a large class of repeated games, for any belief learning model, there are no prior beliefs that satisfy three criteria: learnability, consistency with optimal play, and diversity. Thus, if players can learn to predict one another's behavior, and are capable of responding optimally to their updated beliefs, then each player's beliefs about his opponents must rule out some seemingly natural strategies a priori. In this sense, the assumption of coordinated prior beliefs that ensures convergence to equilibrium in rational learning models does not seem dramatically weaker than a direct assumption of equilibrium play.

ELSEWHERE IN PALGRAVE:
Adaptive Heuristics
Belief Learning

[10]Moreover, the player's beliefs about this parameter must come from the family of Dirichlet distributions, the conjugate family of distributions for multinomial trials—see Fudenberg and Levine (1998).

Deterministic Evolutionary Dynamics
ESS
Stochastic Adaptive Dynamics

References

Benaïm, M. and Weibull, J. W. (2003). Deterministic approximation of stochastic evolution in games. *Econometrica*, 71:873–903.

Binmore, K., Samuelson, L., and Vaughan, R. (1995). Musical chairs: Modeling noisy evolution. *Games and Economic Behavior*, 11:1–35.

Björnerstedt, J. and Weibull, J. W. (1996). Nash equilibrium and evolution by imitation. In Arrow, K. J. et al., editors, *The Rational Foundations of Economic Behavior*, pages 155–181. St. Martin's Press, New York.

Blume, L. E. (2003). How noise matters. *Games and Economic Behavior*, 44:251–271.

Börgers, T. and Sarin, R. (1997). Learning through reinforcement and the replicator dynamics. *Journal of Economic Theory*, 77:1–14.

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.

Brown, G. W. and von Neumann, J. (1950). Solutions of games by differential equations. In Kuhn, H. W. and Tucker, A. W., editors, *Contributions to the Theory of Games I*, volume 24 of *Annals of Mathematics Studies*, pages 73–79. Princeton University Press, Princeton.

Ellison, G. (1993). Learning, local interaction, and coordination. *Econometrica*, 61:1047–1071.

Ellison, G. (2000). Basins of attraction, long run equilibria, and the speed of step-by-step evolution. *Review of Economic Studies*, 67:17–45.

Erev, I. and Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review*, 88:848–881.

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.

Foster, D. P. and Young, H. P. (1990). Stochastic evolutionary game dynamics. *Theoretical Population Biology*, 38:219–232.

Foster, D. P. and Young, H. P. (2003). Learning, hypothesis testing, and Nash equilibrium. *Games and Economic Behavior*, 45:73–96.

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

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

Hart, S. (2002). Evolutionary dynamics and backward induction. *Games and Economic Behavior*, 41:227–264.

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. (2003). Uncoupled dynamics do not lead to Nash equilibrium. *American Economic Review*, 93:1830–1836.

Hofbauer, J. (2000). From Nash and Brown to Maynard Smith: Equilibria, dynamics, and ESS. *Selection*, 1:81–88.

Hofbauer, J. and Sandholm, W. H. (2002). On the global convergence of stochastic fictitious play. *Econometrica*, 70:2265–2294.

Hofbauer, J. and Sandholm, W. H. (2006). Survival of dominated strategies under evolutionary dynamics. Unpublished manuscript, University of Vienna and University of Wisconsin.

Hofbauer, J. and Sigmund, K. (1988). *Theory of Evolution and Dynamical Systems*. Cambridge University Press, Cambridge.

Hofbauer, J. and Swinkels, J. M. (1996). A universal Shapley example. Unpublished manuscript, University of Vienna and Northwestern University.

Jordan, J. S. (1995). Bayesian learning in repeated games. *Games and Economic Behavior*, 9:8–20.

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

Kandori, M., Mailath, G. J., and Rob, R. (1993). Learning, mutation, and long run equilibria in games. *Econometrica*, 61:29–56.

Kandori, M. and Rob, R. (1995). Evolution of equilibria in the long run: A general theory and applications. *Journal of Economic Theory*, 65:383–414.

Maynard Smith, J. and Price, G. R. (1973). The logic of animal conflict. *Nature*, 246:15–18.

Nachbar, J. H. (2005). Beliefs in repeated games. *Econometrica*, 73:459–480.

Nöldeke, G. and Samuelson, L. (1993). An evolutionary analysis of backward and forward induction. *Games and Economic Behavior*, 5:425–454.

Nyarko, Y. (1998). Bayesian learning and convergence to Nash equilibria without common priors. *Economic Theory*, 11:643–655.

Robinson, J. (1951). An iterative method of solving a game. *Annals of Mathematics*, 54:296–301.

Sandholm, W. H. (2006). Pairwise comparison dynamics and evolutionary foundations for Nash equilibrium. Unpublished manuscript, University of Wisconsin.

Sandholm, W. H. (2007). Population games and evolutionary dynamics. Unpublished manuscript, University of Wisconsin.

Schlag, K. H. (1998). Why imitate, and if so, how? a boundedly rational approach to multi-armed bandits. *Journal of Economic Theory*, 78:130–156.

Shapley, L. S. (1964). Some topics in two person games. In Dresher, M., Shapley, L. S., and Tucker, A. W., editors, *Advances in Game Theory*, volume 52 of *Annals of Mathematics Studies*, pages 1–28. Princeton University Press, Princeton.

Smith, M. J. (1984). The stability of a dynamic model of traffic assignment—an application of a method of Lyapunov. *Transportation Science*, 18:245–252.

Taylor, P. D. and Jonker, L. (1978). Evolutionarily stable strategies and game dynamics. *Mathematical Biosciences*, 40:145–156.

Young, H. P. (1993). The evolution of conventions. *Econometrica*, 61:57–84.

Young, H. P. (2004). *Strategic Learning and Its Limits*. Oxford University Press, Oxford.