Will reasoning improve learning?

Nicolaas J. Vriend*

Universitat Pompeu Fabra, Dept. Economics, Ramon Trias Fargas 25-27, 08005 Barcelona, Spain

Abstract

Adding some reasoning capabilities to a naive reinforcement learning model, we get outcomes farther away from the fully introspective game-theoretic approach. This is caused by an asymmetry in the information agents can deduce from their experience, leading to a bias in their learning process. ©1997 Elsevier Science S.A.   © 1997 Elsevier Science S.A.

Keywords: Ultimatum game; Game theory; Reasoning; Reinforcement learning

JEL classification: C7; D8

1. Introduction

One of the games most extensively studied in the literature during the last decade is the Ultimatum Game. The reason that this game is so intriguing seems to be that the game-theoretic analysis is straightforward and simple, while the overwhelming experimental evidence is equally straightforward but at odds with the game-theoretic analysis (see, e.g., Güth et al., 1982; Güth, Tietz, 1990, or Thaler, 1988).

Here is the basic form of the Ultimatum Game: there are two players, A and B, and a pie. Player A proposes how to split the pie between herself and player B. Upon receiving player A's proposal, player B has two options. First, to accept the proposal, which will then be carried out. Second, to reject it, after which both get nothing. Many variants of this basic setup have been considered in the literature, but the details are not essential for this note. There are many Nash equilibria in this game. Every strategy for player A combined with any strategy for player B that accepts that offer but rejects all worse offers is one. But, considering the extensive form variant of the game, there is a unique subgame perfect equilibrium: player A offers the minimum piece, and player B accepts that.[1]

Empirical evidence shows time and again that this is not what happens in most games played in the laboratory. Players A usually offer somewhat less than half the pie to players B. And players B

---

*Tel.: 34 9 3 542 2718; e-mail: vriend@upf.es

[1]Strictly speaking, in case it is a discrete choice problem including zero, there are two subgame perfect equilibria, with player A offering either zero or the smallest possible strictly positive piece to player B.

usually reject small offers. Concerning player A's behavior, there are two main explanations for the anomaly offered in the literature. First, some argued that fairness and reciprocity considerations are the force driving players A to offer more than the game-theoretic approach would suggest (see, e.g., Forsythe et al., 1994). A second explanation found in the literature is that players A are basically following an adaptive, best-reply seeking approach to the behavior of players B. In a multi-period setup where players played the game repeatedly but against different players, some papers showed how it can happen that players A 'unlearn' to play the perfect equilibrium strategy before players B learn that they should play their perfect equilibrium strategy. And once players A do not play that strategy anymore, players B will never learn to play theirs. See Roth, Erev (1995) who apply a reinforcement learning approach, while Gale et al. (1995) use replicator dynamics.[2]

In both explanations we see that the (strategic) interactions between the behavior of players A and players B may lead to complications as far as the analysis is concerned.[3] In this note we will concentrate on the behavior of players A, and we assume for the moment that players B play the perfect equilibrium strategy. What we will show is that if we start with simple stimulus-response agents, learning through naive reinforcement (see next section), and then attribute to them some reasoning capabilities, we may get outcomes that are not closer but farther away from the fully introspective game-theoretic approach. The cause of this is the following: there is an asymmetry in the information that agents can deduce from their experience, and this can lead to a bias in their learning process. It is a simple theoretical point, but I have not seen it in the literature. In Section 5 we will discuss its implications for the modeling of learning behavior. Before reaching those concluding remarks, Section 2 will explain what reinforcement learning is, and the difference between naive reinforcement learning, and reinforcement learning plus reasoning. Section 3 shows how an information asymmetry leads to a distortion of the learning process, and Section 4 discusses and relaxes the assumptions made in Section 3.

2. Actual and virtual reinforcement learning

Standard game theory is completely based on introspection. Given the agents' reasoning capabilities concerning the game, the payoffs and the reasoning of the other players, actually playing games seems inessential.[4] One can imagine a spectrum of modes of individual behavior that differ in the way and extent to which reasoning plays a role. The fully introspective players of standard game theory would be at one of the extremes of this spectrum. At the opposite end of the spectrum we would find boundedly rational agents without any reasoning capabilities. They act like simple stimulus-response creatures that can be conditioned. They do not understand what game is being played, perhaps not

---

[2]For a detailed comparative analysis of replicator dynamics and reinforcement learning, on which we focus in this note, see Börgers, Sarin (1995).

