Learning by imitation

Erdem Başçı*

Department of Economics, Bilkent University, 06533 Bilkent, Ankara, Turkey

Abstract

This paper introduces a learning algorithm that allows for imitation in recursive dynamic games. The Kiyotaki–Wright model of money is a well-known example of such decision environments. In this context, learning by experience has been studied before. Here, we introduce imitation as an additional channel for learning. In numerical simulations, we observe that the presence of imitation either speeds up social convergence to the theoretical Markov–Nash equilibrium or leads every agent of the same type to the same mode of suboptimal behavior. We observe an increase in the probability of convergence to equilibrium, as the incentives for optimal play become more pronounced. © 1999 Elsevier Science B.V. All rights reserved.

1. Introduction

This paper introduces the idea that agents use imitation of social values in their learning process. We describe a learning algorithm which incorporates individual and social learning in recursive decision environments. In numerical simulations, we show that the presence of imitation matters a lot under some economic environments.

* Tel. 00 90 312 2901469; fax: 00 90 312 2665140; e-mail: Basci@bilkent.edu.tr.

In repeated static decision problems, imitation is studied by Ellison and Fudenberg (1993), Schlag (1998) and Offerman and Sonnemans (1998). In his theoretical work Schlag (1998) allows for the imitation of the *action* of one other agent, depending on his performance in the last period. In another theoretical paper, Ellison and Fudenberg (1993) study a similar rule of thumb for social learning. In both studies, historical data beyond one period is ignored by agents. In the experimental study of Offerman and Sonnemans (1998), subjects are observed to imitate *forecasts* of successful players whenever they have access to this information. In the learning algorithm that we suggest here, the agents occasionally imitate the *social values*, which are composed of individual values, attached to particular strategies. This approach, makes the social learning model usable in recursive dynamic settings as well.

In dynamic decision problems, learning by experience is studied by Lettau and Uhlig (1995) and Marimon et al. (1990). Lettau and Uhlig (1995) explore the theoretical performance of an individual learning algorithm for an agent that is faced with a dynamic optimization problem. In contrast, Marimon et al. (1990) run numerical simulations of a similar algorithm, to model learning by experience in a recursive dynamic game, namely, the Kiyotaki–Wright economy.

Here, we introduce learning by imitation in dynamic decision contexts. In numerical simulations, we apply our suggested learning algorithm to the Kiyotaki–Wright economy and contrast our results with the simulation results of Marimon et al. (1990) and the experimental results of Brown (1996) and Duffy and Ochs (1996). We observe that the presence of imitation speeds up convergence and homogenizes behavior. However, there is a possibility of convergence to a suboptimal mode of behavior as well. Such cases are observed when the payoffs from the suboptimal play and optimal play are very close to each other. Moreover, the likelihood of observing this possibility approaches zero, as the incentives for optimal play become more pronounced.

The next section describes the Kiyotaki–Wright model of money. Section 3 gives the details of our learning model in this context. Section 4 presents our main results. The final section concludes with a brief discussion of the results. The detailed algorithm can be found in the appendix.

2. The Kiyotaki–Wright model

In the economy described by Kiyotaki and Wright (1989), there are three indivisible goods: 1, 2, and 3. They are all storable but at a cost in terms of utility. Goods 1 and 3 have the lowest and highest storage cost, respectively. There are three types of infinitely lived agents: 1, 2 and 3. Type 1 enjoys consuming Good 1 but derives no utility from consuming the other goods. Likewise, Goods 2 and 3 are the favorites of Types 2 and 3, respectively. A continuum of each type is present in the economy. Each agent has one unit of storage capacity and is

endowed with one unit of a good to start with. Agents are specialized in production. As soon as consumption takes place, Goods 2, 3 and 1 are produced by Types 1, 2 and 3, respectively. They all get disutility from production.

There is no central clearing house, instead there is a trade round in the beginning of each period. In a trade round, agents are randomly matched in pairs and they either trade by exchanging their inventories or choose not to trade and wait for the next period's match. Trade takes place if and only if they both agree on it. Immediately after the trade round, the consumption decision takes place. If an agent decides to consume, he enjoys some utility, produces one unit of his production good and incurs some loss in the form of disutility from production. Right after this decision, the storage cost of the good left at hand is subtracted from the current utility. Then the agent enters the next period and the economy repeats itself.

Under the above production and preference pattern, this market organization does not allow the agents to produce and consume in every period, due to a mutual inconsistency in needs. Some of the agents have to accept and temporarily hold a good that they dislike, to facilite mutually beneficial trade. This good will serve as a medium of exchange. Which agent accepts which good as a medium exchange will be determined endogenously by the Markov–Nash equilibrium concept.

