ELSEVIER

European Economic Review 42 (1998) 631–639

EUROPEAN ECONOMIC REVIEW

Learning and Evolution: Where Do We Stand?

Learning in games

Drew Fudenberg[a,*], David Levine[b]

[a] *Department of Economics, Harvard University, Cambridge, MA 02138, USA*
[b] *Department of Economics, UCLA, Los Angeles, CA 90024-1477, USA*

Abstract

This essay discusses some recent work on 'learning in games'. We explore non-equilibrium theories in which equilibrium emerges as the long-run outcome of a dynamic process of adjustment or learning. We focus on individual level models, and more specifically on variants of 'fictitious play' in two-player games. We discuss both the theoretical properties of the models and their relationship to regularities observed in game theory experiments. © 1998 Elsevier Science B.V. All rights reserved.

JEL classification: C72; C92; D83

Keywords: Learning in games; Fictitious play; Consistency; Experiments; Game theory

1. Introduction

This essay discusses some recent work on learning in games, with emphasis on our own research and on its relation to game experiments. The basic research agenda is to explore non-equilibrium explanations of equilibrium in games, and to view equilibrium as the long-run outcome of a dynamic process of adjustment or learning. Instead of distinguishing between models of 'learning' and other forms of adjustment, it is easier to make a distinction between models that describe behavior of individual agents and those that start at the aggregate level.

* Corresponding author. E-mail: fudenberg@fas.harvard.edu.

PII S0014-2921(98)00011-7

This essay focuses on individual level models, and more specifically on variants of 'fictitious play' in two-player games. Jorgen Weibull's essay in this issue discusses aggregate models in general and the replicator dynamic in particular; Peyton Young's essay discusses related stochastic models of individual adjustment.

The first question a learning model must address is what individuals know before the game starts, and what it is that they are learning about. A closely connected issue is how much rationality to attribute to the agents. We believe that for most purposes the right models involve neither full rationality nor the extreme naïveté of most stimulus-response models; 'better' models have agents consciously but perhaps imperfectly *trying* to get a good payoff. In our opinion, sensible rules should do reasonably well in a reasonably broad set of circumstances; we believe that rules that do poorly in simple environments are not likely to be used for important decisions.

We will simplify matters by assuming that players know the extensive form of the game and their own payoff functions; they may or may not know their opponents' payoffs. From this point of view, if players do have information about opponents' payoffs they use it to form prior beliefs about how their opponents are likely to play. For example, they might think it is unlikely that opponents will play strategies that are dominated given the presumed payoff functions.[1]

Besides specifying the players' forecast rules, an analysis of learning must also address the issue of whether players try to influence their opponents' play: If two people repeatedly play a two person game against each other, they ought to consider the possibility that their current play may influence the future play of their opponent. Most of learning theory abstracts from these repeated game considerations by explicitly or implicitly relying on a model in which the incentive to try to alter the future play of opponents is small enough to be negligible. This can be justified by appeal to models with a large number of players, who interact anonymously (which is the case in most experiments), with the population size large compared to the discount factor.[2]

---

[1] These two cases fit most laboratory experiments, but in other applications players might have less information. At the other extreme, players might not know anything about their opponents, so that players do not even know if they are playing a game, and are only aware of a subset of their own possible strategies. They might observe only the history of payoffs and their own strategies, and discover new strategies by experimentation or deduction. So far only the genetic algorithm model of learning (for example, Holland, 1975) has attempted to explore a setting like this.

[2] See Fudenberg and Kreps (1990) for a discussion of various large-population models and Ellison (1995) for a caveat. Although the large population stories explain naïve or myopic play; they only apply if the relevant population is large. Such cases may be more prevalent than it first appears, as players may extrapolate between similar games.