[3]For example, observing that the behavior of players A eventually converges to a best-reply to players B cannot be the complete story about players A. What is a best-reply after a coevolutionary process during which players B adapt to the behavior of players A as well, depends upon how players A have been adapting during that process.

[4]Here is a verified anecdote illustrating this. A famous economist who had published a book about the game of Bridge, met a colleague who expressed his surprise about him being a player of this game. "I don't play Bridge", was his reaction, "I wrote a book about it."

even that a game is being played, nor do they understand how payoffs are determined, or what utility is. But they recognize it when they get it. These agents behave adaptively to their environment in that they experiment by trying actions, and actions that led to better outcomes in the past are more likely to be chosen again in the future. This type of behavior is nowadays known as 'reinforcement learning'.

There is a family of stochastic dynamic models of such individual behavior in the scientific literature, for which different backgrounds can be distinguished. The idea was first developed in the psychological literature. See especially Hull (1943), and Bush, Mosteller (1955), on which Cross (1983) is based. Much later, reinforcement learning was independently reinvented twice as a machine learning approach in computer science. See, e.g., Barto et al. (1983), and Sutton (1992) for a survey of an approach called reinforcement learning. The other reinforcement learning approach in computer science is known as Classifier Systems. See Holland (1975) for early ideas on this, or Holland et al. (1986) for a more elaborate treatment. In the economics literature reinforcement learning became better known only recently through Roth, Erev (1995). For expositional convenience, in this note we will follow the Roth, Erev approach to reinforcement learning.[5]

In computer science the main reason for the interest in reinforcement learning is its success in performing difficult tasks. In psychology reinforcement learning is mainly judged on its success in explaining empirical evidence of subjects in experiments. And the fact that this kind of model seems consistent with the robust properties of learning that have been observed in the large experimental psychology literature is also the main line of thought in the application of reinforcement learning models in the experimental economics literature. Following Roth, Erev (1995), our main interest in reinforcement learning, is to use it as a benchmark case.[6] It is the simplest possible type of learning model, assuming the most limited cognitive capabilities, no reasoning or introspection at all, while still allowing for learning. Hence, besides the standard game-theoretic approach, this class of models serves as a second benchmark.

Consider the Ultimatum Game. Assume player A has no understanding of game theory, the structure of the Ultimatum Game, or the behavior of players B. Player A is a boundedly rational agent who behaves adaptively to her environment. She simply tries actions, and is in the future more likely to repeat those actions that led to high payoffs in the past than those that led to low payoffs. One can imagine player A as playing with a multi-armed bandit, where different arms might give different payoffs; and player A does not know at the start which is the best arm to pull. This is the basic reinforcement learning approach (see Roth, Erev, 1995): At time $t = 1$ each player has an initial

---

[5]Reinforcement learning differs from some other important recent models of dynamics with experimentation in the economics literature (see, e.g., Ellison, 1993; Kandori et al., 1993), or Young, 1993). In those evolutionary dynamic models, adaptive behavior is basically a one-step error correction mechanism. The agents have a well-specified model of the game, they can reason what the optimal action would be, given the actions of the other players, completely independent from any payoff actually experienced, and they play a best-response strategy against the frequency distribution of a given (sub-)population of other players. The evolutionary dynamics consist of a coevolutionary adaptive process, players adapting to each others' adaptation to each other . . . , plus experimentation in the form of trembling. The very first task for a reinforcement learning algorithm, on the other hand, is to learn what would be good actions. The agents do not have a well-specified model of their environment, and they do not know which action would be the best response. See Vriend (1994) for a more extensive discussion of these issues.