The objective function of a Type i agent is given by

$$\mathrm{E} \sum_{t=0}^{\infty} \beta^t (U_{it} - D_{it} - c_{i1t} - c_{i2t} - c_{i3t}),$$

where U_{it} is the utility from consumption, D_{it} is the disutility from production, c_{ijt} are the costs of storage of the jth good, and $\beta \in (0,1)$ is the discount factor. In the above equation, each one of U_{it}, D_{it} and c_{ijt} is a history contingent random variable that takes on the value of zero, unless the corresponding event of consumption, production or storage of a particular good has taken place. When the event of consumption of the own good, or of production or storage of Good j takes place, these random variables take on the values of $U_i > 0$, $D_i > 0$ and $c_{ij} > 0$, respectively. For each type of agent, Good 1 has the lowest storage cost and Good 3 has the highest, i.e. $c_{i1} < c_{i2} < c_{i3}$ for all $i = 1, 2, 3$.

The strategies of each player consists of trade offer and consumption policy functions. Since any agent can hold only one of the three goods at a point in time, after two agents are matched, there are nine possible pre-trade states for each one of them. The trade offer function maps these nine states to the set $\{0, 1\}$. Here, 1 stands for 'offer trade' and 0 for 'do not offer trade'. Similarly, after every trade round, there are three possible pre-consumption states and the consumption function maps these to $\{0, 1\}$, 0 meaning 'do not consume' and 1 meaning 'consume'. Clearly, the value of the objective function depends not only on the agent's strategies but also on those of the other agents and the initial distribution of goods over agents.

Kiyotaki and Wright (1989) have studied the steady-state Nash equilibrium strategies for all possible ranges of parameters. They have shown that, if the storage cost difference between Goods 3 and 2 is sufficiently large, compared to the discounted net utility obtained from the consumption good of a Type 1 agent, or more specifically for parameter values satisfying

$$(c_{13} - c_{12}) > 0.5\frac{\beta}{3}(U_1 - D_1), \tag{1}$$

the unique equilibrium is the 'fundamental' one. In a fundamental equilibrium, Type 1 agents do not accept Good 3 as a medium of exchange and tend to keep their production good, Good 2, in hand waiting for a suitable match in the next period. For any set of parameter values satisfying condition (1), the unique equilibrium inventory distribution is given in the left panel of Table 1. The term fundamental equilibrium is used in the sense that only the good with the lowest storage cost, i.e. Good 1, serves as a medium of exchange.

In a trade round, each agent obviously accepts his own consumption good. Furthermore, Good 1 is acceptable by Type 2 agents since it has the same marketability as Good 3 but its storage cost is lower. Type 3 agents, knowing that their production good, i.e. Good 1, is accepted by both Types 1 and 2, will not use Good 2 as a medium of exchange. For Type 1 agents, however, their production good, i.e. Good 2, has limited marketability. Therefore, they would accept Good 3 in exchange for Good 2 if and only if the marketability advantage of Good 3 over Good 2 exceeds the storage cost disadvantage. This idea is captured by inequality (2) below,

$$(c_{13} - c_{12}) < \beta(U_1 - D_1)(p_{31} - p_{21})/3, \tag{2}$$

where p_{ij} denotes the proportion of Type i agents entering the trade round with Good j. In this inequality, $p_{21}/3$ and $p_{31}/3$ are the probabilities of Type 1 agent being matched to an agent of Types 2 and 3, respectively, carrying Good 1 in his inventory.

Table 1
Type distribution over goods at the pre-trade stage

Agent	Fundamental equilibrium			Speculative equilibrium		
	Good 1	Good 2	Good 3	Good 1	Good 2	Good 3
Type 1	0	1	0	0	0.71	0.29
Type 2	0.5	0	0.5	0.59	0	0.41
Type 3	1	0	0	1	0	0

Note: The table reports the equilibrium densities of agents over stocks of goods when they enter the trade round under fundamental and speculative parameters.

Since these are the only types who would accept Goods 2 and 3, respectively $(p_{31} - p_{21})/3$ indicates the extent of the marketability difference between Goods 3 and 2 in the next period. Multiplication by the utility value and discounting converts marketability to current utility terms. The left-hand side is the difference between the costs of holding Goods 3 and 2. The inequality, then, is a condition for Good 3 to be acceptable by Type 1 agents.[1] Notice from Table 1 that $p_{31} = 1$ and $p_{21} = 0.5$, which means under condition (1) that inequality (2) is not satisfied and, hence, the Type 1 agents will not accept Good 3 as a medium of exchange in the fundamental equilibrium. As a consequence, in the fundamental equilibrium, only Good 1 will be used by Type 2 agents as a medium of exchange.

