ELSEVIER

European Economic Review 46 (2002) 791–799

EUROPEAN ECONOMIC REVIEW

www.elsevier.com/locate/econbase

Bounded Rationality and Learning

On the limits to rational learning

H. Peyton Young *

Department of Economics, Johns Hopkins University, Baltimore, MD 21218, USA

Abstract

This paper summarizes recent work of Foster and Young (2001), which shows that some games are unlearnable in principle by perfectly rational players. That is, under any learning rule – including Bayesian updating of common priors – the players' strategies fail to come close to Nash equilibrium with probability one. Furthermore at least one of them is unable to predict the behavior of the other in an asymptotic sense. This result can be interpreted as an "uncertainty principle" that applies to some kinds of interactive learning problems.

JEL classification: C72; D83

Keywords: Learning; Prediction; Rationality; Nash equilibrium

1. Introduction

This paper is concerned with the foundations of rational learning, and the conditions under which rational learning leads to equilibrium behavior from out-of-equilibrium conditions. If a system is composed of rational agents using the information revealed by their opponents' play, can the players learn to predict their opponents' future behavior, and does the learning process converge to equilibrium? Variants of this question have been discussed in both the macro and microeconomics literature. For example, Grandmont (1998) examines the convergence properties of macroeconomic learning models, and argues that their stability in the neighborhood of a rational expectations equilibrium depends crucially on the agents' *expectations about their stability*, from which he concludes that convergence is, at best, indeterminate. Similarly, Blume and Easley (1998) examine the learning problem in a competitive market setting, and argue that convergence to a price equilibrium by rational agents will only happen if beliefs are aligned in a particular way.

* Tel.: +1-410-526-6118; fax: +1-410-516-7600.
E-mail address: pyoung@jhu.edu (H.P. Young).

Similar results have been obtained for repeated games. For example, Kalai and Lehrer (1993) show that convergence to Nash equilibrium in an infinitely repeated game hinges on whether the ex ante beliefs put positive probability on the opponents' actual strategies (more generally, whether the beliefs put positive probability on any event that has positive probability under the strategies). As various authors have argued (and as we shall show below), this condition places severe restrictions on the players' initial beliefs (Nachbar, 1997, 1999, 2001; Jordan, 1991, 1993, 1995; Miller and Sanchirico, 1997; Foster and Young, 2001).

In this paper, we demonstrate a class of learning environments in which convergence to equilibrium behavior fails to occur for *any* learning process, including the Bayesian updating of objectively correct priors. An essential property of these environments is that all of the equilibria are fully mixed. When beliefs are close to such an equilibrium (but not yet at it), the behavior of perfectly rational agents becomes increasingly erratic: due to their insistence on strictly optimizing at all times, their responses keep shifting from one pure action to another in a way that is difficult to predict without knowing their payoff functions exactly. The result is that the learning process never catches up with the increase in complexity of the agents' behavior, and the agents are unable to predict the behavior of the system.

Lest the reader be misled, we hasten to point out that this impossibility result concerns the behavior of the *individual agents* in the system, not the behavior of the system *as a whole*. There are well-known learning processes, such as fictitious play, for which the *time average* of play converges to the mixed Nash equilibrium in 2×2 games (Miyasawa, 1961; Monderer and Shapley, 1996). There are other games in which the *average behavior of a population* of players mimics Nash equilibrium from the observer's standpoint (Harsanyi, 1973; Fudenberg and Kreps, 1993; Fudenberg and Ellison, 2000). (Indeed Nash (1950) suggested such an interpretation in his Ph.D. Dissertation.) Our concern, by contrast, is with the behavior of individuals. At this level, rationality may never lead them to play close to Nash equilibrium strategies, and they may never learn to predict the behavior of their opponents with any degree of accuracy.

2. Rational learning

We begin by recalling the elements of neoclassical learning theory as applied to repeated games. Given is a finite collection of players $i = 1,2,\ldots,n$, each of whom chooses an action from a finite set of available actions each period. Let X_i denote the action space of player i, and let $X = \prod X_i$. Each agent i is endowed with a von Neumann-Morgenstern utility function $u_i(x)$ over plays $x \in X$. Each discounts future payoffs by a fixed discount factor $\delta_i < 1$. *Rationality* means that each agent chooses a contingent plan of action that maximizes his discounted expected utility at each point in time t, given his beliefs about the probability of future play paths, and conditional on the observed history of play up to that time.