[6]The question they address is, for which classes of games can the experimental results be explained, 'on average', by such a benchmark model, which they call a 'low rationality' model (see also Erev, Roth, 1995).

propensity to play his kth pure strategy given by some real number $q_k(1)$. If a player plays his kth pure strategy at time t, and receives a payoff of z, then the propensity to play strategy k is updated by setting $q_k(t + 1) = q_k(t) + z$, while for all other pure strategies j, $q_j(t + 1) = q_j(t)$. The probability that the player plays his kth pure strategy at time t is $p_k(t) = q_k(t) / \Sigma q_j(t)$, where the sum is over all of the player's pure strategies j. We will call this naive reinforcement learning, or learning through actual reinforcement. Notice that the reinforcement learning model can be summarized by the following three steps. First, choose an action using the probabilities given above. Second, deduce information about payoffs from your experience. And third, update the propensities to choose actions.

There are two natural ways to extend this basic reinforcement learning mechanism. Both concern only the second of these three steps listed above, i.e., the information gathering, and both have been used in the literature. First, instead of learning only on the basis of his own actions and payoffs, an agent can learn from observing other agents' actions and payoffs, or from a supervisor teaching him about such a hypothetical agent. He can, then, reinforce his propensities to choose an action as if he had tried those actions, and experienced the payoffs himself. This is called learning through vicarious reinforcement (see Bandura, 1977 or Lin, 1992).[7] Second, instead of learning on the basis of actions and payoffs actually experienced by himself or others, a human player can learn by reasoning about counterfactuals. That is, he can use his imagination concerning unchosen actions and foregone payoffs, and he can, again, reinforce his propensities to choose actions as if he had actually tried those actions, and experienced the corresponding payoffs. We will call this learning through virtual reinforcement (see Holland, 1990).

We will not pursue the track of vicarious reinforcement here, but will concentrate on virtual reinforcement, because that got some attention in the economics literature recently.[8] The basic reinforcement learning model as used by Roth, Erev (1995) has been criticized frequently as being 'too dumb' to model human players, since human players are simply smarter than rats. Through introspection an intelligent player can deduce more information from his experience than only the payoff for the action actually chosen. And in particular, it is postulated sometimes, that in any sensible learning model players should use all available information (see, e.g., Camerer, Ho, 1996).

Clearly, virtual reinforcement requires that one relaxes the restriction on player A's cognitive capabilities, and assumes that player A has some reasoning capabilities. In the Ultimatum Game the arms of the bandit can be ordered naturally from low offers to high offers. For example, when an offer x has been accepted by player B, player A can reason that player B would have accepted also all offers $x' > x$. Or when an offer x has been rejected, player A can reason that offers $x'' < x$ would have been rejected as well. Thus, player A can imagine her foregone payoffs for those unchosen strategies. This is information available to player A. The reasoning process to gather the additional information concerning foregone payoffs seems correct as such,[9] and adding the virtual updating will speed up player A's learning process. So far so good. The point, however, is that there is a second effect of the virtual updating, which is the following. There is an asymmetry in the information obtained by

---

[7] Notice that this is more sophisticated than imitation. It is only the propensity to choose an action that is updated on the basis of an other agent's action and consequence.

[8] Two recent examples are Stahl (1996), and Camerer, Ho (1996). Since both limit their attention to games in normal form in which there is no information asymmetry, they could avoid the issues raised in this note.

[9] We want to stress once more at this point, that the reasoning allowing for virtual reinforcement learning concerns only the second of the three steps listed above. That is, the reasoning as such concerns only the gathering of information, and not directly the updating of propensities, or the choice of actions.

Table 1
Basic assumptions

i)	The pie has size Π.
ii)	The only possible offers x to player B are $0, 1, 2, \ldots, \Pi$.
iii)	Player B plays the perfect equilibrium strategy, and accepts every offer.
iv)	Player A tries every action equally often, say n times.
v)	In case of acceptance the payoff to player A is simply $\Pi - x$. In case of rejection it would be 0.
vi)	Reinforcement takes place as explained above: the payoff realized by playing offer x is added to the propensity to choose offer x.
vii)	Only actual reinforcement learning takes place.

playing strategy x. If offer x is accepted, player A can reason what player B would have done with offers $x' > x$, but player A does not get information about what player B would have done with offers $x'' < x$. Similarly, if offer x is rejected, player A can reason what player B would have done with offers $x'' < x$, but not what player B would have done with offers $x' > x$.[10] In the next section, we will analyze what the postulate that players should use all the available information would imply for the reinforcement learning process, and whether including virtual updating in a reinforcement learning model is such a self evident improvement.