A more interesting situation arises as the unique equilibrium when parameter values satisfy,

$$(c_{13} - c_{12}) < (\sqrt{2} - 1)\frac{\beta}{3}(U_1 - D_1). \tag{3}$$

This equilibrium is called 'speculative', since in addition to Good 1 being accepted by Type 2 agents, Good 3, which has the highest storage cost, becomes acceptable by Type 1 agents. For any parameter value satisfying condition (3), the equilibrium inventory distribution is also given in Table 1.

In the speculative equilibrium, the strategies of Types 2 and 3 agents are the same as those in the fundamental equilibrium. From Table 1, it is clearly seen that $p_{21} = 0.59$ and $p_{31} = 1$, hence, inequality (2) is satisfied this time, thanks to condition (3). Therefore, Type 1 agents have the marketability advantage of carrying Good 3, rather than Good 2, in their inventory. This marketability advantage exceeds the storage cost disadvantage, and now Type 1 agents are willing to exchange Good 2 for Good 3, which has the highest storage cost. Therefore, in this case both Goods 1 and 3 are utilized as media of exchange in the economy.

3. The learning algorithm

In this section, we suppose that the agents operating in a Kiyotaki–Wright economy do not know the equilibrium strategies and start off by acting according to some randomly held beliefs regarding the values of the possible actions. While they interact, they will have opportunities of both learning by experience and by imitation.

---

[1] For the derivations and the proofs of the statements here, the reader may refer to the original Kiyotaki and Wright (1989) paper.

Table 2
Parameters used in simulations

Good	Produced by	Consumed by	Net utility	Storage cost
1	Type 3	Type 1	1.5	0.1
2	Type 1	Type 2	1.5	0.20 or 0.29
3	Type 2	Type 3	1.0	0.3

Number of agents per type	20
Discount factor	0.90
Maximum number of periods	2000

Note: The set of parameter values corresponding to ones used in Brown's (1996) experimental study (Set I) and a variant of it (Set II) are shown in the table. In Set I, $c_2 = 0.20$ and in Set II, $c_2 = 0.29$. Under both sets the unique equilibrium is the speculative one. However, Set II provides relatively stronger incentives for Type 1 agents to accept Good 3 as a medium of exchange.

To run numerical simulations, we need to have a finite number of agents of each type, which we take as 20. The two sets of parameter values that we use in our simulations, are reported in Table 2.[2] Both Sets I and II parameters satisfy condition (3), so that under both, the unique equilibrium is the speculative one. The first set is, basically, the same as the ones used by Brown in his experimental study, with the exception of our explicit use of a discount factor ($\beta = 0.9$) and a disutility term ($D_i = 0.1$ for all i), which is not reported in the table. Set II differs from Set I only in the storage cost of Good 2. This makes inequality (3) and, therefore, Eq. (2) more strict under Set II parameters, and increases the incentives for Type 1 agents to use the speculative strategy of accepting Good 3 as a medium of exchange.

Each of our Type 1, 2 and 3 agents start period zero by holding Goods 2, 3 and 1 in their inventories, respectively. Then they enter their first trading round.

In our algorithm, as in Marimon et al. (1990), learning and decision making takes place by means of classifier systems. A classifier system is a list of condition-action statements and an associated real number for each statement. Each condition-action statement is called a classifier and the real number associated with it is called its strength.[3] There are three main steps in a learning

---

[2] In Brown (1996), as in our Table 2, the net payoff assigned to Type 3 agents is 0.5 points below those assigned to Types 1 and 2. The reason for this is, most probably, to economize on the total cost of the experiment. This asymmetry in utilities is expected to cause no problems in learning, since the problem faced by Type 3 agents is the simplest among all the three.

[3] The classifier systems were introduced to the artificial intelligence literature by Holland (1975). Booker et al. (1989) is also a good reference. For applications in economics, Arthur (1991), Arthur et al. (1997), Beltrametti et al. (1997), Marimon et al. (1990), Sargent (1993) and Vriend (1996) are some good references.

algorithm utilizing a classifier system: activation, selection and update. These are described as Steps 1, 3 and 4 below. The intermediate Step 2, which is the contribution of the present paper, constitutes the imitation step.

1. Recognize your current state and determine the classifiers whose 'condition' parts are satisfied in the current state (activation).
2. (Execute this step with probability p_{imit}) Pick one of the activated classifiers randomly and replace its strength with an average of other agents' strengths attached to this classifier (imitation).
3. Pick one classifier among the activated ones according to their strengths, follow its advice and bear the consequences (selection).
4. According to the consequences, update the strength of the classifier responsible from these (update).
5. Go to step 1.