Consider the following information structure. At the beginning of the game, each player i is informed of his own utility function $u_i(x)$; he is also is informed of the

distribution from which the utility functions of the other players are drawn. (We can think of a utility function as a point in R^X, and a distribution over utility functions as a probability distribution on R^X.) We shall assume that these distributions are independent among the n players, so that a player's realized utility function conveys no additional information about the utility function of any other player.

All plays are publicly observed and the agents remember the plays from all previous periods. A player's *strategy* is a contingent plan of action, including randomized plans of action, conditional on each possible history of play to each point in time t. A player's *belief at time t* is a probability distribution over the opponents' strategies from time t on, conditional on the history of play through time $t-1$. A player is *rational* if, at each time t, his strategy from t on optimizes his expected discounted utility given his updated belief at t. (This is also known as "sequential rationality".)

The updating of beliefs can be either simple or complex. A very simple scheme (the basis of fictitious play) is to believe that the opponent will play a fixed probability distribution p^{t-1} in each period from t on, where the distribution p^{t-1} is the observed empirical frequency distribution of his play up through time $t-1$. More complex schemes involve theorizing about how the opponent updates his beliefs, and still more complex ones involve theorizing about how the opponent theorizes about one's own updating scheme, etc. While these schemes may be very sophisticated and complex, however, they are not necessarily better specified than naïve schemes, nor are they necessarily more "rational". By *rationality* we shall simply mean that an agent always optimizes given his updated beliefs.

The question we pose is whether there are any beliefs that lead to good prediction when players act rationally. In other words, can rational players "learn" the process generating the data if they watch the process long enough and are sufficiently patient? We are going to show that for some games – including some very simple 2×2 games – there is no such scheme no matter how patient the players are.

3. Good prediction

To state this result precisely we need to pin down the concept of good prediction. In general, let z^t be a random variable whose distribution at time t is given by some probability distribution $P(z^t = z) = p^t(z \mid z^1, \ldots, z^{t-1})$ that may depend on realizations in previous periods. Assume that at the beginning of period t agent i knows the sequence $z^1, \ldots, z^{t-1}$ and on that basis predicts that the probability that $z^t = z$ next period is $q_i^t(z \mid z^1, \ldots, z^{t-1})$. Unlike most of the literature we shall treat the belief formation process as a black box: all we assume is that a prediction function $q_i^t(z \mid z^1, \ldots, z^{t-1})$ is defined for every agent i, every time period t, and every possible initial sequence $z^1, \ldots, z^{t-1}$.

Assume now that z^t takes its values in a finite set Z. (In the case of a finite game, z^t takes its values in the finite action space X.) Then we can define agent i's *error in prediction at time t* to be

$$\sum_z [q_i^t(z \mid z^1, \ldots, z^{t-1}) - p^t(z \mid z^1, \ldots, z^{t-1})]^2.$$

We say that agent i is a *good predictor* if the mean square error in i's prediction goes to zero with probability one, that is, if for almost all realizations of the process $(z^1,\ldots,z^t,\ldots)$,

$$\lim_{T\to\infty}(1/T)\sum_{i=1}^{T}\sum_{z}[q_i^t(z\,|\,z^1,\ldots,z^{t-1})-p^t(z\,|\,z^1,\ldots,z^{t-1})]^2=0. \tag{1}$$

Notice that this definition allows for occasional large errors in prediction, so long as they eventually become infrequent. This gives i scope to make substantial mistakes early on in the learning process, and even to continue to make substantial mistakes from time to time throughout the learning process. Both of these assumptions seem natural in a model of learning.

4. The Kalai–Lehrer framework

In a seminal paper, Kalai and Lehrer (1993) exhibit conditions under which players can in fact learn to predict their opponents' behavior with increasing accuracy. To illustrate the concept, consider a simple coordination game in which both parties get payoff 1 if they choose the same direction (left or right) and zero otherwise:

	L	R
L	1,1	0,0
R	0,0	1,1

(2)

In one-shot play this game has three equilibria: both play L, both play R, and both randomize fifty–fifty. In repeated play there are many other equilibria, but perhaps the simplest are the constant strategies: always play R, always play L, or always play fifty–fifty. Denote these repeated-game strategies by $\bar{\mathrm{R}}$, $\bar{\mathrm{L}}$, and $\bar{\mathrm{F}}$, respectively. Assume that each player begins with a prior belief that puts positive probability on each of these three possibilities and no others. After each play, the players choose best responses to their updated beliefs. It is easy to arrange the beliefs so that they begin by playing, say, (R, L) in period one. (Here agent 1's choice is listed first, agent 2's second.) After seeing this outcome, agent 2 can now eliminate $\bar{\mathrm{L}}$ as a possible strategy for agent 1, and agent 1 can eliminate $\bar{\mathrm{R}}$ as a possible strategy for agent 2. In period 2, agent 1's best response is L and agent 2's best response is R. At this point they can eliminate both $\bar{\mathrm{R}}$ and $\bar{\mathrm{L}}$ as possible strategies of the opponent, and each concludes that the opponent is playing $\bar{\mathrm{F}}$.