2. Fictitious play

We present the standard model of fictitious play, which has a single agent in each player role.[3] Denote a strategy by player i by s^i, and the set of i's strategies by S^i; we use $-i$ to denote the player other than i. The best response correspondence is denoted by BR^i. In fictitious play, player i has an exogenous initial weight function $\kappa_0^i : S^{-i} \to \Re_+$. This weight is updated by adding 1 to the weight of a strategy each time it is played. The probability that player i assigns to player $-i$ playing s^{-i} at date t is given by

$$\gamma_t^i(s^{-i}) = \frac{\kappa_t^i(s^{-i})}{\sum_{\tilde{s}^{-i} \in S^{-i}} \kappa_t^{-i}(\tilde{s}^{-i})}.$$

This corresponds to Bayesian inference when player i believes that his opponents' play corresponds to a sequence of i.i.d. multinomial random variables with a fixed but unknown distribution, and player i's prior beliefs over that unknown distribution take the form of a Dirichlet distribution. However, the key feature of this specification is that all observations are weighted equally, so beliefs converge to the empirical distribution. In fitting experimental data it is better to allow geometrically declining weights, as shown for example by Cheung and Friedman (1994).

Fictitious play is any rule that assigns $\rho_t^i(\gamma_t^i) \in BR^i(\gamma_t^i)$.

Proposition (Fudenberg and Kreps, 1993). (a) *If s is a strict Nash equilibrium, and s is played at date t in the process of fictitious play, s is played at all subsequent dates. That is, strict Nash equilibria are absorbing for the process of fictitious play.* (b) *Any pure strategy steady state of fictitious play must be a Nash equilibrium.* (c) *If the empirical marginal distributions converge, the strategy profile corresponding to the product of these distributions is a Nash equilibrium.*

This convergence result above supposes players ignore cycles, even if, because of these cycles, the empirical *joint* distribution of the two players' play is correlated, and does not equal the product of the empirical marginal distributions. As a result, fictitious play can give payoffs below the minmax values, as shown by examples of Fudenberg and Kreps (1990, 1993), Jordan (1993) and Young (1993).[4]

---

[3] Recent work by Kanivokski and Young (1994) examines the dynamics of fictitious play in population of players.

[4] In these examples, players switch their strategies in every period, and manage to persistently mis-coordinate. Fudenberg and Levine (1994) and Monderer et al. (1994) show that the realized time-average payoffs under fictitious play cannot be much below what the players expect, and, in particular, must be individually rational, if the time path of play exhibits '*infrequent switches*'.

Fictitious play is *consistent*, meaning that it does as well as playing a best response to the time average when the opponent's play is generated by i.i.d. draws from a fixed distribution, but this is a very weak property. Instead, we ask that learning rules be *universally consistent*, meaning roughly that they should do as well as playing a best response to the time average regardless of the rule used to generate the opponent's play. Formally, let $\bar{\gamma}_t^{-i}$ be the empirical distribution of play of i's opponents.

Definition. A behavior rule ρ^i is *ε-universally consistent* if for any ρ^{-i}

$$\lim_{T\to\infty} \max_{\sigma^i} u^i(\sigma^i, \bar{\gamma}_t^{-i}) - \frac{1}{T}\sum_t u^i(\rho_t^i(h_{t-1})) \leq \varepsilon$$

almost surely with respect to the stochastic process induced by ρ^i, ρ^{-i} for every ρ^{-i}.

The existence of such rules was originally shown by Hannan (1957) and Blackwell (1956). Fictitious play is not universally consistent, but we will see that a simple modification of it is.

3. Stochastic fictitious play

In fictitious play, players can only randomize if exactly indifferent. To develop a sensible model of learning to play mixed strategies, one should start with an explanation for mixing. One such explanation is Harsanyi's (1973) purification theorem, which explains a mixed distribution over actions as the result of unobserved payoff perturbations. A second explanation for mixing is that it is descriptively realistic: some psychology experiments show that choices between alternatives that are perceived as similar tend to be relatively random. A third motivation for supposing that players randomize is that stochastic fictitious play allows behavior to avoid the discontinuity inherent in standard fictitious play; consequently, it allows good long-run performance in a much wider range of environments.

