NICOLAAS J. VRIEND

'FROM EXCHANGE IT COMES TO TEARS'. A DUTCH 'FOLK THEOREM' RECONSIDERED

ABSTRACT. A Dutch *'folk theorem'* holds that *'from exchange it comes to tears'*. This seems to contradict the basic idea found in economics that exchange and trade can make both sides better off. We show that the *'folk theorem'* has a better theoretical foundation than sometimes thought, as it is vindicated by the equilibrium of an exchange game with two-sided asymmetric information. We, then, explain the practical value of such *'folk wisdom'* in the real world by showing why players might be unlikely to learn such an equilibrium strategy.

KEY WORDS. Bayes–Nash equilibrium, exchange, *'folk theorem'*, information asymmetry, learning, nonconvergence

JEL CLASSIFICATIONS. C72, D82.

1. INTRODUCTION

Children in Holland are instilled with the warning that *'from exchange it comes to tears'*. But during my first lectures in economics at the University of Amsterdam, the professors assured us, with a superior and ironic smile, that although this *'folk theorem'* might be amusing, it is, of course, incorrect. As economists show time after time, exchange and trade can make both sides *better off*. An example is the dismissal of the doctrine of unequal exchange in the context of the theory of comparative advantage (see, e.g., Krugman and Obstfeld [1994]). Now, who was right? The *'folk theorem'* or the economists?

In Section 2, we will show that the *'folk wisdom'* has a better theoretical foundation than sometimes thought. In particular, it is vindicated by the Bayes–Nash equilibrium of an exchange game in which each player owns an object that he can offer (or not) in exchange for the object of the other player, with the value of each object being private information of the current owner. This leaves us with the following question. Why do Dutch children seem to

Theory and Decision **55:** 315–338, 2003.

need such persistent paternalistic advice in these matters? Even if children perhaps should not be expected to reach equilibrium through some introspective process, they might arrive at it through a process of learning, as they play this game very frequently. As we will see in Section 3, as learning takes place through a process of social interaction, it is not obvious at all that it will lead to equilibrium.

2. GAME-THEORETIC ANALYSIS

Let us consider the following two-player game[1]. Each player draws a lottery ticket from a hat (each player from a different hat), which contains tickets with numbers between 1 and 100. The number on each ticket indicates the monetary prize the player will get from the organizer in exchange for his ticket. However, before handing in their ticket, the two players may decide to exchange the tickets with each other. Importantly, the players know the number on their own ticket, but *not* the number on the other player's ticket. If and only if both players agree, exchange will take place. Otherwise each player simply gets the prize found on his own ticket. Should a player who finds himself with a ticket with number 40 agree to an exchange for a ticket with an unknown number in the range from 1 to 100? And if his ticket has number 10? And what about number 2?

If the set of possible actions is $A = \{keep\ own\ ticket,\ propose\ to\ exchange\ ticket\}$, and H_i is the collection of player i's information sets, with a specific information set h for player i being identified by the value of player i's own ticket, where $h \in \{1, 2, \ldots, 100\}$, then the strategy set for player i is $S_i = \prod_{h=1}^{100} A(h)$. Any strategy $s_i \in S_i$ with $A(100) = keep$ weakly dominates all corresponding strategies $s_i' \in S_i$ for which $A(100) = exchange$ and $s_i = s_i'$ for each $h = 1, 2, \ldots, 99$. Once eliminated those 2^{99} strategies for both players, the same argument applies to $h = 99$ for both players, and so forth till we reach $h = 1$. We will have, then, iteratively eliminated all 2^{100} strategies except two; saying to *keep* any ticket with a value greater than 1, and to *keep* or *exchange* a ticket with number 1. Notice that this iterative process of weakly dominated strategies happens to be path independent, and it also does not depend on the specific form of the distributions of tickets in the hats.

For convenience let us now assume the distribution of tickets in each hat is uniform, e.g., each hat contains exactly 100 tickets numbered from 1 to 100. The following simple reasoning process applies. If our player 1's potential trading partner, player 2, is prepared to exchange any ticket, the expected value of player 1's ticket will be the average of all possible values from 1 to 100: $(1+100)/2=50.5$. If our player 1 is risk-neutral, and only wants to maximize the expected payoff in dollars, the highest number on his own ticket which he is prepared to trade is 50. Now, if player 2 is rational, and knows player 1 is rational, then his expected value of the ticket player 1 offers him in exchange is only 25.5, i.e., $(1+50)/2$. Hence, the highest number he will offer to exchange will be 25. Since player 1 is still rational, knows that his opponent is rational, and knows that his opponent knows that he is rational, he is only prepared to exchange tickets with numbers from 13 down to 1. Just like the iterative elimination of dominated strategies above, this process of unraveling is based on the assumption of successive degrees of rationality, and is a reasoning process that will stop only when we are landed at 1. That is, both players may be prepared to exchange a ticket with number 1, but not any ticket with a higher number.