In our economy, regarding Step 1, we assume that each agent has a private classifier system that is complete, consisting of trade and consumption classifiers. By completeness we mean that in each possible state and for each possible action, that can be taken at that state, there is one distinct classifier. This assumption means that, the agents are able to recognize fully both their consumption state, i.e. the type of good that they have in hand before consumption, and their trade state, i.e. the type of own good and the trading partner's good, in that period. Since there are only three goods, there are three consumption states and nine trade states. In a consumption round, for each good there are two possible classifiers, one recommending consumption and the other advising to keep the good. This makes altogether six consumption classifiers per agent. Likewise, there are two trade classifiers for each of the nine trade states offering to and not to trade, which makes a total of 18 trade classifiers.

In principle, we let our agents follow the recommendation of the classifier with the higher strength, in Step 3. However, we also allow for totally random action with a 5% probability. This both captures the 'trembling hand' notion of Selten (1975) and, also, is essential as a device for making the agents try the seemingly bad strategies. As a result, since always two classifiers are activated in this model, the seemingly bad classifier is chosen with 2.5% probability in every decision situation. The presence of such *mistakes* is known to be helpful in the learning process, as they serve as a non-deliberate device for experimentation.[4]

In Step 3, we draw the initial classifier strengths from an iid. normal density function with unit mean and variance. The initial strengths may be interpreted as prior beliefs regarding the value of each condition-action statement. The

---

[4] For example, see Marimon et al. (1990) for mistakes arising from the mutation and crossover operations of the genetic algorithm and Başçı and Orhan (1998) for mistakes arising from trembling hands.

update of consumption classifier strengths takes place according to the formula,

$$S_{t+1}^c = S_t^c + \frac{1}{\tau_t^c + 1}(U_t - D_t - c_t + \beta S_{t+1}^r - S_t^c), \tag{4}$$

where S_t^c is the consumption classifier strength of an agent, who has chosen this particular rule in period t. τ^c is an experience counter showing how many times this particular rule has been chosen in the past. U_t, D_t are the utilities and disutilities from consumption and production in this period. If consumption did not take place, these are taken as zero. c_t is the storage cost of inventory held from period t to $t + 1$. Finally, S_{t+1}^r is the strength of the trade offer classifier, which is selected in the trade round to take place the next period. This update is done only after the trade decision in time $t + 1$ is made. Therefore, the effects of trembling hands on S_{t+1}^r are incorporated.

The presence of the discount factor, β, multiplying the strength of the next period's chosen trade classifier, S_{t+1}^r, is to establish the link between the asymptotic values of the classifier strengths and the optimal values coming from the Bellman's equation faced by the agent in equilibrium. Once the expected value of the term in parenthesis is zero for all classifiers, then we can say that Bellman's equation is satisfied for this consumer, where the optimal value function at a given state is equal to the maximal classifier strength at that state. Once the expected value in the parenthesis becomes equal to zero, the expected value of the change in the classifier strengths $(S_{t+1}^c - S_t^c)$ is also zero, an indicator that learning has taken place.[5] Immediately after the strength update, the experience counter is also updated according to $\tau_{t+1}^c = \tau_t^c + 1$. The initial values of the experience counters are set to 1 for all of the classifiers.

The update of the trade classifier strength is made similarly according to,

$$S_{t+1}^r = S_t^r + \frac{1}{\tau_t^r + 1}(S_t^c - S_t^r), \tag{5}$$

where τ^r is the experience counter for this particular trade classifier. There is no utility and discounting term in the equation because the consumption decision is made right after the trade round. Hence, the strength of the chosen consumption classifier summarizes all future payoffs afterwards. If our trade classifier has sent us to a *strong* consumption state, it is rewarded by an increase in its strength. Otherwise, its strength is reduced. Again, the update of the experience counter comes immediately after the strength update, $\tau_{t+1}^r = \tau_t^r + 1$.

---

[5] For dynamic problems with a recursive structure this modification of the strength update formula has been suggested by Lettau and Uhlig (1995), in order to establish a correspondence with the appropriate dynamic programming problem. Their proofs are based on stochastic approximation theory, an introductory account of which can be found in Sargent (1993).

In Step 2, we introduce imitation in the form of occasionally adopting the *social strength* attached to a randomly selected one of the activated classifiers before a decision is to be made at an observed state. The social strength of a rule is calculated as the experience weighted average strength, over all agents of the same type as the agent under consideration. Imitation is assumed to take place, at a rather low probability, every period. The probability of imitation, p_{imit}, is constant across agents and over time. When two agents are matched, two separate imitation coins are tossed. The agent who observes the 'imitate' side of the coin, tosses another coin. Accordingly, with equal probabilities either one of the two active classifier's strength is replaced with its social counterpart, given by

$$S_{it}^l = \frac{\sum_{k=1}^{20} \tau_{kt}^l S_{kt}^l}{\sum_{k=1}^{20} \tau_{kt}^l}, \tag{6}$$

where i is the index of the agent to imitate, l is the index of the classifier that is selected at the imitation step. This adoption is assumed to be done subconsciously. The implicit mechanism could be as follows. Each agent keeps talking about his perceived value regarding this particular classifier, based on his experience with it. The more experienced ones talk more confidently, so that their influence on the social average is greater. Occasionally and subconsciously, an individual agent is affected by this social value and adopts it and treats this number as if it was due to his own experience. Therefore, no update of experience counters is made at the imitation stage.