All three motivations lead to models where the best-response function BR^i are replaced by 'smooth best-response functions' $\overline{BR}^i$, with

$$\overline{BR}^i(\sigma^{-i})(s^i) = \text{Prob}[s^i \,|\text{opponents expected to play } \sigma^{-i}].$$

Definition. The profile σ is a *Nash distribution* if $\overline{BR}^i(\sigma^{-i}) = \sigma^i$ for all i.

This distribution may be very different from any Nash equilibria of the original game if the smoothed best-response functions are very far from the original ones. But in the context of randomly perturbed payoffs, Harsanyi's

purification theorem shows that ('generically') the Nash distributions of the perturbed game approach the Nash equilibria of the original game as the support of the payoff perturbations become concentrated about 0.

Smooth fictitious play is any model in which behavior is determined by applying some smooth best-response function to the beliefs defined earlier. While smooth fictitious play is random, the time average of many independent random variables has very little randomness, so the asymptotic dynamics resembles that of the continuous time 'near best-response' dynamics $\dot{\theta}^i = \overline{BR}^i(\theta) - \theta^i$. Consequently, any stationary point of the system must be a Nash distribution. Moreover, if a point or cycle is an unstable orbit of the continuous time dynamics, we expect that the noise would drive the system away, so that the stochastic system can only converge to stable orbits of the continuous-time system. These intuitions can be verified using the techniques of 'stochastic approximation'. For example,

Proposition (Benaim and Hirsch, 1996). *Consider a two-player smooth fictitious play in which every strategy profile has positive probability at any state θ. If θ^* is an asymptotically stable equilibrium of the continuous time process, then regardless of the initial conditions $P[\theta_t \to \theta_*] > 0$.*

Fudenberg and Levine (1995) show that universal consistency can be accomplished by a smooth fictitious play procedure in which $\overline{BR}^i$ is derived from maximizing a function of the form $u^i(\sigma) + \lambda v^i(\sigma^i)$, and at each date players play $\overline{BR}^i$ applied to their current beliefs. For example, if $v^i(\sigma^i) = \sum_{s^i} -\sigma^i(s^i)\log \sigma^i(s^i)$, then

$$\overline{BR}^i(\sigma^{-i})[s^i] \equiv \frac{\exp((1/\lambda)u^i(s^i, \sigma^{-i}))}{\sum_{r^i} \exp((1/\lambda)u^i(r^i, \sigma^{-i}))}.$$

This is 'logistic fictitious play': each strategy is played in proportion to an exponential function of the utility it has historically yielded; it corresponds to the logit decision model that has been extensively used in empirical work. Notice that as $\lambda \to 0$ the probability of playing any strategy that is not a best response goes to zero. Note also that this is not a very complex rule. Since it is not difficult to implement universally consistent rules universal consistency may be a useful benchmark for evaluating learning rules, even if it is not exactly descriptive of experimental data in moderate size samples and even if individual players do not implement the particular procedures discussed here.

4. Unobserved 'opponents' strategies

The results on universal consistency go through if players do not know the strategic form, and do not observe the opponent's play at the end of each round,

but only their own realized payoffs. Intuitively, observing one's own payoffs is enough to identify the best action given the empirical frequency of opponent's actions and hence is enough for universal consistency; observing the extra information about opponent's actions can result in more efficient (faster) learning but will not improve long-run performance. This idea has since been tested by Er'ev and Roth (1997) in the context of the reinforcement learning models discussed below

In fact, there is a simple variation on logistic fictitious play that yields asymptotically the same behavior as if the opponent's play was observed. For example, the following logistic rule is universally consistent:

$$\rho^i(h_{t-1})[s^i] \equiv \frac{\exp((1/\lambda)\bar{u}_t^i)(s^i)}{\sum_{r^i}\exp((1/\lambda)\bar{u}_t^i(r^i))},$$

where the 'average utilities' $\bar{u}_t^i$ are defined as

$$\bar{u}_t^i(s^i) = \bar{u}_{t-1}^i(s^i) \text{ if } s^i \neq s_t^i,$$

$$\bar{u}_t^i(s^i) = \bar{u}_{t-1}^i(s^i) + \frac{1}{\rho^i(h_t(s^i))\kappa_{t-1}(s^i)}[\tilde{u}_t^i(s^i) - \bar{u}_{t-1}^i].$$

