Universität Hannover

Diskussionspapiere des Fachbereichs Wirtschaftswissenschaften

# Cournot or Walras?
# Agent Based Learning, Rationality, and Long Run Results in Oligopoly Games

—

# Cournot oder Walras?
# Agentenbasiertes Lernen, Rationalität und langfristige Resultate in Oligopolspielen

Thomas Riechmann*

Diskussionspapier Nr. 261
ISSN 0949 – 9962
August 2002

*University of Hannover, Faculty of Economics, Königsworther Platz 1, 30 167 Hannover, Germany, e–mail: riechmann@vwl.uni-hannover.de

**Abstract**

Recent literature shows that learning in oligopoly games might in the long run result in the Cournot *or* in the Walras equilibrium. Which outcome is achieved seems to depend on the underlying learning dynamics. This paper analyzes the forces behind the learning mechanisms determining the long run outcome. Apart from the fact that there is a difference between social and individual learning, the key parameter is shown to be the degree of rationality of the learning agents: Learning the Cournot strategy requires the agents to acquire a large amount of information and to use sophisticated computational techniques, while the Walras strategy can be shown to be a particular 'low rationality result'.

**Zusammenfassung**

Neuere Veröffentlichungen zeigen, dass Oligopolspiele auf lange Frist sowohl im Cournot– als auch im Walras–Gleichgewicht enden können. Welches Ergebnis erreicht wird, scheint von den zugrundeliegenden Lerndynamiken abzuhängen. Dieses Papier widmet sich der Frage, welches die Kräfte hinter den Lernmechanismen sind, die das langfristige Resultat bestimmen. Neben dem Unterschied zwischen sozialem und individuellem Lernen kann als wichtigster Einflussfaktor der Grad der Rationalität der lernenden Agenten identifiziert werden: Um die Cournot–Strategie zu lernen, benötigen die Agenten viele Informationen und komplizierte Techniken, während das Walras–Gleichgewicht als „Niedrig–Rationalitäts–Resultat" identifiziert werden kann.

**Key words:** Agent Based Economics, Oligopoly Games, Learning, Rationality, Spite Effect, Evolutionary Algorithms

**JEL classifications:** C63 – D43 – D83

# 1 Introduction

The Cournot model of oligopolistic quantity choice is one of the oldest and one of the best analyzed and most widely understood models in game theory. For years, things seemed to be very clear: As long as there is a finite number of players in this game, the result will be the Cournot–Nash equilibrium. But, since the work on evolutionary learning in the Cournot game by Vega-Redondo (1997), at the latest, things are not so clear any more: Vega-Redondo shows that under the regime of evolutionary forces, the unique long run outcome of the game will no longer be the Cournot–, but the Walrasian competitive market equilibrium. This result holds for any finite number of players. The question remains: What is it that decides whether the outcome of learning in the Cournot model is Cournot or Walras? This paper gives some answers to this question.

It has been shown that the driving force leading to the surprising result of Vega-Redondo (1997) is to be found in the underlying dynamics, which are forms of the usual replicator dynamics. Thus, it can be concluded that it is the type of learning in the model that determines the type of outcome. A paper in this direction of thought is the one by Vriend (2000), who argues that the most important influence is the difference between individual and social learning. Following Vriend, social learning processes lead to the Walrasian outcome, while individual learning tends to converge to Cournot. This paper will show that it is indeed the type of learning which determines the outcome, but that a distinction into social and individual learning is not enough. While social learning will inevitably lead to the Walrasian outcome, under the regime of different types of individual learning both, Cournot or Walras, can be the result. The second force influencing the quality of the result is the degree of rationality of the learning firms or agents: If agents are smart, they will learn to play Cournot, if they are not, the result will be Walras.

This paper proceeds as follows. First, the model of this paper will be briefly introduced, which is a simple model of the Cournot type. This model will provide the economic background for every type of learning dynamics analyzed throughout this paper.

Then, in a second step, the field of social learning will be visited. The section starts with an informal and intuitive review of the main driving force of evolutionary dynamics in this model, the spite effect. Making use of the concept of spite, it is shown that the Walrasian outcome does indeed represent the only stable symmetric Nash equilibrium of the Cournot game under the regime of social learning.

After that, the paper briefly describes the concept of stochastic stability, which is the major technical concept underlying the results of Vega-Redondo (1997). A look at the structure of agent based models relying on dynamics generated by evolutionary algorithms (EAs)[1] shows that these algorithms are capable of closely re-

[1] Evolutionary algorithms are a family of simulation methods based on the principles of the Darwinian evolution. The most prominent member of the EA family is the genetic algorithm (Holland, 1992; Goldberg, 1989), which has been successfully used in economics before, see e.g. Dawid (1999); Riechmann (2001b).

sembling the evolutionary dynamics of replicators, Vega-Redondo's 1997 results are based on. In the following, based on these algorithms, an inductive method is developed which opens the chance of analyzing the long run results even of models which cannot be described analytically. This method still keeps the spirit of the concept of stochastic stability. Thus, at the end of the section on social learning, the EA based method for the analysis of long run results is applied to a variant of the evolutionary Cournot model which closely resembles the original setup by Vega-Redondo (1997). It is shown that the method does indeed reproduce Vega-Redondo's results.

In the next section, the paper turns to individual learning. First, the general structure of individual learning models in contrast to the structure of social learning models is illustrated. After that, three models of individual learning are introduced representing processes of learning by agents with different degrees of rationality. For each of the models, the information and the abilities agents need in order to 'learn' are explicitely accounted of. The model representing a high degree of rationality is a model with agents learning by using the best response technique. The outcome of this learning method is Cournot. The second model, in contrast, is a model of naive low rationality learning, a learning method often applied in textbook cobweb models. Cobweb learning in the model of this papers and for a given, well behaved, parameter set, leads to convergence towards the Walras equilibrium. The third model is a model of medium rationality: Agents learn by computing a best response to last period's market price, at the same time taking account of the fact that they are able to influence the current market price. This type of learning is shown to result in the Cournot equilibrium in the long run. The results of these three models are then used to stress the central hypothesis of this paper: The more rational agents are, the more likely they are to learn to play Cournot. The paper ends with a summary.

## 2 The Model

The model is a simple variant of the standard textbook Cournot model of quantity choice in oligopolies.

The demand of the market in period $t$, $D_t$, is time invariant and exogenously given as

$$D_t = A - B\,p_t \tag{1}$$

with $A$, $B$ as positive parameters and $p_t$ giving the equilibrium price in period $t$.