3. Effect of the information asymmetry

The strategy of this note is to start with an example, making some extremely simplifying assumptions (Table 1) in order to make our point as transparent as possible. In Section 4 the basic assumptions of this section will be discussed and relaxed. We want to stress here that the effect of the information asymmetry is independent of the behavior of player B.

Proposition 1. Under assumptions i) to vii), the most reinforced offer will be $x = 0$.

Proof. After player A has tried each possible offer n times, the propensity to choose a certain offer x will have increased with n times a payoff of $(\Pi - x)$. Formally, the total reinforcement for any offer x is given by $r(x) = n \cdot (\Pi - x)$, which has a maximum at $x = 0$.

Next, we assume that player A reasons about counterfactuals as well. Hence we keep assumptions i) to vi), but instead of vii), we make the assumption given in Table 2.

Proposition 2. Under assumptions i) to vi) plus vii'), the most reinforced offer will be offer x satisfying $x > (\Pi - 2)/2$ while $(x - 1) < (\Pi - 2)/2$.

Table 2
Assumption virtual reinforcement

vii')	Player A reasons that player B's behavior has a reservation value property. If offer x is accepted, offers $x' > x$ would have been accepted too. Therefore, player A applies both actual and virtual reinforcement.

[10]While intuition might suggest that the second asymmetry offsets the first, it turns out they reinforce each other (see below).

Proof. After player A has tried each possible offer n times, the propensity to choose a certain offer x will have increased with n times a payoff of $(\Pi - x)$ through the actual reinforcement. In addition, the virtual reinforcement for an offer x will be: x times this actual reinforcement (we will give the intuition below). Hence, the total reinforcement for offer x is: $r(x) = n \cdot (x + 1) \cdot (\Pi - x)$. Taking first differences gives: $r(x + 1) - r(x) = n \cdot (\Pi - 2x - 2)$. Hence, $\Delta r(x) < 0$ if $x > (\Pi - 2)/2$.

Notice that the only difference with the case of actual reinforcement only is the factor $(x + 1)$. Since $(x + 1) \geq 1$, all propensities are updated more often now. That is, the virtual updating in addition to the actual updating makes the learning process faster. But there is also a second effect: the information asymmetry. Table 3 helps to explain this.

Consider the first line in Table 3, i.e., the offer $x = 0$. The propensity of this offer will be updated only when $x = 0$ has actually been accepted. If other offers $x > 0$ are accepted, player A will not know whether $x = 0$ would have been acceptable too. Next, consider the offer $x = 1$. This propensity will be reinforced actually when $x = 1$ has been accepted, and in addition virtually when $x = 0$ has been accepted. One can continue this up to the largest possible offer $x = \Pi$. If $x = \Pi$ has been accepted by player B, player A will update this propensity actually. But player A will also reason after every offer $x < \Pi$ that an offer $x = \Pi$ would have been acceptable as well. Hence, $x = \Pi$ will be updated $(\Pi + 1)$ times as often as $x = 0$. In other words, the direct effect of the information asymmetry is that some offers are reinforced more often than other offers, although they are actually tried the same number of times. What does this mean for the final propensities to choose these offers? As shown in proposition 2, applying both actual and virtual reinforcement, the most reinforced offer x will be slightly below the 50–50 offer. The example in Fig. 1 for $n = 1$, and $\Pi = 6$ illustrates this. The graph shows the difference between the two forms of learning. In one case only actual reinforcement takes place, while the other case allows for virtual reinforcement as well. In the first case the most reinforced offer is 0, while in the second case the most reinforced offers are 2 and 3.

Recall that reinforcement learning is a directed random search process; trial-and-error based on the player's experience. What actions a player chooses depends upon two things. First, how well she knows the payoff of certain actions. Second, how good the payoffs generated by those actions are. Of course, if the payoff of a certain action is not good, it is unlikely that a player will get to know that action well. But now with the reasoning about unchosen actions, players get additional information about payoffs 'for free'. As we showed, this information is asymmetric, and leads to a bias in the learning process.