Instead of this introspective unraveling process by rational players, we could also search directly for the equilibrium strategy in this game with imperfect information. Define m_i as the highest number that player i (with $i=1,2$) is prepared to exchange. To value the various available strategies, player 1 must imagine what player 2 will do. The probability that player 2 has a ticket in hand with a number smaller than m_2 equals $m_2/100$, which is by definition the probability that player 2 is prepared to exchange tickets. Clearly, this means that the probability that player 2 is not ready to exchange will be $(100-m_2)/100$. Now, suppose player 1 has a ticket in hand with the number v_1. In case he decides to renounce the possibility of a trade, his payoff will be v_1. But if he opts for the trade, then his expected payoff will be the probability that player 2 agrees with the trade times the expected value of player 2's ticket, plus the probability that player 2 rejects times

v_1: $(m_2/100)\cdot E[v_2|m_2]+(100-m_2)/100\cdot v_1$. Obviously, player 1 will offer his ticket for trade if and only if $v_1 \leqslant [v_2|m_2]$. The right-hand side is exactly the m_1 as defined above. Hence, $m_1 = E[v_2|m_2]$, where the latter expectation is $(1+m_2)/2$. For player 2 we can derive exactly the same, and find that $m_2=(1+m_1)/2$. The strategies and beliefs of the players are mutually compatible if both equations are satisfied. Solving the two equations for m_1 and m_2 gives $m_1=m_2=1$. In other words, in a Bayes–Nash equilibrium, where the players' beliefs and strategies are mutually consistent, and where each player chooses a best-reply given the strategy of the other player, both players decide to keep any ticket with a number greater than 1.

Hence, even with a number 2 ticket in hand, a rational player decides to refrain from exchange. But that is exactly what the '*folk theorem*' says[2]. The analysis shows that the '*folk theorem*' should be taken as extreme as it appears. It would be incorrect to interpret the saying '*from exchange it comes to tears*' as a simple call for prudence because sometimes a player will discover to end up at the right side of a trade, and sometimes at the wrong side. If other players adhere to the equilibrium strategy, whenever trade takes place, a player will always discover he is at the same side of the deal; the wrong one. Hence, we can see the '*folk theorem*' as the embodiment of a social norm with a self-enforcing property, in the sense that its wisdom is justified by the equilibrium strategy of our exchange game, implying that deviations do not pay[3].

This exchange game is obviously related to a number of other games with asymmetric information analyzed in the literature. These games, however, differ in some important aspects from our exchange game. Moreover, it is our exchange game that corresponds precisely to the real world situation in which the '*folk theorem*' is employed: two children who want to exchange their toys, with each owner knowing the value of his own toy (the hidden defects, broken doors, stuck wheels, etc.), while lacking such information concerning the toy to be received.

Akerlof [1970] analyzed the '*market for lemons*' concerning the market for second-hand cars. In that market there is a one-sided information asymmetry. The seller knows the value of a car, whereas the potential buyer does not. This leads to an unraveling of the market as sketched above, in which each round the better cars disappear from the market, and only the lemons remain. This process of adverse selection could continue until no second-hand cars are offered for trade anymore[4]. There is a couple of differences between the market for lemons and our exchange game. A first difference is that car-dealers rarely cry after a trade. They do not suffer from the information asymmetry as they know exactly what their customers' money is worth. Hence, the '*folk wisdom*' '*from exchange it comes to tears*' does not apply to car-dealers. In our game, the information asymmetry is two-sided. Notice that the fact that the information asymmetry is bilateral does not mean that the effect of the unilateral asymmetry disappears. The disadvantage for a player of not knowing the value of the ticket to be received in exchange is not compensated by the relative advantage that the other player is in the same situation; quite the contrary. The '*folk theorem*' points to the fact that, as both sides are buyers, each side is on the verge of tears. Another difference is that car-dealers, in order to halt the unraveling process, take resort to formal warranties, and informal guarantees in the form of their reputation in the long run. Obviously, remedies such as warranties or the seller's reputation could be important in the exchange game we consider as well. Hence, our analysis applies to those situations where (for one reason or another) such remedies are not or cannot be effectively implemented. In particular, it seems not realistic to expect such remedies with children; either because it might be legally infeasible or too costly, or because children might be too myopic.

Kessler [2001] considers a market for lemons with two-sided asymmetric information, in the sense that some sellers are uninformed about the quality of the cars they sell. Notice that this is different from our two-sided information asymmetry because in Kessler both sides are uninformed about the same object.

The exchange game we consider is also related to a theme pursued by Myerson and Satterthwaite [1983]. They consider an exchange game with two-sided information asymmetry, and show the impossibility of a trading mechanism that is incentive compatible, individually rational, and *ex post* efficient. There are, however, some differences with our exchange game. In Myerson and Satterthwaite there are two players, of which only one owns an object, which the other wants to buy for money. The value that each player attaches to the object is private information. In our game, both players own an object, with its value being private information to the owner. That is, each player knows exactly how much his own object is worth to the other player, while not knowing the value of the object to be received. In Myerson and Satterthwaite, however, each player knows what he would get in case of exchange, while not knowing how much his own object is valued by the other player. Hence, whereas Myerson and Satterthwaite focus on efficiency (the player that values the object most owning it in the end), in our game efficiency is not an issue, as the value of a given ticket is identical for both players.

3. SOCIAL INTERACTION AND LEARNING

Given the game-theoretic analysis, some readers might wonder why it is that children seem to need advice from adults in these matters, and in such a persistent way. A first reason might be that the fact that an exchange game like this appears as an assignment in a graduate textbook suggests it is not realistic to assume that children themselves will be able to figure out the equilibrium strategy through *introspection*. However, one might presume that it would not be too difficult for children to *learn* their way to equilibrium if they play a large number of these exchange games.

As we will show in this section, although children may be learning in the real world, for such a process to reach the equilibrium is not as straightforward as the game-theoretic analysis might suggest. The main cause of this is that it is not a simple single-agent decision problem that the players need to learn to solve. Instead learning takes place through social interaction. A player's optimal action depends on the actions of the other players. Hence, we have a coevolutionary process because while a player is learning about the other players, these other players are learning about him. We will analyze how this feature may prevent a population of players from converging to the equilibrium strategy by examining a number of different learning schemes.