Notice first that neither player actually is playing $\bar{\mathrm{F}}$, because in periods 1 and 2 they did not play F (their choices were deterministic). Notice further that *any* choice of strategy from period 3 on is a best response to the belief that the opponent is playing $\bar{\mathrm{F}}$, but the only best responses that are consistent with good prediction are those that are "close" to $\bar{\mathrm{F}}$ in the sense of (1). (Such strategies are, however, quite unsatisfactory as solutions to the coordination game.)

More generally, although neither player's initial belief need put positive probability on the actual strategy of the opponent (because of "start-up" problems early in the game or lapses later in the game), it may still the case that they *eventually* learn to predict if

their best responses are suitably aligned vis-à-vis the beliefs of the other. The technical condition that guarantees this is *absolute continuity*: any event (collection of infinite play sequences) that has positive probability under the players' best response strategies should have positive probability under their beliefs. Adapting a theorem of Blackwell and Dubins (1962) on the merging of opinions, Kalai and Lehrer (1993) show that this condition guarantees that all players will be good predictors with probability one.

5. Critiques of the Kalai–Lehrer framework

The preceding example should make clear, however, that absolute continuity will only hold if, among all best response strategies, players choose particular ones that are serendipitously aligned with the opponents' beliefs. Various authors have argued that this seriously limits the applicability of the Kalai–Lehrer result. For example, Miller and Sanchirico (1997) show that for a suitably defined distribution on the players' belief spaces, the probability is zero that any two beliefs drawn at random will satisfy the absolute continuity condition.

An alternative critique is due to Nachbar (1997, 1999, 2001), who proposes conditions on the support of players' beliefs that capture the idea that they are genuinely uncertain about the strategies their opponent will use. These conditions include the idea that if one believes that a strategy is possible for an opponent, then one should believe that strategic variants of similar complexity should also be possible. The beliefs should also be mutually consistent in the sense that they include at least one ε-best response of the opponent to his beliefs, for every small $\varepsilon > 0$. Nachbar shows that for many two-person games – including the simple coordination game given above – there is no set of beliefs satisfying these conditions (plus several others) that results in good prediction by both players.

6. Learning rules

The impossibility results described above are framed in terms of "plausible" or "probable" beliefs about the opponent's strategic behavior. In this paper we shall remain entirely agnostic about the structure of agents' beliefs. Instead we shall view the learning problem as one of making conditional forecasts. In general, let $x^t \in X$ denote the play in period t and let $h^{t-1} = (x^1, \ldots, x^{t-1})$ be the *history through* $t-1$. Let Δ_i denote the set of probability mixtures over i's possible actions. Define a *forecasting rule* for agent i to be a function f_i that maps each initial history h^{t-1} to a conditional probability distribution over the opponent's choice of action next period, $f_i(h^{t-1}) \in \Delta_{-i} = \prod_{j \neq i} \Delta_j$. Define a *response rule* for agent i to be a probability distribution $g_i(\cdot \mid h^{t-1}, f_i) \in \Delta_i$, where $g_i(X_i \mid h^{t-1}, f_i)$ is the probability of choosing x_i next period, given the history to date and conditional on the forecast generated by the forecasting rule f_i. A *learning rule* for agent i is a forecasting rule f_i together with a response rule g_i. (Jordan (1993) was the first to pose the learning problem in these terms.)

A response rule is *optimal* for a given forecasting rule if, in each period, the agent puts positive probability only on actions that maximize expected discounted payoffs from that time on, as computed via the forecasting rule. We say that an agent is *rational* if, at each point in time, his response rule is optimal given his forecasting rule.

Bayesian updating fits into this framework, because the updating of beliefs produces a one-step-ahead forecast at each stage of the game. Conversely, any forecasting rule can be interpreted as a Bayesian scheme in the following way: simply define the prior probability of any play path to be the product of the conditional probabilities generated by the forecasting rule along that path. It follows that forecasting rules are equivalent to Bayesian updating schemes. Nevertheless, framing the problem in terms of forecasting rules is useful because we do not need to address the issue of whether some beliefs are more "reasonable" than others, as is common in Bayesian reasoning. *Any* priors will do for our analysis, so long as they do not directly contradict the information available to the players.