This sort of rule can be viewed as a kind of 'stimulus-response' dynamic. In a stimulus–response model, the probability an action is played depends on a 'propensity', in this case the 'average utility'. The propensity with which and action is played increases if the utility received when it is used exceeds an 'aspiration level', and decreases if it is less than the 'aspiration level'. When the utility exceeds the aspiration level, we refer to a 'positive reinforcement', when it is less than the aspiration level, we refer to a 'negative reinforcement'. Here the aspiration level is simply the average utility to date. This model differs from a typical stimulus–response model because of the presence of ρ^i in the denominator; typically stimulus–response models weight all observations equally.

5. Smooth fictitious play compared to stimulus response

The stimulus–response model was developed largely in response to observations by psychologists about human behavior and animal behavior: choices that lead to good outcomes are more likely to be repeated and behavior is random. Many learning processes share both of these properties.[5] However, the

[5] See also Camerer and Ho (1996), who construct a general learning model that has variants of fictitious play and Erev–Roth's reinforcement model as special cases. They conclude that neither model does as well as a combination of the two.

original 'stimulus–response' model supposes that aspiration levels are fixed. Such a procedure has odd long-run properties, such as 'probability matching' instead of optimization in simple prediction tasks.[6] At one time psychologists believed this was characteristic of human behavior, but probability matching does not seem to persist when subjects have enough experience.

Roth and Er'ev (1995) and Er'ev and Roth (1997) study a more sophisticated stimulus response model in which the aspiration level follows a linear adjustment process. They also place a minimum probability on every strategy being played. They conclude that their model fits better than does deterministic fictitious play. In a similar vein, Bereby-Meyer and Erev (1997) show that the speed of learning in a simple prediction task is influenced by whether the payoffs are presented as rewards or as losses from a fixed endowment. This finding is not consistent with logistic fictitious play, but is consistent with an adjustable reference point. An important issue is whether we should expect to observe this type of framing effect when subjects have thought carefully about the problem.

In contrast, Cheung and Friedman (1994) have had some success in fitting modified fictitious play models to experimental data. Their model (in games with two actions) has three parameters per player: the probability that a player uses his first action is a linear function (two parameters) of the exponentially weighted history (one more parameter for the exponential weight). The median exponential weighting over players and experiments is about 0.3 or 0.4; best response corresponds to weight zero and classic fictitious play is weight one.

Van Huyck et al. (1996) report on an experimental study based on a simple 2×2 coordination game. They report that they can reject the hypothesis that players are using historical performance of strategies (as they would in stimulus–response type models) in favor of the hypothesis that they are using forecast performance of strategies (as they would in smooth fictitious play type models). Indeed, in their data the model of exponential fictitious play fits the data quite well in all but one of 12 sessions.[7]

While the stimulus-response versus belief-based learning debate continues on, we can at least identify several key issues. First, both models implicitly assume stationarity; they do not look for patterns in the data such as convergence to a particular equilibrium. Real players are more sophisticated than this. This raises the question whether using these models (and various ad hoc devices such as exponential weights) results in a good fit by 'accident' because the particular ad hoc assumptions happen to be a closer fit to a model in which players recognize patterns in the data. Also, belief-based learning focuses attention on

[6] This shows that it need not be consistent, let alone universally consistent.

[7] Er'ev and Roth (1997) also find weak evidence in favor of this view: their model fits better on the assumption that players give equal weight to all data than under the usual reinforcement assumption that subjects only receive reinforcement from their own experience.

the formation of priors and the idea that players may be able to form 'good' priors by reasoning about the game, while stimulus-response models are more mechanical.

6. Use of prior information

An experiment by Prasnikar and Roth (1992) on the 'best-shot' game, in which two players sequentially decide how much to contribute to a public good, shows the importance of prior information. Prasnikar and Roth ran two treatments of this game. In the first, players were informed of the function determining opponents' monetary payoffs. Here, by the last few rounds of the experiment the first movers had stopped contributing, which is the backwards induction solution. In the second treatment, subjects were not given any information about the payoffs of their opponents. In this treatment even in the later rounds of the experiment a substantial number of player 1's contributed 4 to the public good, while others played 0.