In Section 2, we explained that the exchange game is closely related to Akerlof's market for lemons. Since the game-theoretic analysis is in both cases driven by an information asymmetry, one might conjecture that there are important similarities in the learning processes in these two situations. Therefore, our analysis of the learning task faced by the players in the exchange game might shine some light on the issue of learning in market for lemons as well[5].

The learning schemes we consider are: best-response behavior, fictitious play, learning direction theory, hill-climbing, and reinforcement learning. The pseudo-code for each scheme can be found in the appendix. In each learning scheme, the player's choice in the first period is chosen from the range of possible threshold levels, from 1 to 100, with equal probability for each value. For all later periods, each scheme specifies a strategy on the basis of earlier events. To simplify the learning task, given the strategy set for player i S_i, we will assume that he will only use strategies with a so-called threshold property. That is, player i's strategy is characterized by a threshold number m_i, the highest number that player i (with $m \in \{1, 2, \ldots, 100\}$ and $i = 1, 2$) is prepared to exchange. Given player i's strategy, his exchange decision simply follows from checking the value of the ticket distributed to him against it.

If a player adheres to *best-response behavior*, he chooses a best-reply against the observed action of his opponent in the most recent period. In case there are multiple best-replies, each of them is chosen with equal probability. When a player himself did not offer to exchange in a given period while his opponent did want to exchange, he does not observe the value offered by his opponent, and therefore sticks to his current threshold.

With *fictitious play* a player also chooses a best-reply against some belief, but now this belief is based on the average value of all tickets received in exchange in the past. As long as he does not have any information in this respect, a player will continue to choose a threshold at random.

A player who behaves according to *learning direction theory* (see, e.g., Selten and Stoecker [1986]) looks at the outcome of the most recent period, and reasons in which direction a better threshold could have been found. He, then, simply adjusts his current threshold into that direction. That is, if he received a ticket in exchange with a value below his original ticket, he will decrease his threshold with one (as it will reduce his chance of encountering such disadvantageous exchanges), whereas if he received a ticket in exchange with a value even above his current threshold, he will increase his threshold with one (to increase the likelihood of finding such advantageous exchanges). If the value of a ticket received in exchange was in between his own original ticket and his current threshold, he keeps his threshold unchanged (as it is unclear which direction would be wise to choose). This also happens if no new information became available in the latest period because no exchange took place.

A *hill-climbing* player does not reason about thresholds and best-responses. He starts out experimenting in the first two periods, choosing a threshold at random. After that, he always looks back at the two most recent periods. If, during those two periods, he had increased his threshold from one period to the other, and this corresponded with an increase of the value of the ticket he had in his hand at the end of each period, then he will increase his threshold once more, moving one unit further up. If, however, an increased threshold had coincided with a decreased ticket value at the end of the day, then he will decrease his current threshold with one. The opposite applies to the cases in which he had decreased his threshold over the last two periods. If either his threshold or the value of the ticket in his hand at the end of the day had been unchanged during these two periods, then there is no hill to be climbed. In such a case he will either increase or decrease his current threshold with one with equal probability.

With *reinforcement learning* a player experiments in choosing his thresholds, being more likely to choose those thresholds that had been more reinforced (through higher payoffs) in the past. We use some weighted average of the reinforcement, placing more weight on more recent experiences. Given the reinforcements for all thresholds, a player uses the logit rule to choose his current threshold. As can be seen in table A5 in the appendix, the logit rule includes a sensitivity parameter β, with low values of β implying more randomness in a player's choice, and high values of β implying that his choice is more sensitive to differences in reinforcement.

Following Goeree and Holt [2001], we implement a zoom-in feature for this sensitivity parameter, decreasing noise as the iterations go on. Our interpretation of this is an *adaptive* one. As players start playing this game, they are very young and inexperienced. Hence, they choose with lots of noise. But as they gain experience, they will pick their threshold with more and more precision.

We will now analyze to what extent this range of different learning schemes would converge to equilibrium[6]. If they do converge, it would suggest the '*folk theorem*' to be unnecessary, leaving us with the unanswered question why children in Holland are almost 'beaten to death' with it. If they do not converge, however, we will analyze why not, and see how the '*folk theorem*' fits into this.

For each learning scheme, we will first consider a population of 60 players, all learning following the same type of scheme. They play 100,000 periods of the basic exchange game. In each period the players are randomly and anonymously matched with each other. The results we present are based on ten of these independent runs for each learning scheme.

Figures 1a to 5a present the relative frequency distribution for the thresholds chosen by the 60 players in the ten runs. As we want to focus on the convergence issue, the figures are based on the periods 50,001 to 100,000 only. Figures 1b to 5b present for each learning scheme the time-series of the average, the 5th percentile, and the 95th percentile of the thresholds chosen. These

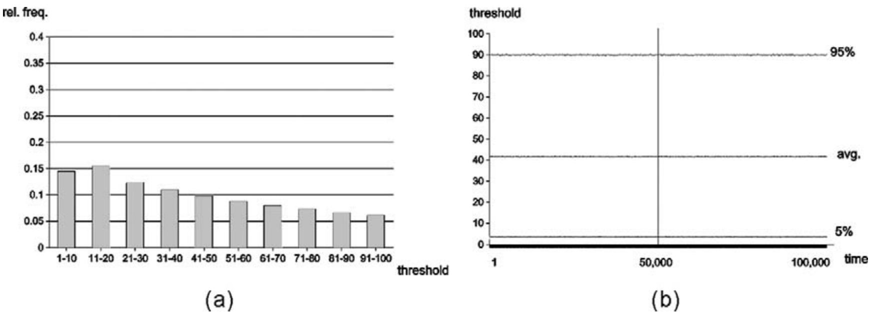

Figure 1. (a) Frequency distribution best-responses, periods 50,001 to 100,000; (b) Time-series best-responses.