4. Results

We have written two versions of a GAUSS program implementing the above algorithm. The first version goes over the algorithm only once and, at the end, presents a time graph of the proportion of Type 1 agents, who would choose to trade their production good (Good 2) for Good 3. The second version runs the algorithm 100 times and reports a different set of statistics. This version reports a distribution over runs of the proportion of Type 1 agents, who would be willing to accept Good 3 as a medium of exchange, in trade state (2, 3), at times 50, 1000 and 2000. The second version can be used to study the probability of social convergence to a particular mode of behavior, while the first version can provide details from a single run.

In a typical run of the first version, with zero imitation probability, the proportion of Type 1 agents, who are willing to accept Good 3 as a medium of exchange, is plotted as a function of time in Fig. 1. The figure indicates that even for Set II parameter values, which provide strong incentives, and even over such a relatively long time horizon, individual experimentation via mistakes was not
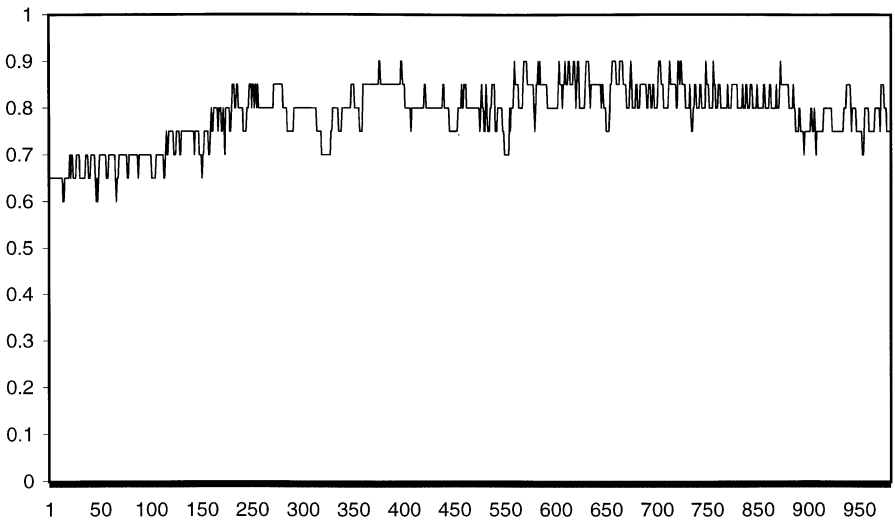
Fig. 1. Learning under strong incentives and without imitation. The proportion of Type 1 agents who have learned to accept Good 3 as a medium of exchange is plotted as a function of time. The imitation rate is zero. Set II parameters, which provide strong incentives for using the speculative strategy of accepting Good 3 are used. In period 1000, only 75% of the agents have learned to play the speculative strategy.

sufficient to teach every agent the optimal behavior. At time period 1000, only around 75% of Type 1 agents were using Good 3 as a medium of exchange. Another observation is that this proportion kept on fluctuating, even in the last 100 periods. This means that some agents were getting favorable results from their fresh experimentation of alternative strategies and switching to them and some agents were getting unfavorable outcomes from their existing strategies and, hence, switching back to their previous attitudes even after quite some experience, on the order of 100 trials per classifier.

For Set II parameter values, again, but with 9% imitation probability this time, Fig. 2 is drawn for a typical run in which all the agents converged to the speculative equilibrium. At around period 600 of this run, all Type 1 agents have started accepting Good 3.[6]

We also include, for comparison, summary results from a run, in which social convergence to speculative equilibrium did not take place, but, still, presence of