Let $s_{i,t}$ denote the quantity firm $i$ supplies in period $t$. Assume that firms must supply non–negative quantities and let firms be restricted by a maximum capacity $s_{\max}$, such that $s_{i,t} \in \mathcal{S} = [0, s_{\max}]$. Aggregate supply in $t$, $S_t$ is given as the sum of the supplied quantities of the $n$ firms involved with the model:

$$S_t = \sum_{i=1}^{n} s_{i,t}\,. \tag{2}$$

From (1) and (2), the equilibrium price in $t$, $p_t$ results as

$$p_t = \frac{1}{B}\left(A - \sum_{i=1}^{n} s_{i,t}\right). \tag{3}$$

Each of the $n$ firms involved has the same cost function $C(\cdot)$ which is quadratic in the quantity produced

$$C(s_{i,t}) = \frac{1}{2}\delta s_{i,t}^2. \tag{4}$$

Fixed costs can be neglected without loss of generality. Marginal cost are constant.

The profit of each firm $i$ in $t$ is given by

$$\pi_{i,t} = p_t s_{i,t} - C(s_{i,t}). \tag{5}$$

Substituting $p_t$ in (5) by (3) shows that the problem is problem of state dependence, or, to put it shorter, that the problem constitutes a game. The profit of firm $i$ depends on the supply strategies of *all* firms in the market:

$$\pi_{i,t} = \frac{1}{B}\left(A - \sum_{j=1}^{n} s_{j,t}\right) s_{i,t} - C(s_{i,t}). \tag{6}$$

Assuming identical equilibrium behavior of all firms and letting them all maximize their profit by selecting the best quantity while considering that every other firm will do the same leads to the usual Cournot–Nash equilibrium. In the case of the model presented here, this means that the optimal quantity $s_{i,t} = s^C \; \forall \; i,t$ is given by

$$s^C = \frac{A}{B\delta + n + 1}. \tag{7}$$

It is important to keep in mind that $s^C$ is the optimal quantity computed by firms knowing that their influence on the market price is non–negligible. If, on the other hand, firms do not care about their influence on the market price and behave as mere price takers instead, the outcome will differ from the Cournot quantity (7). In this case, the outcome will be all firms producing the usual competitive market equilibrium quantity, for convenience labeled as the 'Walras–quantity' in the rest of this paper. This quantity, $s_{i,t} = s^W \; \forall \; i,t$, is

$$s^W = \frac{A}{B\delta + n}. \tag{8}$$

It is easy to recognize that for large populations the difference between the Cournot and the Walras quantity vanishes:

$$\lim_{n\to\infty} \left| s^C - s^W \right| = 0. \tag{9}$$

An interesting question to ask is which of the possible quantities, or even different ones, firms would produce if they developed their quantity decisions over time under the regime of different kinds of learning process.

# 3 Social Learning

Social learning means learning processes with agents learning from one another. Thus, models based on populations of truely interacting agents are models of social learning. Figure 1 displays the general structure of social learning processes for the Cournot game. Each firm is characterized by only one piece of data, which is its strategy or the quantity it plans to supply at the market. Learning takes place in form of an inter–agent process i.e. from firm to firm. By this method, firms update their strategies in order to use them at the market. There, at the market, by the interaction of aggregate demand and aggregate supply, the market price is generated, which is the most important feedback to the agents, signaling the quality of their strategies.

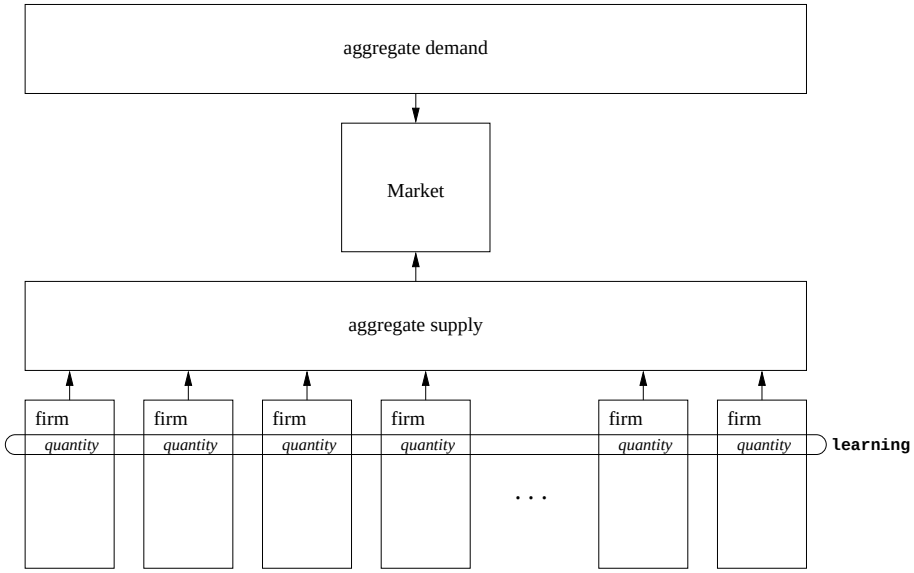


Figure 1: General Structure of Social Learning Models

In recent literature, there are mainly two types of models aiming to describe social learning in games. One type of models are models from evolutionary game theory, having their dynamics based on replicator equations. The other type of models are agent based models, grounding their dynamics on evolutionary algorithms.

## 3.1 Social Learning and the Spite Effect

It has been mentioned above that evolutionary learning in the Cournot model leads to a long run outcome which is Walras. The technical reasons for this will be reviewed below (Secs. 3.2 and 3.3). The intuitive behavioral reason for this outcome is the so called 'spite effect' which is discussed in the following.

The driving force of evolutionary dynamics is relative rather than absolute pay-

off or fitness, i.e. the difference between a player's payoff and the population mean payoff or, alternatively, the proportion of a players payoff at the aggregate payoff of the player's population. At the start of a new period, agents engage in the process of imitation, which is a kind of imitation of the population of the period before: Agents in the current population imitate those agents that performed relatively well in the last period. Imitation here means imitation of success, i.e. the imitation of *better* or *best* strategies of the last period. The proportion of a certain strategy in a population of strategies grows the faster, the larger the difference between this strategy's payoff and the population mean payoff, i.e. relative payoff, is. In order to spread throughout the population, a strategy has to be better than most of the other strategies. Again: It is relative, not absolute payoff that counts. Thus, the kind of behavior evolutionary dynamics implies is maximization of relative, not maximization of absolute payoff. Seen this way, evolutionary dynamics are not truely appropriate for modeling agents' learning to maximize their profits. In fact, all that agents do under the regime of these dynamics is (try to) maximize the difference between their own and the other agents' profits.