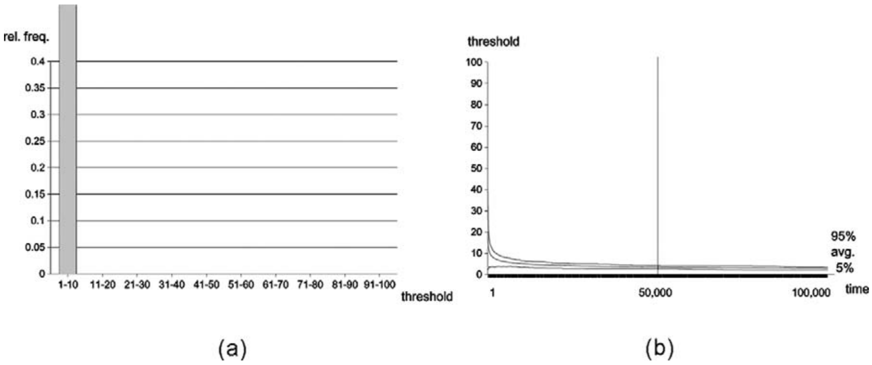

Figure 2. (a) Frequency distribution fictitious play, periods 50,001 to 100,000; (b) Time-series fictitious play.

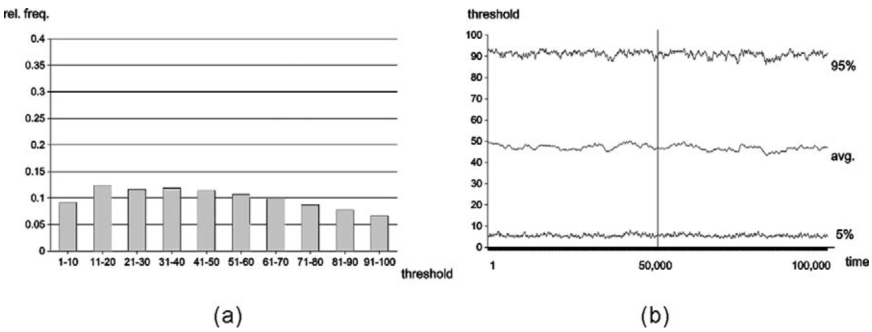

Figure 3. (a) Frequency distribution hill-climbing, periods 50,001 to 100,000; (b) Time-series hill-climbing.

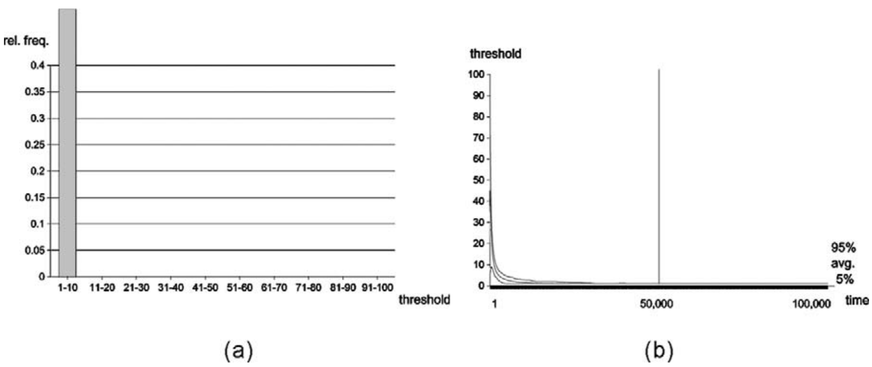

Figure 4. (a) Frequency distribution learning direction theory, periods 50,001 to 100,000; (b) Time-series learning direction theory.

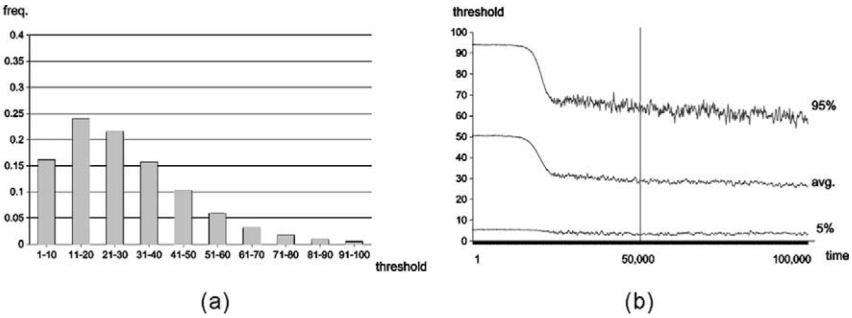

Figure 5. (a) Frequency distribution reinforcement learning, periods 50,001 to 100,000; (b) Time-series reinforcement learning.

three variables are computed for each single period in each run. For presentational reasons, we averaged the variables over blocks of 100 periods.