This is not consistent with backwards induction or even Nash equilibrium, but it is consistent with an (approximate, heterogeneous) self-confirming equilibrium (Fudenberg and Levine, 1997), so these experiments provide evidence that information about other players' payoffs is used as if players are maximized.

Erev and Rapoport (1997) study a market entry game and that 'information about other players' payoff, that should have no effect according to the Roth/Erev model, does influence behavior in this experiment'.

Reflecting on these papers suggests to us that theorists and experimenters interested in learning should develop, characterize, and test models of 'intermediate-cognition' procedural rationality; meaning models in which players are at least trying, albeit imperfectly, to obtain some objective.

Acknowledgements

This essay draws on our forthcoming book *The Theory of Learning in Games* and also on an earlier essay based on a talk at the 1996 North American meeting of the Econometric Society. We thank Ken Binmore and Al Roth for helpful comments on the earlier essay.

References

Benaim, M., Hirsch, M., 1996. Learning processes, mixed equilibria and dynamical systems arising from repeated games. Mimeo.

Bereby-Meyer, Y., Erev, I., 1997. On learning to become a successful loser: A comparison of alternative abstractions of learning processes in the loss domain. Mimeo.

Blackwell, D., 1956. Controlled random walks. Proceedings International Congress of Mathematicians, vol. III, pp. 336–338, North-Holland, Amsterdam.

Camerer, C., Ho, T., 1996. Experience-weighted attraction learning in games: A unifying approach. Mimeo.

Cheung, Y., Friedman, D., 1994. Learning in Evolutionary Games: Some Laboratory Results. Santa Cruz.

Ellison, G., 1995. Learning from personal experience: On the large population justification of myopic learning models. Mimeo. MIT, Cambridge, MA.

Er'ev, I., Roth, A., 1997. On the need for low rationality cognitive game theory: Reinforcement learning in experimental games with unique mixed strategy equilibria. Mimeo. University of Pittsburgh.

Erev, I., Rapoport, A., 1997. Coordination, 'magic', and reinforcement learning in a market entry game. Mimeo.

Fudenberg, D., Kreps, D., 1990. Lectures on Learning and Equilibrium in Strategic-Form Games, CORE Lecture Series.

Fudenberg, D., Kreps, D., 1993. Learning mixed equilibria. Games and Economic Behavior 5, 320–367.

Fudenberg, D., Levine, D.K., 1995. Consistency and cautious fictitious play. Journal of Economic Dynamics and Control 19, 1065–1090.

Fudenberg, D., Levine, D.K., 1997. Measuring subjects losses in experimental games. Quarterly Journal of Economics 112, 479–506.

Hannan, J., 1957. Approximation to Bayes risk in repeated plays. In: Dresher, M., Tucker, A.W., Wolfe, P. (Eds.), Contributions to the Theory of Games, vol. 3. Princeton University Press, Princeton, NJ, pp. 97–139.

Holland, J.H., 1975. Adaptation in Natural and Artificial Systems. The University of Michigan Press, Ann Arbor, MI.

Jordan, J., 1993. Three problems in learning mixed-strategy Nash equilibria. Games and Economic Behavior 5, 368–386.

Kanivokski, Y., Young, P., 1994. Learning dynamics in games with stochastic perturbations. Mimeo.

Monderer, D., Samet, D., Sela, A., 1994. Belief affirming in learning processes. Technion.

Prasnikar, V., Roth, A., 1992. Considerations of fairness and strategy: experimental data from sequential games. Quarterly Journal of Economics 107, 865–880.

Roth, A., Er'ev, I., 1995. Learning in extensive-form games: experimental data and simple dynamic models in the intermediate term. Games and Economic Behavior 8, 164–212.

Van Huyck, J., Battalio, R., Rankin, F., 1996. On the Evolution of Convention: Evidence from Coordination Games. A&M, Texas.