This idea can be clarified by a simple example. Table 1 gives the normal form of a 2–player–2–strategy stage game. According to the usual rules in evolutionary game theory, players are restricted to playing pure strategies only.

| | | Player B | |
|---|---|---|---|
| | | $s_1$ | $s_2$ |
| Player A | $s_1$ | $a, a$ | $c, b$ |
| | $s_2$ | $b, c$ | $d, d$ |

Table 1: Spite Game, absolute payoffs; $a > b > c > d$

Strategy $s_1$ is a strictly dominant strategy for both players. Thus, both players should be expected to play the dominant equilibrium $(s_1, s_1)$. If, however, the vital criterion of success is relative payoff (or relative fitness), things change dramatically: Now each player is better off deviating from $s_1$ and playing $s_2$. While A's payoff of playing $s_1$ against B's $s_1$, $\pi_A(s_1, s_1) = a$ is the maximum absolute payoff, A's relative payoff can be increased: If A switches to $s_2$ (assuming that B still plays $s_1$), her relative payoff is greater than before. $\pi_A(s_2, s_1) = b$ and B's payoff is $\pi_B(s_2, s_2) = c$, which means that by playing $s_2$, A performs better than B (but not better than before). A has lost absolute but at the same time gained relative payoff. In order to make this even clearer, Table 1 is transferred into Table 2 by transforming absolute into relative payoffs. Relative payoffs for player $k$ playing strategy $i$ against player $-k$ playing $j$, $\pi_k^r$, in Tab. 2 are calculated as the absolute payoff to player $k$ minus the mean absolute payoff to players $k$ and $-k$:

$$\pi_k^r(s_i, s_j) = \pi_k(s_i, s_j) - \frac{1}{2}\left[\pi_k(s_i, s_j) + \pi_{-k}(s_i, s_j)\right]. \tag{10}$$

This reflects the most frequently used way of calculating relative payoffs in contin-

| | | Player B | |
|---|---|---|---|
| | | $s_1$ | $s_2$ |
| Player A | $s_1$ | $0, 0$ | $\frac{1}{2}(c-b), \frac{1}{2}(b-c)$ |
| | $s_2$ | $\frac{1}{2}(b-c), \frac{1}{2}(c-b)$ | $0, 0$ |

Table 2: Spite Game, relative payoffs

uous time replicator dynamics (Samuelson 1997, p. 66; Weibull 1995, pp. 72–74).

It is clear to see that in this game, explicitely accounting for relative payoffs, the profile $(s_2, s_2)$ constitutes a dominant equilibrium. Thus, if it is relative payoff that counts as criterion of success, both players will play $s_2$. Note, that in the example, not the game itself has changed, but only the way of measuring success.

In other words: If relative payoff is the measure of success, it pays to hurt yourself (in terms of absolute payoff) as long as by hurting yourself you hurt your opponent even more. This is the true meaning of the term 'spite effect'. Note that this type of spiteful behavior is a result of the dynamics implied, not of the game itself. Deciding to use evolutionary dynamics as a model of social learning in games *automatically* means imposing spiteful behavior to the agents.

Social learning in models of this type means agents learning from one another. The crucial point at social learning is the fact that people need not hold spiteful motives in order to display spiteful behavior. If the only way to find a better strategy is to look around what others do and than to eventually imitate one of the better strategies other agents use, this way of imitation automatically gives rise to the spite effect. Agents in these models do not intentionally maximize relative payoff (maybe because they hate each other) but imitate better strategies just because these strategies are better than the ones the agents applied before. If agents are not allowed to learn by introspection (or if they are not capable of doing so), the imitation of others is the only way of learning that is left. As long as learning by imitation means imitation of other agents' behavior, and not by imitation of one's own behavior from the past, this type of social learning leads to spiteful behavior even if agents do not have spiteful motives. To put it in sufficiently short words: Social learning *means* spiteful behavior. This implication is independent of the degree of rationality of the agents.[2]

Based on this idea, it is straightforward to ask what would be the optimal strategy if behavior is spiteful. The next section (Sec. 3.2) will be devoted to this question. It shows that under the regime of spite agents' optimal strategy is to play Walras.

[2]Of course, the mere fact that agents feel that they must be learning by imitation of others and are thus incapable of introspective learning might be a hint that these agents are not too rational.

### 3.2 Spiteful Behavior in the Cournot Model

In contrast to the game of Tabs. 1 and 2, in most types of games, there is no particular effect of spiteful behavior. In most games, the structure of payoff tables written in absolute and relative payoffs is simply the same. This of course means that in most types of games, spiteful behavior generates the same results 'normal' maximization of absolute payoff does. The Cournot game is a rare exception to this rule. While for 'normal' profit maximization, the only Nash–equilibrium is the Cournot–Nash equilibrium, for spiteful behavior, it is the Walrasian competitive market equilibrium. This can easily be shown using the model introduced in Section 2.

Each firm has absolute payoff depending on its current strategy $s_i$ and the market price $p$ as given in (5) above.[3] Population mean payoff is given by

$$\overline{\pi} = \frac{1}{n} \sum_{i=1}^{n} \pi_i (s_i) \,. \tag{11}$$

Relative payoff of firm $i$ is given by the difference of the firm's payoff and the population mean payoff:

$$\pi_i^r (s_i) = \pi_i (s_i) - \overline{\pi} \,. \tag{12}$$

Assuming symmetric behavior of all firms $j \neq i$, i.e. $s_j = s_{-i} \; \forall \; j \neq i$, (12) becomes

$$\pi_i^r (s_i) = \frac{n-1}{n} \left[ \pi_i (s_i) - \pi_{-i} (s_{-i}) \right] . \tag{13}$$

Maximization of $\pi_i^r$ with respect to $s_i$ yields

$$p = \frac{\partial C (s_i)}{\partial s_i} + (s_{-i} - s_i) \frac{\partial p}{\partial s_i} \,. \tag{14}$$

For the given example, the best response function implicitly given in (14) can be derived explicitely:

$$s_i^\star = \frac{A}{B\delta} - \frac{n}{B\delta} s_{-i} \,. \tag{15}$$

For totally symmetric behavior, i.e. $s_i = s_{-i} = s \; \forall \; i$, (15) gives

$$s^\star = \frac{A}{B\delta + n} \,. \tag{16}$$

It is easy to recognize that the resulting optimal quantity $s^\star$ from (16) is the same as the Walrasian quantity $s^W$ given in (8).