Notice that the fact that offers slightly below half the pie size are most reinforced has nothing to do

Table 3
Actual plus virtual reinforcement

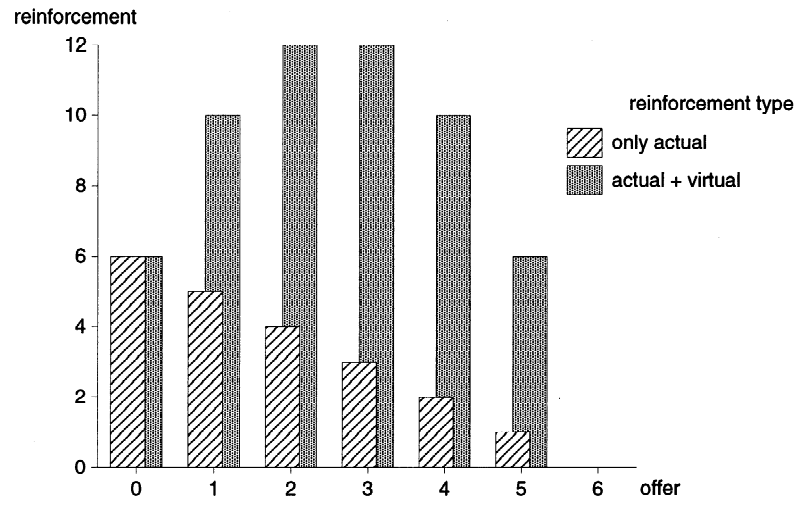| Offer x to player B	Total reinforcement $r(x)$
0	$n \cdot (0 + 1) \cdot (\Pi - 0)$
1	$n \cdot (1 + 1) \cdot (\Pi - 1)$
2	$n \cdot (2 + 1) \cdot (\Pi - 2)$
$\ldots$	$\ldots$ $\ldots$ $\ldots$
$\ldots$	$\ldots$ $\ldots$ $\ldots$
Π	$n \, (\Pi + 1) \, (\Pi - \Pi)$

Fig. 1. Outcomes of two types of reinforcement process.

with fairness or reciprocity.[11] Player A is assumed to be so boundedly rational that such concepts form no part of her cognitive capabilities. It has also nothing to do with best-responses to player B's behavior. In our exposition thus far player B accepts all offers made to him. The fact that offers slightly below half the pie size are most reinforced is due here entirely to the asymmetry in the information player A can deduce from player B's acceptances.

4. Relaxing the assumptions

As said above the simplifying assumptions made thus far were made in order to highlight the information asymmetry effect as such. We will now discuss these assumptions, and show how they can be relaxed without cancelling this effect.

ad i) The are some papers in the literature in which the pie size Π is variable and not known with certainty (see, e.g., Mitzkewitz, Nagel, 1993). Such a relaxation will not annul the effect of the asymmetry as such, but only change the numbers somewhat.

ad ii) Assuming that only integer offers up to Π can be made is without much loss of generality. As long as we assume there exists a minimum slice size, all that is involved is re-scaling that indivisible unit to 1.

ad iii) Assuming player B accepts all offers helps to highlight the information asymmetry effect. In general player B will not accept every offer; not even every offer with equal probability. But this does not deny the asymmetry effect. And if, what seems most plausible, lower offers are rejected more often than higher offers, it only reinforces that effect; moving away from the perfect equilibrium.

---

[11]From a strictly formal point of view, the expressions for the total reinforcement in Table 3 could be re-interpreted as a preference for fairness. The second term measures the payoff derived from the consumption by player B, and the third term the payoff from player A's own consumption, putting slightly more weight on the latter.

Adding the possibility of refusals, one should notice that there is an analogous asymmetry effect with refusals. If an offer x has been refused, player A can reason so would have been offers $x'' < x$, but not necessarily offers $x' > x$. The standard assumption in the literature is that in case of refusal the payoff and hence the reinforcement is zero. Hence, this does not influence our result. If we want to take negative reinforcement in case of refusals into account as well, it only reinforces our argument, since, as can be easily checked, the effect of the information asymmetry in case of rejections works in the same direction as the one discussed for the case of acceptance (against the lower offers).

ad iv) Player A will not try every strategy equally often, but will adapt towards the more reinforced ones. Consider Fig. 1, which gives the reinforcement after trying each action x once. If, for example, from then on the actions $x = 0$ and $x = 1$ are tried less often than $x = 2$ or $x = 3$ the distribution of future reinforcement will get skewed towards the right. Hence, this relaxation will only reinforce the asymmetry effect. Making small offers to player B will get reinforced even less and less.