---

[6] The details of how learning evolves in all of the key classifiers for all types over time are reported in the discussion paper version of this paper, Başçı (1998), which is available from the author upon request. Başçı (1998) also verifies that the pre-trade inventory distribution over types is in conformity with the speculative equilibrium, shown in the right panel of Table 1 here.
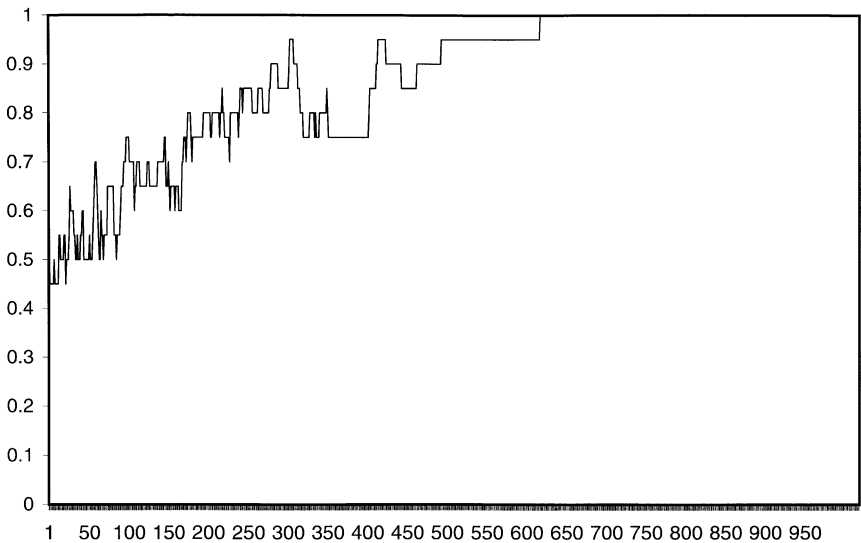
Fig. 2. Learning under strong incentives and with imitation. The proportion of Type 1 agents who have learned to accept Good 3 as a medium of exchange is plotted as a function of time. The imitation rate is 9%. Set II parameters, which provide strong incentives for using the speculative strategy of accepting Good 3 are used. In period 600, all of the agents have learned to play the speculative strategy.

imitation acted in a way to homogenize behavior within the same type. For this run, Set I parameter values were used. Fig. 3 shows that at around the 900th time period, none of Type 1 agents were willing to accept Good 3 as a medium of exchange, even though it would be optimal for them to do so.

Under a positive imitation probability, we have observed two substantially different convergence patterns in two different runs. To analyze the respective probabilities of observing these two outcomes in a particular run, we have used our second program, which generates 100 independent simulation paths for any given imitation probability.

Table 3 reports the results for Set I parameter values, under which the incentives to play 'speculative' are rather weak. The first striking observation from the table is that, without imitation, in none of the 100 runs, full social convergence to a speculative equilibrium was attained, even after 2000 time periods. As soon as the imitation rate is increased to a small positive number, the number of runs, with full social convergence as of period 2000, jumps to 21. A second observation is the tendency to increase in the number of fully convergent runs by period 1000, with the increase in the rate of imitation. The third observation is that, in period 2000 there is a rough tendency toward stabilization of the number of fully convergent runs, at around 40% after the imitation rate of 6% is exceeded.
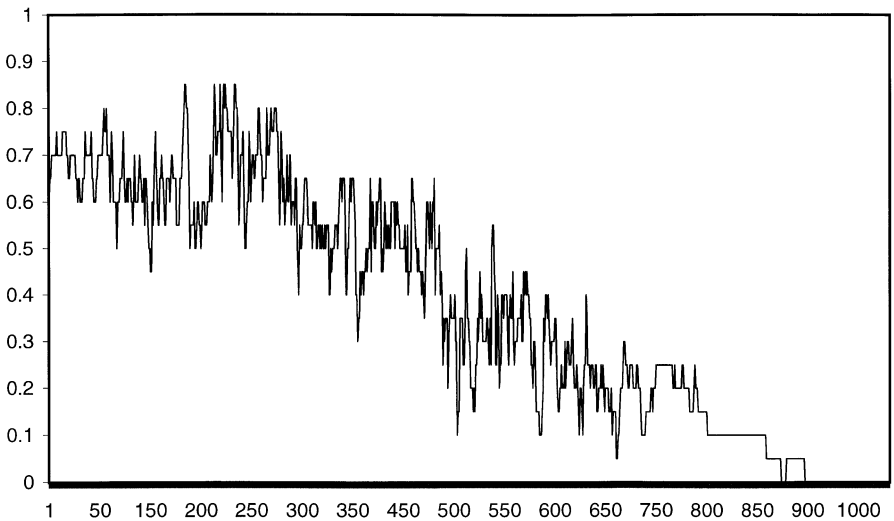
Fig. 3. Learning under weak incentives and with imitation. The proportion of Type 1 agents who have learned to accept Good 3 as a medium of exchange is plotted as a function of time. The imitation rate is 9%. Set I parameters, which provide relatively weaker incentives for using the speculative strategy of accepting Good 3 are used. In period 900, all of the agents are observed to play the fundamental strategy, although it is suboptimal.

Table 3
The probability (%) of convergence to speculative equilibrium (weak incentives)

Behavioral assumptions	Period 50	Period 1000	Period 2000
Rational expectations	100	100	100
Learning with imitation prob = 0.00	0	0	0
Learning with imitation prob = 0.03	0	10	21
Learning with imitation prob = 0.06	0	23	35
Learning with imitation prob = 0.09	0	26	41
Learning with imitation prob = 0.12	0	34	39
Learning with imitation prob = 0.15	0	36	42

Note: The table reports the number of cases out of 100 where *all* of Type 1 agents learned to accept Good 3 as a medium of exchange. Set I parameter values, which provide relatively weak incentives for optimization, are used. Since all the other types learn equilibrium play by time 1000, the reported numbers can also be interpreted as the probability of convergence to the unique Markov–Nash equilibrium in a particular run.