Note that, as the intersection of $n$ best reply functions, the Walrasian equilibrium formed by each of the $n$ payers playing $s^W$, in the Cournot game in relative payoffs, the Walrasian equilibrium *is* a Nash equilibrium.

The general conclusion to be drawn from this is the following: Irrespective of the particular learning method involved: As long as players behave spitefully, the

[3] As the outline given here needs no dynamics, the time indices $t$ are omitted.

best (symmetric) form of behavior they can find is the Walrasian type of behavior. Thus, the result establishes a benchmark result for social learning methods: A sufficiently good learning method should be able to make agents learn to co–ordinate their behavior to the Walrasian solution.

### 3.3 Evolutionary Game Theory

In the field of evolutionary game theory, it has been shown by Vega-Redondo (1997) that evolutionary dynamics in a setting like the model presented above will in the long run lead to the Walrasian equilibrium rather than the Cournot–Nash equilibrium of the game in absolute payoffs. The central concept underlying this result is the concept of 'stochastic stability' (Foster and Young, 1990; Young, 1993).[4] Evolutionary dynamics are population dynamics. In economics, they describe the development of a distribution of behavioral strategies in a population of agents over time. Evolutionary dynamics based on pure imitation, e.g. dynamics described by the usual forms of replicator equations, are known to lead to homomorphic populations, i.e. populations with all agents playing the same strategy. In order to prevent the dynamics from (possibly premature) convergence to such a homomorphic state, a second evolutionary force is introduced: mutation. Mutation means some agents spontaneously changing their strategy. They discard the strategy they adopted in the process of imitation and use a different one instead. Mutation is usually interpreted as mistakes in imitation (Alchian, 1950) or as a form of learning by experiments. In order not to disturb the process of imitational learning too severely, mutation is usually assumed to take place with only a small mutation probability. More formally, evolutionary dynamics as described by replicator equations can be characterized as a population Markov process: Due to the special form of dynamics, the composition of a population only depends on the composition of the population before. Thus, evolutionary *imitation* dynamics establish a Markov process with a number of absorbing states. Every homomorphic population is such an absorbing state: Once a population consisting of only one strategy is reached, no other strategy can be imitated any more. In a way, evolutionary pure imitation dynamics are processes of the continuous dying out of strategies until there is only one strategy left. This single surviving strategy is the final state of the pure imitation process. Adding mutation changes the scene: Now, there is an anti–force against the dying out of strategies. Consequently, the new learning process is still Markov, but it has no absorbing states any more. Nevertheless, the Markov process is ergodic. This means that for a constant mutation probability, the process has a unique long run distribution of states, the so called limit distribution of the process. This is true for every value of the mutation probability. Particularly interesting, of course, is the limit distribution of the process if the mutation probability becomes deliberately small, i.e. for a learning process with *almost no* mutation, a learning process which is nearly the same as learning by pure imitation. The limit distribution for learning

[4] A related concept is the one of 'long run equilibria' by Kandori et al. (1993); Kandori and Rob (1995).

processes with a mutation probability approaching zero establishes the notion of 'stochastically stable states': Every state with a positive mass in the limit distribution of this learning process is such a state. It is intuitively obvious that these stochastically stable states should be monomorphic states, as they are the absorbing states of the process of learning by imitation alone. But, what is appealing about the concept of stochastic stability is that it helps to distinguish between homomorphic states which are visited with positive probability over the long run of the process and such homomorphic states which are almost surely not visited at all. Vega-Redondo (1997) shows that the limit distribution of the evolutionary learning process including imitation and mutation for the Cournot model is in fact a degenerate distribution: It contains only one state with positive mass. This state is the one that establishes the Walrasian market equilibrium.

### 3.4 Agent Based Models and Evolutionary Algorithms

Cournot models have been analyzed by means of agent based models and evolutionary algorithms before. Some examples are the papers by Arifovic (1994); Dawid and Kopel (1998) and Franke (1998). Nevertheless, it seems as if in most of these papers (an exception from this is the paper by Vriend 2000), the true nature of the economic model analyzed has to some extent been misunderstood by the authors themselves. The models are mostly called models of the 'Cobweb' type, indicating that the question addressed is of macroeconomic rather than game theoretic nature. Consequently, what the authors do is analyze the question whether EA dynamics are capable of generating a Walrasian market equilibrium. The fact that this equilibrium is indeed the global attractor of the EA dynamics is thus no surprise to the authors. At a closer look at the models, their true nature becomes evident, though: As it is impossible to conduct EA simulations based on infinitely large populations, in these models, the number of firms in focus is finite. For finite populations, however, models of quantity choice are *by definition* models of the Cournot type. Accordingly, without further knowledge, the simulations should be expected to lead firms to chose the Cournot– rather than the Walrasian quantity.

Nevertheless, the aim of the above cited papers was not to distinguish Walrasian from Cournot outcomes, but rather to determine the form and quality of EA dynamics as such. Consequently, it is not surprising that Arifovic (1994, p. 24) states she found her EA simulations to converge to '*rational expectations equilibrium values*', clearly meaning the Walras equilibrium. (Although, in a Cournot game, the rational expectations equilibrium should of course be the Cournot–Nash equilibrium.) Thus, in a way, at least some authors knew EAs in Cournot games to converge to the Walrasian outcome for long, but in a way they did not know that they knew this.[5]

[5] A second reason for the authors not recognizing the full value of their results might be the following: According to (9), even for relatively small populations, the Walras– and the Cournot–quantity (and the resulting equilibrium prices) become so similar, that in the presence of the 'noise' caused by the mutation operator, it is nearly impossible to tell if a certain simulation result is Cournot

In this paper, the Cournot model will be explicitely faced as such, i.e. as a game involving only finitely many players. It has been shown before, that evolutionary algorithms in agent based economic models are appropriate tools of analyzing evolutionary games (Riechmann, 2001a). Moreover, it has been found that EAs, too, establish Markov processes (Dawid, 1994; Riechmann, 1999) which parallel the evolutionary dynamics of replicator equations without and with mutation. It has already been demonstrated that evolutionary replicator dynamics and the dynamics of replication as used in evolutionary algorithms are equivalent (Riechmann, 2001c). Thus, it should be no surprise that EAs tend to generate the same long run results as replicator dynamics.