ad v) Reinforcement to player A is not necessarily linear in the payoff in pie slices, but this seems a harmless assumption. Changing it would change the numbers somewhat, but not cancel the asymmetry effect.

ad vi) Suppose we still do not update offers $x'' < x$, but instead of making reinforcement additive, each action's propensity is averaged over the number of times it has been reinforced. Let us see what players are supposed to do when averaging reinforcement instead of simply adding up. First, one should divide each action's propensity not only by the number of times it has actually been played, but also by the number of times it has virtually been played and reinforced through the reasoning process. Second, all actions $x'' < x$ should be considered as not reinforced rather than as reinforced with factor 0. If the latter were done, each turn all propensities would be updated, and hence the averaging would make no difference. Notice that in order for the players to accomplish this one needs to attribute them higher reasoning skills. Not only do they need to reason about unchosen actions they also need to realize the existence of the asymmetry effect identified in this note, and they need to compensate correctly for it in their reinforcement. Hence, implementing this relaxation is easy from a technical point of view, but the empirical question is how realistic it is for human players.

ad vii') The result depends upon player A presuming player B's behavior has a reservation value property. One could relax this. As long as, after an acceptance of offer x, offers $x' > x$ are presumed to be more likely to be accepted than offers $x'' < x$, this does not deny the effect of the information asymmetry on the virtual updating as such, but only changes the numbers somewhat. If all offers are updated every time, then the asymmetry effect is avoided if and only if the agent has correct estimates about the probabilities that such offers would have been accepted. However, the idea behind an evolutive approach (instead of the fully introspective game-theoretic approach) was exactly that the agents do not know such things to start with.[12]

5. Concluding remarks

Before concluding, let me stress what this note is not about. First, this note is not meant to present virtual reinforcement as an improvement of the naive reinforcement learning model. The idea of

---

[12]Moreover, relaxation here does not matter as long as the reinforcement of offers $x'' < x$ is biased downwards as a result of the information asymmetry.

virtual reinforcement as such is trivial. And as argued above, even a priori there would be good reasons not to consider it an improvement of the basic reinforcement model, in as far as that model was meant as a benchmark. This note is about the information asymmetry, and its effect on the outcome of the reinforcement process once virtual updating would be considered. Second, whether players actually update virtually or not is an empirical question. But that is not our point. With respect to the example of the Ultimatum Game in extensive form, this note addresses the logical issue: whatever the extent to which players learn through virtual reinforcement, the effect of the information asymmetry would be a bias away from the perfect equilibrium strategy.[13] And if virtual reinforcement turns out to be relatively less utilized than actual reinforcement, this means that one could expect outcomes closer to the perfect equilibrium.

The exact contents of the propositions in this note are linked to the specific assumptions made. But the analysis of these propositions and assumptions leads to a paradoxical observation which is of more general relevance. Take standard game theory and basic reinforcement learning, and consider them as two benchmark cases with respect to reasoning in a spectrum of possible modes of individual behavior. The first based entirely on introspective processes, the latter abstracting completely from such processes. Adding reasoning about unchosen actions, and hence virtual updating, to the basic reinforcement learning model brings you closer to the fully introspective game-theoretic approach as far as the modeling is concerned. But it leads to outcomes that are farther away from the game-theoretic prediction. Hence, reasoning does not necessarily improve learning. This result leads to the conjecture that the asymmetry in the information agents can deduce from their experience is of more general importance when boundedly rational agents are supposed to be endowed with some reasoning capabilities; in particular in sequential-move games.[14] The Ultimatum Game might, then, turn out to be merely a first example. In any case, the example shows that it would be wrong to presume that there is a smooth, continuous, monotonic function from a one-dimensional variable called reasoning to performance. A possible lesson is that one has to be cautious not only with allegedly ad hoc models of learning and adaptive behavior, but in particular also with so-called 'self evident improvements' thereof.

Acknowledgments

---

[13]At this point we can only recall that offers somewhat below 50% are common in laboratory experiments. Whether the effect analyzed here might be an explanation for that evidence is a completely open question. We also avoid other empirical questions, like whether players perhaps update virtually, but in addition use additional reasoning processes. Starting from the basic reinforcement learning model, we analyze the effect of only one reasoning step; a step that has been advanced in the literature as a self-evident improvement.