Taking into account the similar behavior of the number of runs that we have observed to converge to the suboptimal fundamental behavior, we can express the following conjecture: For any positive imitation rate, the probability of full convergence to a speculative equilibrium, in a particular run, is around 40% and the probability of full convergence to the (suboptimal) fundamental behavior is close to 60%. The justification for this belief comes from the observation that, in period 2000 the sum of these two extreme cases starts exceeding 90%, for high imitation probabilities, and has a tendency to increase further, approaching 100% as the number of periods is further increased. Of course, these numbers are specific to Set I parameter values and the 5% trembling hand probability. A second conjecture, which may be posed, is that higher imitation rates speed up reaching either one of the two limit points of social convergence, mentioned above.

To see the effects of incentives for optimization on the probability of convergence to the Markov–Nash equilibrium, we run the same program with Set II parameter values. Table 4 presents the results for this case, which yield stronger incentives to play 'speculative'. Although the same qualitative observations can be made, the probability of convergence to the speculative equilibrium, in this case, is observed to be around 90%, which is substantially above the 40% observed under Set I parameter values. Investigation of the series, which are not reported in Table 4, reveals that the probability of convergence to the 'fundamental' behavior is around 10% in this case (For details and further robustness and sensitivity checks, see Başçı, 1998).

Table 4
The probability (%) of convergence to speculative equilibrium (strong incentives)

Behavioral assumptions	Period 50	Period 1000	Period 2000
Rational expectations	100	100	100
Learning with imitation prob = 0.00	0	0	0
Learning with imitation prob = 0.03	0	26	68
Learning with imitation prob = 0.06	0	58	84
Learning with imitation prob = 0.09	0	70	89
Learning with imitation prob = 0.12	0	74	85
Learning with imitation prob = 0.15	0	77	85

Note: To be compared with Table 3. Here, Set II parameter values, which provide stronger incentives for Type 1 agents to use speculative strategies, are used.

5. Concluding remarks

In this paper we introduced a learning algorithm which allows the agents to imitate other agents' values in dynamic game settings with a recursive structure. Imitation in this context is the pooling of the diverse group experience, summarized by individual values attached to certain strategies, into social values for these strategies, and letting the individual agents occasionally adopt these social values in a random fashion. We applied this algorithm by running numerical simulations in the context of the Kiyotaki and Wright (1989) model of money as a medium of exchange.

In contrast to the previous work of Marimon et al. (1990), which studied learning by experience in the Kiyotaki–Wright framework, we allowed for imitation. As a consequence, under speculative equilibrium parameter values, we observed a positive probability of full social convergence to the unique Markov–Nash equilibrium in a given run. This probability is observed to be zero, under the absence of imitation. Our second observation has been that the probability of convergence to the speculative equilibrium approaches one as the incentives to accept Good 3 as a medium of exchange increase.

Once we set the imitation rate to zero, our algorithm qualitatively mimics the experimental results of Brown (1996) and Duffy and Ochs (1996). In this case only a fraction of agents of the same type are observed to learn the optimal play in reasonable time horizons. Neither Brown (1996) nor Duffy and Ochs (1996) have allowed their subjects to communicate with each other. Since sharing attitudes, regarding the values of alternative strategies, was not possible, their institutional setup allowed for individual learning by experience only. As a result, only a portion of their subjects learned to play optimally. We would expect that, in experimental studies, letting the subjects communicate among themselves will have a non-negligible effect on the outcomes.

Based on our experience with the learning algorithm in Kiyotaki–Wright environment, we can make the following conjectures about the performance of the model under discrete choice recursive decision environments in general:

1. Learning dynamic optimization takes time, and if incentives to maximize are not strong enough it may not even take place.
2. Imitation in the form of adopting social values leads to uniformity behavior across agents of the same type.
3. Imitation speeds up convergence to the uniform behavior mentioned in (2).
4. With imitation, although the convergence, mentioned in (2), is normally to the optimal behavior, if the incentives to maximize are not very strong, social convergence to a suboptimal mode of behavior can be observed as well.

These findings could be important, for instance, in macroeconomic models, where imitation is a potentially significant factor in affecting the adjustment dynamics that emerge after a major structural change takes place in the economy.