Nevertheless, EAs provide a suitable way of analyzing the long run results of learning processes, even if these learning processes are *not Markov* and can thus not be analyzed with the help of the concept of stochastic stability, e.g. learning processes based on agents with certain forms of personal memory. What can be done is the following: Set up an agent based EA driven model of the respective economics and code it as a computer simulation. Then run the computer simulation many times[6] and let each simulation run for a long time, i.e. very many rounds. Eventually, the last population of agents in every one of these simulations represents a result of the process after many periods of learning. The aggregate of all populations of the many different runs of the simulation then represents (due to the law of large number) a kind of mean learning result. In order to check if this result is a 'long run' result, it can be tested if two aggregate populations i.e. one recorded in period 900 000 and one recorded in period 1 000 000, are identical. If this is the case, this should be an evidence that both of these populations represent a kind of limit distribution of the learning process. As, of course, it is not completely sure if these distributions are truely the limit distribution, they will simply be called 'long run distributions' in this paper.

There is a problem to this method, though: Due to technical restrictions, it is not possible to set up simulations for mutation probabilities 'approaching zero'. The best thing that can be done in EAs is to use very small mutation probabilities in order to record the long run results of the respective simulations. Thus, the long run distributions do represent an approximation for limit distributions, but not for the limit distribution of a process whose mutation probability approaches zero.

### 3.5 EA Simulations

In order to show that an agent based model with EA dynamics is able to fully resemble the results analytically obtained by Vega-Redondo (1997) (and, for a broader range of dynamics, by Schenk-Hoppé 2000), the following model is set up. The economic model is the one introduced above (Sec. 2). The agent based setup is a

or Walras.

[6] Of course, in order to produce sensible results, each simulation run has to be started with a different initial setting of the random number generator

simple evolutionary algorithm with each agent fully characterized by her one shot strategy, i.e. the quantity $s_{i,t}$. Strategies are coded as real valued numbers.

Learning by imitation is modeled by a selection operator, which displays extreme evolutionary pressure: In every period, agents adopt the strategy of the agent that performed best in the last period, i.e. the agent with the highest payoff in the period before. This kind of 'imitate the best' replication displays more selective pressure than the usual 'biased roulette wheel'–replication (Goldberg, 1989), which is equivalent to replication in the usual discrete time non overlapping generations replicator dynamics (Samuelson, 1997, pp. 63). Nevertheless, this type of replication has all characteristics needed to belong to the class of replication processes capable of generating Vega-Redondo's (1997) results.[7] As usual in EAs, replication is a process of drawing with replacement agents (i.e. strategies) from the old population and copying them into the new. This process is repeated $n$ times, i.e. there is one draw per agent in the population. The replication probability $P_{i,t}$ is the probability of an agent playing strategy $s_{i,t}$ to be drawn and thus replicated into the next population. For the 'imitate the best' replication used in this paper, replication probabilities are given by

$$P_{i,t} = \begin{cases} 1 & \text{for} \quad \pi_{i,t} = \max_j \left\{ \pi_{j,t} \right\}, \\ 0 & \text{for} \quad \pi_{i,t} < \max_j \left\{ \pi_{j,t} \right\}. \end{cases} \tag{17}$$

It can be seen from (17) that this type of replication is a quasi–deterministic process. The driving force, though, is relative rather than absolute payoff. This means that (17) implies spiteful behavior of the agents.

Learning by mistakes or experiments, i.e. mutation, is modeled by agent's switching to a random quantity from the definition set of quantities $\mathcal{S}$. Let $s'_{i,t}$ denote the strategy agent (firm) $i$ learned by imitation. Then, with the small mutation probability $\varepsilon$, the agent switches to a different strategy $s_{i,t}$, which is randomly drawn and i.i.d. in $\mathcal{S}$:[8]

$$s_{i,t} = \begin{cases} s'_{i,t} & \text{with probability } 1-\varepsilon, \\ s \sim \text{i.i.d.} \in \mathcal{S} & \text{with probability } \varepsilon. \end{cases} \tag{18}$$

Altogether, the EA consists of many rounds, which themselves consist of a 'playing mode' and a 'learning mode' (Binmore and Samuelson, 1994; Binmore et al., 1995). The playing mode means firms selling their quantity at the market, thus collecting their payoffs and experiencing the quality of their strategies. During the learning mode, agents update their strategies by replication and mutation. EAs are generally a mere repetition of these two modes.

Note, that the learning behavior this type of dynamics implies to the agents is a very 'low rationality' kind of learning. All the information the agents need is the information which strategy was the best in the last period. Agents do not even

[7] In fact, this is the same process as the one used by Vega-Redondo (1996, pp. 128).

[8] This type of mutation is used by e.g. Binmore et al. (1995).

need to remember what they did themselves last period. All that agents need to do then, is imitate the previously best strategy. This is the reason why social learning processes modeled by EAs or replicator dynamics are processes for *very* boundedly rational agents.

In order to check for the long run results of this learning process, 5 000 EA simulations with the operators described above are run for 1 000 000 periods each, using a mutation probability of $\varepsilon = \frac{1}{100\,000}$. In each simulation, the populations in period 900 000 and in the last period, i.e. period 1 000 000, have been recorded. Then out of the 5 000 period–900 000 and the 5 000 period–1 000 000, one aggregate population each is generated. If these giant populations should both represent the limit distribution, they should not differ from each other. Thus, these populations, in form of probability distributions of supplied quantities, are tested for equality. The result is, that these populations are indeed extremely similar. Then, the period–1 000 000 population is plotted as a histogram representing the long run distribution of quantities. This plot is given in Fig. 2. The economic parameters

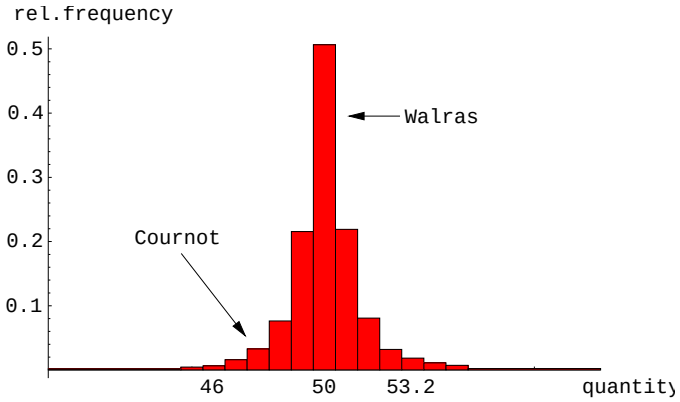


Figure 2: Long Run Distribution of Results in a Model of Social Evolutionary Learning

underlying the simulation models are: $A = 1\,000$, $B = 10$, $\delta = 1$, $n = 10$. Consequently, the Walras–quantity should be $s^W = 50$, the Cournot–quantity $s^C = 47.6$. It is obvious that the Walras–quantity is the most frequent one, while the Cournot–quantity is not too frequent. This result is perfectly in line with the theoretical results by Vega-Redondo (1997) and shows that an EA simulation is an appropriate tool to simulate the respective evolutionary dynamics.[9]

[9] Statistical characteristics of the distribution given in Fig. 2 are: mean = 50.07, median = 50.00, std.dev. = 2.23.