The frequency distributions show that only fictitious play (Figure 2a) and learning direction theory (Figure 4a) converge to all thresholds being in the 1–10 class. For the other learning schemes the modal threshold class is 11–20. With hill-climbing (Figure 3a) the distribution is almost uniform, whereas the best-response distribution (Figure 1a) is slightly more skewed in favor of the lower thresholds, and the distribution for reinforcement learning (Figure 5a) even more so.

The time-series for best-responses (Figure 1b) and hill-climbing (Figure 3b) are rather constant, showing no trend after the first observation covering a block of 100 periods, with an average over the last 50,000 periods of 41.62 for the former, and 46.56 for the latter. With reinforcement learning (Figure 5b) we see a sudden change after about 20,000 periods, which is due to the decrease of the noise term in the logit choice function[7]. In addition, the series show a weak but significant trend. Therefore, we run the reinforcement learning scheme also for ten million periods. After about one million periods, the average threshold stabilizes at a level of 21.36 (whereas it was 27.95 over the periods 50,001 to 100,000). The frequency distribution for the ten million period run is slightly more skewed towards the left, with the modal class still being the 11–20 one. With fictitious play (Figure 2b) and learning direction theory (Figure 4b) we see an initial learning effect before they become constant at very low threshold levels. It is, however, only the learning direction theory scheme that really converges to the equilibrium. The average threshold over the periods 50,001 to 100,000 is 1.00, whereas with the fictitious play scheme it is 3.12.

Hence, we observe, with one exception, that we do not get convergence to equilibrium, and in general we stay far away from it. Before we further explain this nonconvergence to equilibrium, we will first explore the robustness of this finding in various directions.

First, of all the learning schemes considered, fictitious play has the most serious memory requirements, computing the expected value of tickets to be received on the basis of all past exchanges. What happens if the memory of the players were limited in the following way? Using the fictitious play learning scheme, if a player has not traded for more than 30 periods, he will forget the expected value of tickets to be received, start with a random threshold, and update his beliefs from scratch again (see also the pseudo-code in the appendix). Figures 6a and 6b show the frequency distribution and time-series for the limited memory fictitious play scheme. As we see, apart from the learning direction theory scheme, the only learning scheme that gets close to the equilibrium will not do

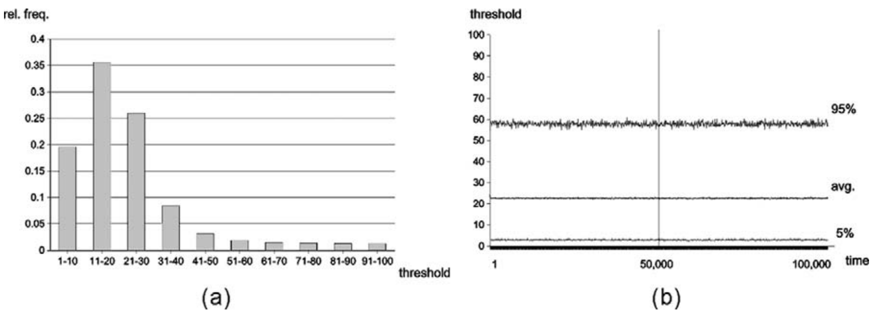

Figure 6. (a) Frequency distribution limited memory fictitious play, periods 50,001 to 100,000; (b) Time-series limited memory fictitious play.

so anymore once we consider a memory limits, with an average threshold of 22.61 in the last 50,000 periods.

Second, could it be that the nonconvergence of some schemes is simply due to some runs being '*stuck*' far away from equilibrium? To answer this question, we look at the differences between the ten runs of each scheme, focussing on the average threshold used over the periods 50,001 to 100,000. Table 1 presents for each learning scheme the run with the lowest average, and the run with the highest average. As we see, all runs are placed within a very narrow band, and there is hardly any difference in the average threshold applied across the runs. The same holds for the margins occurring for the 5th and 95th percentiles (not shown here). In other words, the nonconvergence is a robust phenomenon across runs.

Third, but what about robustness across individual players? In the nonconverging learning schemes there is no single individual that converges to equilibrium, not even in the weak sense of choosing all thresholds within the 1–10 class. We do not show any graphs or tables here, but the frequency distributions of individual players look very similar to the ones presented per learning scheme above. What this shows is that the nonconvergence is not due to some individuals being '*stuck*' far away from equilibrium while the others actually converge[8].

Fourth, the behavior we examined above arose in homogeneous populations, in which each player was learning according to the same type of scheme. It could be that the outcomes would be

TABLE 1

Average thresholds periods 50,001 to 100,000 for ten runs.

	Run with lowest avg.	Run with highest avg.
Best-response	41.58	41.65
Fictitious play	2.86	3.42
Hill-climbing	44.56	48.19
Learning direction theory	1.00	1.01
Limited memory fictitious play	22.58	22.68
Reinforcement learning	27.72	28.29

different in a *heterogeneous* population. In particular, one could conjecture the following two effects. It could be that players choosing thresholds closer to equilibrium might pull others' thresholds down. But it could also be the other way round. Therefore, we analyze the behavior of heterogeneous populations of 60 players, in which each of the six learning schemes considered (best-response behavior, fictitious play, learning direction theory, hill-climbing, reinforcement learning, and limited memory fictitious play) is used by ten players. Figures 7a and 7b show the results. We observe that the higher threshold classes (above 30) have virtually the same relative frequencies as the average of the six homogeneous populations, but there is an enormous drop in the frequency of the 1–10 class, and corresponding increases in the 11–20 and 21–30 classes. As a result, the latter is now the modal threshold class. The average threshold chosen during the periods 50,001 to 100,000 is also *higher* in the mixed population than in the average of the six homogeneous populations (30.56 against 23.81). The run with the lowest average threshold chosen in the mixed population has an average of 29.43 (and the highest 31.93). Hence, if we look at Table 1, we see that even the lowest run with a mixed population has a higher threshold than the average of the highest runs of the six learning schemes in homogeneous populations (which is 24.21). In other words, we see that it is not true that some players smart enough to learn the equilibrium strategy (like the learning