Acknowledgements

Appendix A. The algorithm

The algorithm of the GAUSS program, which generates one simulation path in our artificial Kiyotaki–Wright economy, is given below. A modified version generates 100 independent runs and reports statistics across runs.

```
Set parameters related to utility, disutility, cost, discount factor,
  mistake and imitation probabilities.
Distribute the agents their production goods as their initial
  allocation.
Set all initial classifier strengths iid. from N(1,1).
Set initial experience counters to unity.
For time = 1 to 1000
  Randomly match agents in pairs – by randomly sorting the
    60 agents and matching the adjacent ones.
  For each pair of agents, N = 1 to 30
    Determine the pre-trade state for both.
    With probability pimit, let first agent imitate by adopting
      the social strength of either DOTRADE classifier or
      NOTRADE classifier in that state with probability 50%.
    With probability pimit, let second agent imitate by adopting
      the social strength of either DOTRADE classifier
      or NOTRADE classifier in that state with probability 50%.
    Pick the stronger classifier's trade advice for both of the
      two agents.
    If the first agent's hand trembles (with probability 5%),
      randomly select between two active trade offer classifiers.
    If the second agent's hand trembles (with probability 5%),
      randomly select between two active trade offer classifiers.
    If t > 1, update the strength of the most recently used
      consumption classifier by using the most recently chosen
      trade classifier's strength for both agents.
```

If both offered trade swap their inventories.

With probability pimit, let first agent imitate by adopting
  the social strength of either DOCONSUME classifier
  or NOCONSUME classifier in that state with probability 50%.

With probability pimit, let second agent imitate by adopting
  the social strength of either DOCONSUME classifier
  or NOCONSUME classifier in that state with probability 50%.

Pick the stronger classifier's consumption advice for both
  of the two agents.

If the first agent's hand trembles (with probability 5%),
  randomly select between two active consumption classifiers.

If the second agent's hand trembles (with probability 5%),
  randomly select between two active consumption classifiers.

If the hand trembles with probability 2.5%, choose the
  contrary advice for both agents.

If decided to consume, let them produce the good they can.

Update the strength of the most recently used trade classifier
  by using the most recently chosen consumption classifier's
  strength for both agents.

Record the proportion of agent who would choose DO consume
  and DO trade at this time for all possible states.

Record the end of period stock distribution over agents.

Prepare statistics and report them.

References

Arthur, W.B., 1991. Designing economic agents that act like human agents: a behavioral approach to bounded rationality. American Economic Review 81, 352–359.

Arthur, W.B., Holland, J.H., LeBaron, B., Palmer, R., Tayler, P., 1997. Asset pricing under endogenous expectations in an artificial stock market. In: Arthur, W.B., Durlauf, S.N., Lane, D.A. (Eds.), The Economy as a Complex System II. Addison-Wesley, Reading, MA.

Başçı, E., 1998. Learning by imitation in the Kiyotaki–Wright model of money. Discussion Paper no. 98-18, Bilkent University.

Başçı, E., Orhan, M., 1998. Reinforcement learning and dynamic optimization. Unpublished manuscript, Bilkent University.

Beltrametti, L., Fiorentini, R., Marengo, L., Tamborini, R., 1997. A learning-to-forecast experiment on the foreign exchange market with a classifier system. Journal of Economic Dynamics and Control 21, 1543–1575.

Booker, L.B., Goldberg, D.E., Holland, J.H., 1989. Classifier systems and genetic algorithms. Artificial Intelligence 40, 253–282.

Brown, P.M., 1996. Experimental evidence on money as a medium of exchange. Journal of Economic Dynamics and Control 20, 583–600.

Duffy, J., Ochs, J., 1996. Emergence of money as a medium of exchange. American Economic Review (forthcoming).

Ellison, G., Fudenberg, D., 1993. Rules of thumb for social learning. Journal of Political Economy 101, 612–643.

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

Kiyotaki, N., Wright, R., 1989. On money as a medium of exchange. Journal of Political Economy 97, 927–954.

Lettau, M., Uhlig, H., 1995. Rules of thumb and dynamic programming. American Economic Review (forthcoming).

Marimon, R., McGrattan, E., Sargent, T.J., 1990. Money as a medium of exchange in an economy with artificially intelligent agents. Journal of Economic Dynamics and Control 14, 329–373.

Offerman, T., Sonnemans, J., 1998. Learning by experience and learning by imitation of successful others. Journal of Economic Behavior and Organization 34, 559–575.

Sargent, T.J., 1993. Bounded Rationality in Macroeconomics. Oxford University Press, Oxford.

Selten, R., 1975. Re-examination of the perfectness concept for equilibrium points in extensive form games. International Journal of Game Theory 4, 25–55.

Schlag, K.H., 1998. Why imitate and if so, how? Journal of Economic Theory 78, 130–156.

Vriend, N.J., 1996. Rational behavior and economic theory. Journal of Economic Behavior and Organization 29, 263–285.