### 3.6 Summary: Social Learning

The optimal choice of behavior in social learning variants of the Cournot game is the Walrasian behavior. Both types of social learning models discussed in this paper lead to this strategy. Consequently, it can be concluded that both methods represent sensible and sufficiently good forms of learning behavior. Moreover, it has been shown that EA simulations do even resemble the transitory dynamics analytically formulated evolutionary games display. Based on this notion, EAs provide a method of analyzing evolutionary learning dynamics even of such systems which are out of the scope of the traditional analytical methods.

## 4 Individual Learning

In contrast to social learning, individual learning does not require the interaction of agents during the learning mode. Agents clearly do interact while going to the market, selling their quantity and thus generating the market price as the main device of information. But after that, when it comes to updating their strategies, agents act for themselves and isolated from each other.

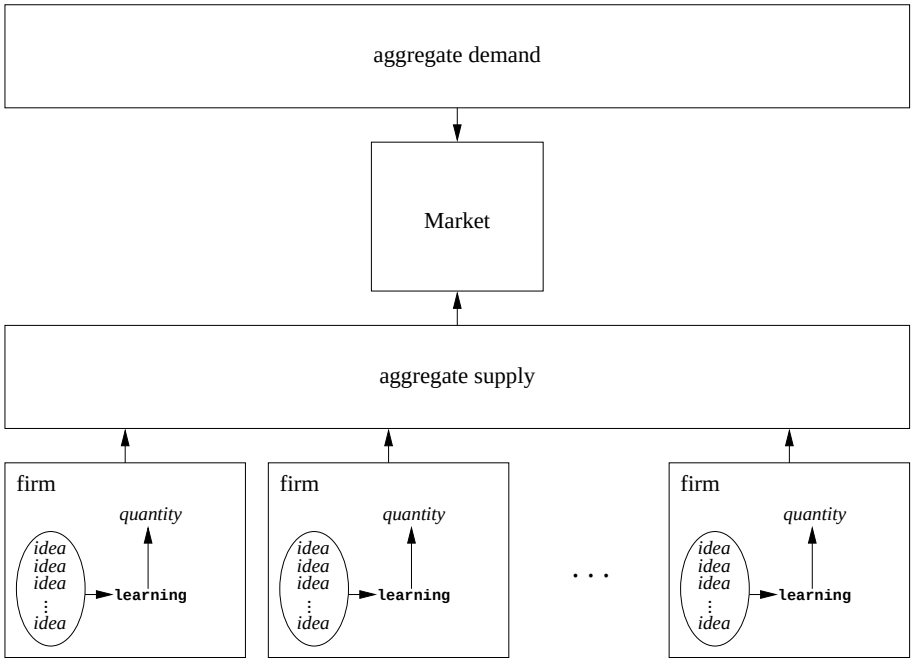


Figure 3: General Structure of Individual Learning Models

Figure 3 displays the general structure of the models for individual learning. Each firm is characterized by a set of *ideas*, i.e. a pool of potential quantities the firm could use at the market. The process of learning is applied in order to select

one of these ideas as the one to really use in the market. Agents in these models do not learn by imitating one another but rather by introspection, i.e. isolated from each other. In the market, the interaction of the agents forms the aggregate supply, which together with the exogenously given aggregate demand generates the market price. The market price is the main piece of information flowing back to the agents and thus enabling them to experience the quality of their strategy and thus to develop a new strategy for the next period. In the following, three different kinds of learning, i.e. determining a supposedly good strategy for the market, will be analyzed.

The main difference between individual and social learning in the Cournot model is the fact, that for individual learning, there is no spite effect. (See Vriend 2000 for further discussions of this point). With individual learning, the game in focus is the 'regular' Cournot game in absolute payoffs. Consequently, the optimal strategy is the Cournot strategy, which in these models forms the only stable symmetric Nash equilibrium.

But, the mere absence of the spite effect does not automatically mean that the outcome of individual learning processes cannot be Walras or even must be Cournot. The opposite is true: It will be shown that it even takes a large amount of rationality (or: sophistication) of the agents to individually learn to play Cournot. Apart from the effect of the presence or absence of spiteful behavior, there is an additional force influencing the quality of the results: the level of rationality of the agents. The more the agents know and the more sophisticated methods they use to determine their strategy, the more likely it is that the result will be Cournot. In other words: With individual learning, the Walrasian strategy turns out to be a kind of 'low rationality behavior', whereas playing the Cournot strategy requires a remarkable amount of knowledge and behavioral sophistication. In the following, three types of individual learning will be considered in order to stress this hypothesis. All of these learning methods represents learning by agents with a different degree of rationality.

### 4.1 Best Response Learning

One of the most sophisticated forms of individual learning is best response learning.[10] In order to compute a best response to the opponents' strategies of last period or, alternatively, to last period's market price, an agent needs the following information: Form and parameters of the aggregate demand function, i.e. $A$ and $B$ in our model, as well as the knowledge that $D_t$ is linear in the market price; form and parameters of the cost function, i.e. $\delta$, and the knowledge that $C(\cdot)$ quadratic; her own quantity chosen in the preceeding period, $s_{i,t-1}$ (i.e. some form of memory); and last period's aggregate supply $S_{t-1}$ or, alternatively, the market price of the last period, $p_{t-1}$.[11] Equipped with all this knowledge and information, an agent can compute her reaction function describing the best response to every possible

[10] For a textbook version of this adjustment process, see Fudenberg and Levine (1998, pp. 8–10).