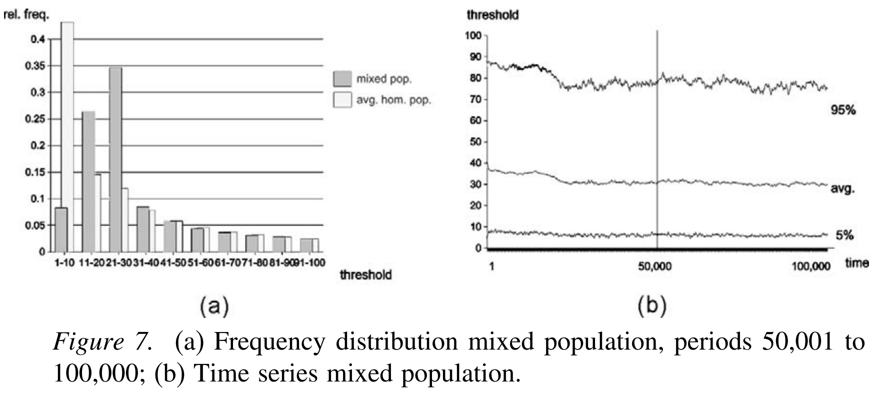

Figure 7. (a) Frequency distribution mixed population, periods 50,001 to 100,000; (b) Time series mixed population.

direction theory players) are sufficient to induce equilibrium play in the population[9].

What is it in the structure of this game that makes that a wide range of learning schemes (with one exception) does not converge to the equilibrium? As the equilibrium is approached the frequency of exchanges goes down. What matters is not so much that this might slow down the learning process, but that the cases in which the feedback to the players points towards the equilibrium become rarer and rarer. If a player does not want to exchange himself, he will not receive feedback about the value offered in exchange by the other player. But if the other player does not want to exchange, then there are multiple best-responses for him. Any threshold is as good as any other. Sooner or later most learning schemes respond to this. Hence, there is some kind of endogenous random drift[10]. What is more, as the frequency of exchanges goes down, the possibility of memory limits becomes more relevant, leading players away from equilibrium. Finally, if, for whatever reason, other players choose thresholds away from equilibrium, then the best a player can do is choosing thresholds higher than the equilibrium one as well. Hence, a fraction of *'less smart'* players may induce all players to stay away from equilibrium.

Why, then, does learning direction theory appear more successful at finding the equilibrium strategy in this exchange game? Learning direction theory still does not use any feedback if no exchange takes place. But when it receives exchange feedback, it reacts very directly and myopically to single successes or failures, instead of accumulating evidence to estimate (implicitly or not) some expected payoff. In some situations this impulsive behavior may lead to undesirable outcomes, but in the exchange game it seems to help avoiding to get 'trapped' for too long in exchanges. What is more, learning direction theory incorporates more reasoning about alternative strategies than some other learning models do. For example, reinforcement learning or fictitious play only update the expected payoff for the thresholds actually used, whereas learning direction theory reasons that if, for example, the end ticket value after an exchange was high, a higher threshold would have been even better, as it would have made the exchange occurring even more likely. This extra reasoning increases the learning speed, and may again help avoid getting trapped in exchanges for too long.

Besides the obstacles inherent in the incentive structure of the game that make it difficult for the learning schemes considered to reach equilibrium, in reality there might be additional factors playing a role. First, in reality the game is played with overlapping generations. Each period some new players enter the population, choosing thresholds at random, while others leave. Second, in reality interactions are probably not determined randomly but endogenously. For example, players who want to exchange may stop bothering players who refuse to do so. Or players may avoid players from whom they got low values only. Third, the only effect of non-equilibrium play upon other players we considered was through the payoffs. But in reality players might learn from the *behavior* of other players as well. Fourth, in reality risk-aversion or regret might provide additional incentives for the players to adhere to the equilibrium strategy. It is only for this last point that more equilibrium play would seem obvious, but the third factor (behavioral learning) deserves some more attention.

Let us first assume that an agent observes the behavior of all other players. Following models of technology adoption (see, e.g., Arthur [1989]), assume that the probability to adopt a certain behavior is a function of the proportion of the population displaying that behavior. In particular, let us assume that this adoption function is monotonically increasing. This leads to a dynamic process for which in principle any state could be a fixed point (depending on the exact shape of the adoption function), including the two most extreme outcomes, and for many adoption functions there might be multiple stable fixed points. Hence, in terms of our exchange game, behavioral learning as such (even when the adoption function is monotonically increasing) might lead to nobody ever exchanging, as well as everybody always exchanging, and everything in between. None of these possibilities can be singled out without making additional assumptions. One such possible assumption would be to combine behavioral learning with one of the other learning theories considered as follows. If a player exchanges, he learns from the feedback provided by the resulting payoff. But if no exchange takes place because a partner refuses to do so, a player learns from his partners behavior. It is not clear, however, what a player would learn from this observation. Instead of imitating his playfellow, behavioral learning could imply the opposite[11]. For example, if a player observes somebody hesitating to exchange, that might increase his own eagerness to exchange, since the refusing player must be somebody with a relatively high value (after all, it is a zero-sum game). In other words, the adoption curve for behavioral learning might even be decreasing instead of increasing.