[14]Whether this phenomenon is related to the issue of the dynamic (in)stability of backward induction (Cressman, Schlag, 1996), or to problems related to fictitious play in such games (Hendon et al., 1996) is an open question.

References

Bandura, A., 1977. Social Learning Theory. Prentice-Hall, Englewood Cliffs, NJ.

Barto, A.G., Sutton, R.S., Anderson, C.W., 1983. Neuronlike Adaptive Elements That Can Solve Difficult Learning Control Problems. IEEE Transactions on Systems, Man, and Cybernetics 13(5), 834–846.

Börgers, T., Sarin, R., 1995. Learning through Reinforcement and Replicator Dynamics (mimeo).

Bush, R.R., Mosteller, F., 1955. Stochastic Models for Learning. Wiley, New York.

Camerer, C., Ho, T.H., 1996. Experience-weighted Attraction Learning in Games: A Unifying Approach (mimeo).

Cressman, R., Schlag, K., 1996. The Dynamic (In)Stability of Backwards Induction. Discussion Paper No. B-347, University of Bonn.

Cross, J.G., 1983. A Theory of Adaptive Economic Behavior. Cambridge University Press, Cambridge.

Ellison, G., 1993. Learning, Local Interaction, and Coordination. Econometrica 61, 1047–1071.

Erev, I., Roth, A.E., 1995. On the Need for Low Rationality, Cognitive Game Theory: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria (mimeo).

Forsythe, R.J., Horowitz, J.L., Savin, N.E., Sefton, M., 1994. Fairness in Simple Bargaining Experiments. Games and Economic Behavior 6, 347–369.

Gale, J., Binmore, K., Samuelson, L., 1995. Learning to Be Imperfect: The Ultimatum Game. Games and Economic Behavior 8, 56–90.

Güth, W., Schmittberger, R., Schwartz, B., 1982. An Experimental Analysis of Ultimatum Bargaining. Journal of Economic Behavior and Organization 3, 367–388.

Güth, W., Tietz, R., 1990. Ultimatum Bargaining Behavior: A Survey and Comparison of Experimental Results. Journal of Economic Psychology 11, 417–440.

Hendon, E., Jacobsen, H.J., Sloth, B., 1996. Fictitious Play in Extensive Form Games. Games and Economic Behavior 15, 177–202.

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

Holland, J.H., 1990. Concerning the Emergence of Tag-mediated Lookahead in Classifier Systems. Physica D 42, 188–201.

Holland, J.H., Holyoak, K.J., Nisbett, R.E., Thagard, P.R., 1986. lnduction: Processes of Inference, Learning, and Discovery. MIT Press, Cambridge, MA.

Hull, C.L., 1943. Principles of Behavior. Appleton-Century-Crofts, New York.

Kandori, M., Mailath, G.J., Rob, R., 1993. Learning, Mutation, and Long Run Equilibria in Games. Econometrica 61, 29–56.

Lin, L.J., 1992. Self-Improving Reactive Agents Based on Reinforcement Learning. Planning and Teaching. Machine Learning 8(3/4), 293–321.

Mitzkewitz, M., Nagel, R., 1993. Experimental Results on Ultimatum Games with Incomplete Information. International Journal of Game Theory 22, 171–198.

Roth, A.E., Erev, I., 1995. Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term. Games and Economic Behavior 8, 164–212.

Sutton, R.S., 1992. Introduction: The Challenge of Reinforcement Learning. Machine Learning 8(3/4), 225–227.

Stahl, D.O., 1996. Boundedly Rational Rule Learning in a Guessing Game. Games and Economic Behavior 16, 303–330.

Thaler, R.H., 1988. The Ultimatum Game. Journal of Economic Perspectives 2, 195–206.

Vriend, N.J., 1994. Artificial Intelligence and Economic Theory. In: E. Hillebrand, J. Stender (Eds.), Many-Agent Simulation and Artificial Life. IOS Amsterdam, pp. 31–47.

Young, H.P., 1993. The Evolution of Conventions. Econometrica 61, 57–84.