[11] Due to (2) and (3), $p_{t-1}$ can be calculated from $S_{t-1}$ and vice versa.

aggregate supply $s_{i,t} = B^s(S_{t-1})$ or to every possible market price $s_{i,t} = B^p(p_{t-1})$:

$$s_{i,t} = B^s(S_{t-1}) = \frac{A - S_{t-1} + s_{i,t-1}}{B\delta + 2} \tag{19}$$

$$= B^p(p_{t-1}) = s_{i,t-1} + \frac{p_{t-1}}{(B\delta + 2)\,B}. \tag{20}$$

These highly informed agents can thus compute their best response to the given circumstances. For the model discussed in this paper, it is known that this type of 'classical' best response learning will quickly lead the dynamics into the Cournot equilibrium. Moreover, under the regime of these dynamics and for the parameters chosen in the above example, the Cournot–Nash strategy is a best response to itself, while the Walras–strategy is not. Thus, the Cournot–Nash strategy is asymptotically stable under best response dynamics, while the Walras–strategy does not even establish a (Nash–) equilibrium.

A slightly different variant of generating a similarly sophisticated best response is presented by Vriend (2000). Vriend lets each of his firms apply a *classifier system* (CFS, Holland 1992; Goldberg 1989) in order to find a best response to a series of combinations of supplied quantities and market prices connected with these quantities. Classifier systems are techniques from the field of artificial intelligence, which are known to be capable of efficiently computing optima of complicated mathematical functions. Thus, it is no surprise that Vriend finds his model to converge towards the Nash–Cournot equilibrium.

## 4.2 Cobweb Learning

If agents do not know or simply neglect the state dependent nature of the problem, i.e. the fact that they do have an influence on the market price, the outcome of individual learning will be the Walras equilibrium. The reason for this is straightforward: If agents do not think or do not know they can influence the market price, the best thing they can do is compute a best response to last period's equilibrium price. Moreover agents assume that even the reactions of other agents do not change the price. Consequently they expect the price of the current period to be the same as the price of the period before: $p_t^e = p_{t-1}$. This myopic expectation can be interpreted as a symptom of the agents' bounded rationality. Consequently, agents adjust their quantity to the seemingly given price. This is exactly the way agents are supposed to update their strategies in the classical cobweb model (dating back to the seminal work by Leontief 1934), sometimes labeled as 'naive' or simply 'cobweb' learning. This kind of behavior results in following the well–known rule 'Select exactly the quantity that equals your marginal costs and the market price', i.e.

$$s^\star_{i,t+1}: \quad p_t = \frac{\partial C\left(s^\star_{i,t+1}\right)}{\partial s^\star_{i,t+1}}. \tag{21}$$

Of course, this quantity is the Walras quantity as given in (8):

$$s^\star_{i,t+1} = s^W. \tag{22}$$

For the given model, the cobweb reaction is given as

$$s_{i,t} = \frac{p_{t-1}}{\delta}. \tag{23}$$

The required knowledge for the generation of a cobweb response is form and parameters of the cost function, or at least the function of marginal costs, i.e. $\delta s_{i,t}$ in the given model, as well as last period's market price, $p_{t-1}$. In order to find a cobweb response, agents must be capable of computing a function like the one given in (23). This act of finding a best answer to a given price can alternatively be modeled by an EA, which represents a particularly 'low rationality' method to solve this intra–agent learning task. It has been mentioned before that there exists a broad range of papers showing that EA learning in Cobweb models leads to the Walrasian outcome, i.e. Arifovic (1994); Dawid and Kopel (1998); Franke (1998).

It is a standard textbook issue[12] to prove that the Walras–equilibrium under the regime of these dynamics is asymptotically stable at least for the model and the parameter set[13] presented above.

There are more variants of the cobweb model, which mainly differ in the way agents form their expectations about the current market price. All of these models are known to eventually converge to the Walras equilibrium, even if the number of players is finite. The key to this behavior seems to be the level of ignorance of the agents: As long as agents do ignore their personal influence on the price, they reach the Walras solution.

### 4.3 State Dependency Learning

In order to make the above point clearer, a model of 'medium rationality' will be set up. Agents know that they can influence the market price with their decision. And, although they do not know the exact form of the aggregate demand function, they will be equipped with the knowledge of the elasticity of demand, or, to be more exact, the slope of the demand function $\frac{\partial D_t}{\partial s_{i,t}}$. Thus, agents are not too smart, but at least they know that the problem they face is a state dependent one. Consequently, just in order to give a name to this learning method, this type of learning will be labeled 'state dependency learning'.

Again, agents try to find a best response to the market price, but this time they do take into account their influence on the price. (What they still neglect, though, is the fact that there is an influence of all other agents on the price, as well.)

Thus, what agents do is try to find a quantity $s_{i,t}$ which will maximize their expected profits $\pi^e_{i,t+1}$ while keeping in mind that the current price will be changed by their quantity decision. Thus, they optimize under the assumption that the current price results as last period's price minus their own change in supply times the slope of the demand curve:

$$\max_{s_{i,t}} \pi^e_{i,t} \quad = \quad p^e_t s_{i,t} - C(s_{i,t}) \tag{24}$$

[12] See e.g. Chiang (1984, pp. 561).

[13] The parameter set establishes a so called 'cobweb stable' situation.

$$\text{s.t.} \quad p_t^e = p_t - \alpha \Delta s_{i,t}$$
$$\text{with } \alpha = \frac{\partial D_t}{\partial s_{i,t}}$$
$$\text{and } \Delta s_{i,t} = s_{i,t} - s_{i,t-1}\,.$$

Thus, the information needed in order to compute this kind of response is last period's market price, the price elasticity of demand, and form and parameters of the cost function. In the simulations constructed in order to generate the long run distribution of strategies, each agent tries to solve (24) with the help of an evolutionary algorithm. Simulations were run for the same parameter set as before, i.e. for $A = 1\,000$, $B = 10$, $\delta = 1$, $n = 10$. Again, these parameters establish the following theoretical benchmark results: In the model, the Walrasian quantity is $s^W = 50$, and the Cournot quantity is $s^C = 47.6$. For these parameters, 1 000 simulations were run over 5 000 periods with a mutations probability of $\varepsilon = \frac{1}{10\,000}$ in order to record the long run distribution of learned quantities. Figure 4 shows the results.[14] The figure makes clear that under the regime of state dependency learning, in the long run, agents tend to play Cournot, not Walras.