Some preliminary insights concerning the empirical relevance of our analysis can be derived from a pilot experiment with undergraduate students from an Experimental Economics class that we organized at Queen Mary in 2001. Sixteen subjects played eight rounds of the exchange game described in this paper, with random matching in each round. Only one player made all his choices in agreement with the equilibrium strategy, the fifteen others did not. Hence, there clearly is some scope for learning. Assuming that players use a threshold strategy, i.e., that they are prepared to exchange any ticket up to a certain value, we define a player as *learning* if he decides to keep in some round a ticket with a value that he would have exchanged earlier on. Similarly, we define *unlearning* as offering in some round a ticket with a value that he would have decided to keep earlier on. A player displays *constant* behavior if he shows neither learning nor unlearning. The behavior of all sixteen players was consistent with constant behavior. In other words, for none of the sixteen players did we observe one single instance conforming to either learning or unlearning, and in fact the average value offered for exchange did not come down over the eight rounds. Hence, we cannot refute the hypothesis that all players used a constant threshold value.

Given the general nonconvergence of the learning processes considered, the reason for the parents to give such persistent advice is clear. If their children employ any non-equilibrium strategy, other strategies might come along that capitalize on them, leading to losses, since after all it is a zero-sum game. Now, some people might argue that even if learning as such should not be expected to converge to equilibrium play, the following evolutionary argument for convergence should still be convincing. If some players systematically make losses, they would disappear from the population, as they would need a never ending inflow of resources to continue playing the exchange game. Although the argument seems correct, it might provide us with an additional explanation for the existence of the *'folk theorem'*. In this exchange game between children, it is the parents who would need to provide such an inflow of resources. Hence, we have some kind of principal-agent problem. There is little incentive for the children to change their behavior. But it is costly for the parents, and therefore they spend so much time and energy on keeping the *'folk theorem'* alive.

Finally, one could see the continuing reminders of the folk theorem as a form of virtual reinforcement learning (see Vriend [1997]). Parents teach children that exchange leads to the outcome tears. Getting this information over and over again plays the same role as actually experiencing this bad outcome. In other words, what the folk wisdom essentially does is speeding up this reinforcement learning process, as it allows to learn even without any exchanges actually taking place.

4. CONCLUDING REMARKS

This paper provides two contributions to the literature. First, we showed that a piece of Dutch *'folk wisdom'* concerning exchange can be explained and justified by the game-theoretic analysis of a specific exchange game with two-sided asymmetric information. Second, we explained that keeping this *'folk wisdom'* alive is important because learning by boundedly rational agents is inherently slow in this type of situation.

Years ago I was stranded in Brussels on a Sunday night, with just a little bit of Belgian money. To add insult to injury, it turned out that these coins had been taken out of circulation some time ago. Undeterred, we started searching for somebody naive enough to exchange our money. Although that was not easy, in the end we succeeded. Trying in vain to suppress a big smile, we approached a booth to buy some food, whereupon the smile disappeared quickly. As it turned out, we had received only some almost worthless

French centimes. But tears? No. After all, this had been a typical equilibrium exchange; our ticket with the lowest possible value for their ticket with the lowest value.

ACKNOWLEDGMENTS

I wish to thank Jonathan Haskel, Steffen Huck, Hugo Keuzenkamp, Paola Manzini, Marco Mariotti, an anonymous referee, and the editor for helpful discussions or comments on earlier versions of this paper. The usual disclaimer applies.

APPENDIX. PSEUDO-CODE

Tables A1 to A5 present the pseudo-code for each of the following five learning schemes: best-responses, fictitious play, learning direction theory, hill climbing, and reinforcement learning. Table A6 lists a variant of fictitious play in which players forget everything about expected values of tickets to be received when they have not exchanged any ticket in the last 30 periods.

TABLE A1

Best-response behavior.

```
procedure BESTREP;
begin
 for each period do
 begin
  if first period the threshold is random
  else
  begin
   if exchanged in previous period then
    threshold := end_ticket;
   if opponent did not want to exchange then
    threshold is random;
  end;
 end;
end;
```

TABLE A2

Fictitious play.

```
procedure FICT;
begin
 for each period do
 begin
  if never exchanged before then
   threshold is random
  else if exchanged at least once in past
   then threshold := expected value tickets
    received through exchange;
 end;
end;
```

TABLE A3

Learning direction theory.

```
procedure LDT;
begin
 for each period do
 begin
  if first period then threshold is random
  else
  begin
   if exchanged in previous period then
   begin
    if end_ticket < start_ticket then
     threshold := threshold - 1;
    if end_ticket > threshold then
     threshold := threshold + 1;
   end;
  end;
 end;
end;
```

TABLE A4

Hill climbing.

```
procedure HILL;
begin
 for each period do
 begin
  if first two periods then threshold is random
  else
  begin
   if threshold[t-1] > threshold[t-2]
      and end_ticket[t-1] > end_ticket[t-2]
      then delta := 1;
   if threshold[t-1] > threshold[t-2]
      and end_ticket[t-1] < end_ticket[t-2]
      then delta := -1;
   if threshold[t-1] < threshold[t-2]
      and end_ticket[t-1] > end_ticket[t-2]
      then delta := -1;
   if threshold[t-1] < threshold[t-2]
      and end_ticket[t-1] < end_ticket[t-2]
      then delta := 1;
   if threshold[t-1] = threshold[t-2]
      or end_ticket[t-1] = end_ticket[t-2]
      then with equal probability either
      delta := -1 or delta := 1;
   threshold := threshold+delta;
    (with min. threshold being 1 and max.
     threshold 100)
  end;
 end;
end;
```