7. Robust learning rules

We now ask whether there exist any learning rules that are *robust*, i.e., that lead to good prediction irrespective of the realized payoffs. By insisting that learning be robust, we avoid the temptation to cook the learning procedure to suit a particular combination of payoffs. For example, in our discussion of the Kalai–Lehrer framework applied to game (2), we relied on both players knowing that $\bar{R}$, $\bar{L}$, and $\bar{F}$ are equilibria of the game. But in real-world games the players may be completely ignorant of their opponents' payoffs, and thus they may have no way of knowing what strategies constitute an equilibrium. Consider, for example, the following game in which the prizes are consumption goods:

The Soda Game

	L	R
L	coke, coke	sprite, seven-up
R	seven-up, sprite	pepsi, pepsi

(3)

Unless the players' preferences are aligned in a particular way, randomizing fifty–fifty will not be a Nash equilibrium of the one-shot game. In fact the players may not even know what *kind* of a game this is. For example, if player 1 likes coke and pepsi better than seven-up and sprite, whereas player 2's preferences are the reverse, this is like a matching pennies game which has a unique interior Nash equilibrium in the one-shot version. But if both like dark drinks better than light drinks then it is a coordination game. Thus, without more information about the opponent's preferences, neither player is sure what game they are playing or what its Nash equilibria are. From a practical point of view this is exactly the kind of situation where the players *need* to learn. We are going to show, however, that this is precisely the kind of situation where rational learning is effectively impossible.

8. Statement of the main result

We shall first formulate our result in terms of a concrete case, then show how it generalizes. Consider the following *perturbed version of matching pennies*:

$$\begin{array}{cll} & \text{L} & \text{R} \\ \text{L} & 1+\alpha_{11}, 1+\beta_{11} & -1+\alpha_{12}, 1+\beta_{12} \\ \text{R} & -1+\alpha_{21}, 1+\beta_{21} & 1+\alpha_{22}, 1+\beta_{22} \end{array} \tag{4}$$

Assume that the eight random variables $\{\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}, \beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}\}$ are drawn i.i.d. from a continuous density whose support is a small interval $[-\lambda/2, \lambda/2]$, where $\lambda > 0$ is the *range* of the perturbations. Call the resulting joint distribution ν.

Theorem 1 (Foster and Young, 2001). *Let G be perturbed matching pennies with perturbation distribution ν and perturbation range $\lambda > 0$. Given any rational learning rules and any discount factors $0 \leqslant \delta_1, \delta_2 < 1$, if λ is small enough the* ν-probability *is one that on almost all play paths one or both players fail to learn to predict in the sense of* (1).

Corollary 1. *For all sufficiently small λ there exist no priors such that rational Bayesian players learn to play perturbed matching pennies with positive probability.*

Suppose instead that the payoff errors are drawn i.i.d. from a density whose support is the whole real line (e.g., a normal distribution). Then with positive probability the realized payoffs fall into the range of payoffs covered by Theorem 1. Hence learning fails to occur *with positive probability*. Indeed the same holds for any size finite game that is perturbed by i.i.d. normally distributed random errors: there is always a range of payoffs such that some 2×2 subgame dominates the other strategies (for both players), and this 2×2 subgame has payoffs close to those of matching pennies. Thus we have the following:

Corollary 2. *Let G be any finite game that is perturbed by i.i.d. normally distributed random errors. Given any rational learning rules and any discount factors $0 \leqslant \delta_1, \delta_2 < 1$, there is a positive probability that on almost all play paths one or both players fail to learn to predict in the sense of* (1).

The force of these results is that they hold for *any* forecasting rules and *any* strategies that are optimal given the forecasts. These may include very sophisticated strategies. For example, an agent might try to estimate the opponent's utility function by "testing" her reactions to different sequences of actions. For a sufficiently patient player this would be rational if the loss of payoff while gathering information were outweighed by payoff gains later on from exploiting the information. The above results show that, no matter how forward-looking the players are, no such strategy guarantees good prediction.

Notice, however, that the theorem does not claim that *both* players fail to learn to predict. For example, in matching pennies, suppose that player 1 predicts that player 2 will play H every period. Then 1's best response is to play H every period. Suppose further that player 2 predicts that player 1 will play H every period. Then 2's best response is to choose T every period. Thus 2's prediction is always correct but 1's is not. Notice further that the theorem does not say that the players are systematically wrong in their predictions *given the information available*. On the contrary, if they begin with correct priors about the distribution of the opponents' payoffs, and if their strategies constitute a Bayesian Nash equilibrium of the underlying game of incomplete information, then at each point in time their predictions will be correct *conditional* on the distribution of types revealed by play so far. Moreover there are circumstances under which these posteriors converge to a Nash equilibrium of the one-shot game (Jordan, 1991; Nyarko, 1998). Nevertheless, the predictions are almost surely not correct against a *given* opponent.