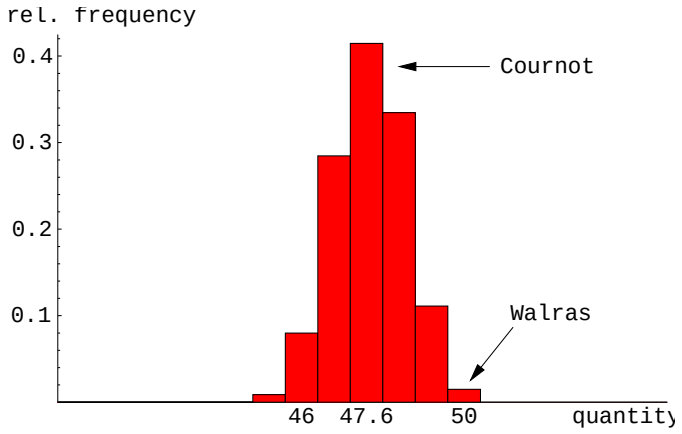


Figure 4: Long Run Distribution of Results in a Model of Individual State Dependency Learning

Note that agents in this model posses a medium level of rationality. They need more information than for cobweb learning, but not as much as for best response learning. But agents are aware of the fact that they have a non–negligible influence on the market price (although they ignore the fact that other agents have this influence as well). This seems to be the key information needed to generate Cournot instead of Walras outcomes in models with individual learning.

[14] Statistical analysis results in: mean=47.68, median=47.75, std.dev.=0.89.

# 5 Summary: Degrees of Rationality and Learning Results

It should have become clear that the long run result of learning in the Cournot game crucially depends on the type of learning and, if learning means individual learning, on the degree of the agents' rationality. Social learning, due to the spite effect, always leads to the Walrasian outcome. With individual learning, agents need a positive minimum degree of information and analytical sophistication to be able to learn to play Cournot. The degree of rationality, in this paper, is measured by three things: The amount of information needed to conduct the respective learning method, the question if agents need a personal memory, and the computational abilities an agent needs.

Table 3 summarizes the results of the paper.

| Learning Method | Level of Rationality | Necessary Information | Memory needed | Necessary computational abilities | Result |
|---|---|---|---|---|---|
| **social** | | best quantity | no | imitation of others | Walras |
| **individual** | | | | | |
| cobweb | low | cost function: marginal costs<br>price | no | maximization | Walras |
| state dependency | medium | agg. demand: elasticity<br>cost function: form and parameters<br>price | quantity | maximization or EA | Cournot |
| best response | high | agg. demand: form and parameters<br>cost function: form and parameters<br>agg. supply or price | quantity | maximization or CFS | Cournot |

Table 3: Summary

It can be clearly seen that social evolutionary learning requires the least level of rationality. Agents only need a minimum of information and capabilities. This type of learning leads to the Walras outcome. Best response learning, as the other extreme case, needs agents that can aquire a large amount of information, hold some personal memory and have command of some very sophisticated technical methods. This is clearly the most 'high rationality' learning scheme and, consequently, leads to the Cournot outcome.

To put the results of the paper into one sentence: The more sophisticated agents are, the more likely they are to learn the Cournot strategy.

## References


Alchian, A. A. (1950). Uncertainty, evolution, and economic theory. *Journal of Political Economy*, 58:211–221.

Arifovic, J. (1994). Genetic algorithm learning and the cobweb–model. *Journal of Economic Dynamics and Control*, 18:3–28.

Binmore, K., Gale, J., and Samuelson, L. (1995). Learning to be imperfect: The ultimatum game. *Games and Economic Behavior*, 8:56–90.

Binmore, K. and Samuelson, L. (1994). An economist's perspective on the evolution of norms. *Journal of Institutional and Theoretical Economics*, 150:45–63.

Chiang, A. C. (1984). *Fundamental Methods of Mathematical Economics*. McGraw–Hill, Auckland, Bogota et. al., 3 edition.

Dawid, H. (1994). A Markov chain analysis of genetic algorithms with a state dependent fitness function. *Complex Systems*, 8:407–417.

Dawid, H. (1999). *Adaptive Learning by Genetic Algorithms*. Springer, Berlin, Heidelberg, New York, 2 edition.

Dawid, H. and Kopel, M. (1998). On economic applications of the genetic algorithm: A model of the cobweb type. *Journal of Economic Dynamics and Control*, 8:297–315.

Foster, D. and Young, H. P. (1990). Stochastic evolutionary game dynamics. *Theoretical Population Biology*, 38:219–232.

Franke, R. (1998). Coevolution and stable adjustments in the cobweb model. *Journal of Evolutionary Economics*, 8:383–406.

Fudenberg, D. and Levine, D. K. (1998). *The Theory of Learning in Games*. MIT Press Series on Economic Learning and Social Evolution. MIT Press, Cambridge, MA.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison–Wesley, Reading, Massachusetts.

Holland, J. H. (1992). *Adaptation in Natural and Artificial Systems*. MIT Press, Cambridge, MA, London, 2 edition.

Kandori, M., Mailath, G. J., and Rob, R. (1993). Learning, mutation, and long run equilibria in games. *Econometrica*, 61:29–56.

Kandori, M. and Rob, R. (1995). Evolution of equilibria in the long run: A general theory and applications. *Journal of Economic Theory*, 65:383–414.

Leontief, W. W. (1934). Verzögerte angebotsanpassung und partielles gleichgewicht. *Zeitschrift für Nationalökonomie*, 5:670–676.

Riechmann, T. (1999). Learning and behavioral stability — An economic interpretation of genetic algorithms. *Journal of Evolutionary Economics*, 9:225–242.

Riechmann, T. (2001a). Genetic algorithm learning and evolutionary games. *Journal of Economic Dynamics and Control*, 25(6–7):1019–1037.

Riechmann, T. (2001b). *Learning in Economics. Analysis and Application of Genetic Algorithms*. Physica–Verlag, Heidelberg, New York.

Riechmann, T. (2001c). Two notes on replication in evolutionary modelling. Diskussionspapier 239, Universität Hannover, Fachbereich Wirtschaftswissenschaften.

Samuelson, L. (1997). *Evolutionary Games and Equilibrium Selection*. MIT Press Series on Economic Learning and Social Evolution. MIT Press, Cambridge, MA, London.

Schenk-Hoppé, K. R. (2000). The evolution of Walrasian behavior in oligopolies. *Journal of Mathematical Economics*, 33:35–55.

Vega-Redondo, F. (1996). *Evolution, Games, and Economic Behavior*. Oxford University Press, Oxford, UK.

Vega-Redondo, F. (1997). The evolution of Walrasian behavior. *Econometrica*, 65:375–384.

Vriend, N. J. (2000). An illustration of the essential difference between individual and social learning, and its consequences for computational analyses. *Journal of Economic Dynamics and Control*, 24:1–19.

Weibull, J. W. (1995). *Evolutionary Game Theory*. MIT Press, Cambridge, MA, London.

Young, H. P. (1993). The evolution of conventions. *Econometrica*, 61:57–84.