TABLE A5

Reinforcement learning.

```
procedure REINFORCE;
begin
 for each period do
 begin
  if first period then β:=0.00001 else β:=β·1.0005;
   (with max. β being 100)
  if first period then for each threshold m_j
   do avg_payoff[m_j]:=100;
  choose each threshold m_j with probability
    exp(β·avg_payoff[m_j])/Σ_i exp(β·avg_payoff[m_i]);
  if threshold m_j was used in this period
   then avg_payoff[m_j]:=0.95*avg_payoff[m_j]
    +0.05*end_ticket;
 end;
end;
```

TABLE A6

Limited memory fictitious play.

```
procedure LM_FICT;
begin
 for each period do
 begin
  if never exchanged before or not exchanged
   in last 30 periods then
  begin
   clear memory concerning average value
    tickets received;
    threshold is random;
  end;
  else if exchanged at least once
   in last 30 periods
    then threshold:=average value tickets
     received in exchange
      since last clearance of memory;
 end;
end;
```

NOTES

1. See, e.g., Osborne and Rubinstein [1994].
2. It is interesting to note that the 'folk theorem' is much older than the theoretic analyses of games with asymmetric information.
3. The game considered is a zero-sum game. One could, however, perhaps more realistically, assume that children may be bored with their own toy (although, of course, one could also argue that children are more attached to their own toys). That is, if player i currently holds a ticket with value v_i, the value to agent i himself is only $\alpha \cdot v_1$, with $0 < \alpha < 1$. If α is not too small, this does not fundamentally change the situation the players are in, and it does not change the flavor of the analysis.
4. See also Gresham's law that '*bad money drives out good*'.
5. This leaves the question open as to how important the differences (indicated in Section 2) between the exchange game and markets for lemons are in this respect. Notice that the issue of learning by boundedly rational agents in markets for lemons has received (surprisingly) little attention in the literature. A recent exception is Feltovich [2003].
6. We will focus on convergence, and we will skip the issue of accumulated losses during the learning process. One reason is that they will depend to some extent on some arbitrary parameters determining the speed of learning. In any case, as we will see below, considering accumulated losses would only strengthen our conclusion.
7. Initially choices are essentially random. Notice that this implies that initial beliefs of expected values are irrelevant; by the time a player becomes really discerning he has tried all thresholds many times.
8. This still leaves open the question whether the distribution in the population in a given period corresponds to the distribution over time for a given player, or that the players move in synchrony. As our primary focus is on the (lack of) convergence issue, we do not provide a full analysis of this. However, the time-series figures suggest that, apart from the initial learning effects, the distribution of choices is relatively steady over time, at least when averaged over blocks of 100 periods.
9. Given the different performances of these schemes, it would be an interesting question to ask whether the players would be able to *learn* which learning scheme to use.
10. This nonconvergence related to the expansion of the best-response set as equilibrium is approached was also observed experimentally in a Hotelling location game with 4 players (see Huck et al. [2002]).
11. Experimental evidence in Cournot oligopolies suggests that players may be rather reluctant to imitate (see Bosch-Domenèch and Vriend [2003]).

REFERENCES

Akerlof, G. (1970), The market for lemons, *Quarterly Journal of Economics* 84, 488–500.

Arthur, W.B. (1989), Competing technologies, increasing returns and lock-in by historical events, *Economic Journal* 99, 116–131.

Bosch-Domènech, A. and Vriend, N.J. (2003), Imitation of successful behaviour in Cournot markets, *The Economic Journal* 113, 495–524.

Feltovich, N. (2003), *Slow Learning in the Market for Lemons: A Note on Reinforcement Learning and the Winner's Curse* (mimeo) Paper presented at CIEF2003.

Goeree, J.K. and Holt, C.A. (2001), Ten little treasures of game theory and ten intuitive contradictions, *American Economic Review* 91(5), 1402–1422.

Huck, S., Müller, W. and Vriend, N.J. (2002), The East End, the West End, and King's Cross: On clustering in the four-player Hotelling game, *Economic Inquiry* 40(2), 231–240.

Kessler, A.S. (2001), Revisiting the lemons market, *International Economic Review* 42(1), 25–41.

Krugman, P.R. and Obstfeld, M. (1994), *International Economics. Theory and Policy (3rd ed.).* HarperCollins, New York.

Myerson, R.B. and Satterthwaite. (1983), Efficient mechanisms for bilateral trading, *Journal of Economic Theory* 29, 265–281.

Osborne, M.J. and Rubinstein, A. (1994), *A Course in Game Theory.* Cambridge, MA: MIT Press.

Selten, R. and Stoecker, R. (1986), End behavior in sequences of finite prisoner's dilemma supergames. A learning theory approach, *Journal of Economic Behavior and Organization* 7, 47–70.

Vriend, N.J. (1997), Will reasoning improve learning?, *Economics Letters* 55(1), 9–18.

Address for correspondence: Nicolaas J. Vriend, Queen Mary, University of London, Department of Economics, Mile End Road, London, E1 4NS, UK. E-mail: <n.vriend@qmul.ac.uk>, www.qmul.ac.uk/∼ugte 173/.