9. The impossibility of learning equilibrium

The problem of good prediction is related to the problem of learning to play equilibrium in a repeated game, but it is not quite the same thing. Intuitively, if every player is making good predictions and is choosing an optimal strategy given his prediction, then play should be close to Nash equilibrium play. It is conceivable however, that even if players are not making good predictions, their responses to their (false) beliefs could still be close to equilibrium. Here we shall argue that this cannot happen in games like perturbed matching pennies that are close to being zero-sum and have only interior Nash equilibria.

The idea of the proof generalizes an argument of Jordan (1993) to the case of nonmyopic players, and runs as follows. Under the assumption of perfect rationality, a player who is not exactly indifferent (given his forecast of the opponent's future behavior) will choose a pure action next period. However, there are ν-almost no payoff realizations that make a player *exactly* indifferent at a given point in time for a given forecast. Since a forecast at time t can condition only on the observed histories up through $t-1$, which are finite in number, there are ν-almost no payoff realizations such that this player is indifferent at time t under *any* forecasting rule. Since the number of periods is countable, there are ν-almost no payoff realizations for which this player would be indifferent at any time under any forecasting rule. In other words, no matter what the players' forecasting rules are (equivalently no matter what their initial beliefs), the ν-probability is one that in every period *both* players will almost surely choose pure actions.

Suppose now that the game is perturbed matching pennies. Given any discount rates less than unity, if the perturbation range λ is sufficiently small, then the repeated game has only mixed equilibria. In fact, the probability placed on each action in each period is bounded away from zero. Thus, in every period, the ν-probability is one that neither player is playing close to any Nash equilibrium of the repeated game. This result may be stated more formally as follows.

Theorem 2 (Foster and Young, 2001). *Let G be perturbed matching pennies with perturbation distribution ν and perturbation range $\lambda > 0$ as in* (4). *Given any rational learning rules and any discount factors $0 \leqslant \delta_1, \delta_2 < 1$, if λ is small enough the ν-probability is one that in almost every time period both players' strategies are far from any Nash equilibrium of the repeated game.*

References

Blackwell, D., Dubins, L., 1962. Merging of opinions with increasing information. Annals of Mathematical Statistics 38, 882–886.

Blume, L.E., Easley, D., 1998. Rational expectations and rational learning. In: Majumdar, M. (Ed.), Organizations with Incomplete Information: Essays in Economic Analysis. Cambridge University Press, Cambridge, UK, pp. 61–109.

Foster, D.P., Young, H.P., 2001. On the impossibility of predicting the behavior of rational agents. Proceedings of the National Academy of Sciences of the USA, Vol. 98 (22), pp. 12848–12853.

Fudenberg, D., Ellison, G., 2000. Learning purified mixed equilibria. Journal of Economic Theory 90, 84–115.

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

Grandmont, J.-M., 1998. Expectations formation and stability of large socioeconomic systems. Econometrica 66, 741–781.

Harsanyi, J., 1973. Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points. International Journal of Game Theory 2, 1–23.

Jordan, J.S., 1991. Bayesian learning in normal form games. Games and Economic Behavior 3, 60–91.

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

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

Kalai, E., Lehrer, E., 1993. Rational learning leads to Nash equilibrium. Econometrica 61, 1019–1045.

Miller, R.I., Sanchirico, C.W., 1997. Almost everybody disagrees almost all the time: The genericity of weakly merging nowhere. Department of Economics Discussion paper 9697-25, Columbia University, New York.

Miyasawa, K.,1961. On the convergence of the learning process in a 2×2 non-zero-sum two-person game. Economic Research Program, Research Memorandum no. 33, Princeton University, Princeton, NJ.

Monderer, D., Shapley, L.S., 1996. Fictitious play property for games with identical interests. Journal of Economic Theory 68, 258–265.

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

Nachbar, J.H., 1999. Rational Bayesian learning in repeated games. Discussion paper, Department of Economics, Washington University, St. Louis, MO.

Nachbar, J.H., 2001. Bayesian learning in repeated games of incomplete information. Social Choice and Welfare 18, 303–326.

Nash, J., 1950. Non-cooperative Games. Ph.D. Dissertation, Princeton University, Princeton, NJ.

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